

Assessment in the Language Classroom

Liying Cheng & Janna Fox

Series Editor: Andy Curtis



Assessment in the Language Classroom

Assessment in the Language Classroom

Teachers Supporting
Student Learning

Liying Cheng

*Professor of Language Education,
Queen's University, Canada*

Janna Fox

*Professor of Applied Linguistics & Discourse Studies,
Carleton University, Canada*

Applied Linguistics for the Language Classroom

Series Editor: Andy Curtis



palgrave

© 2017 Liying Cheng and Janna Fox

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 or 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.

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.

First published 2017 by
PALGRAVE

Palgrave in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of 4 Crinan Street, London, N1 9XW.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN: 978–1–137–46483–5 paperback

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

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

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

Contents

<i>List of Figures and Tables</i>	vi
<i>Series Editor's Introduction</i>	viii
<i>Acknowledgements</i>	xii
<i>Introduction</i>	xiii
1 Why Do We Assess?	1
2 What Do We Assess?	30
3 How Do We Assess?	62
4 How Do We Develop a High-Quality Classroom Test?	102
5 Who Are We Assessing? Placement, Needs Analysis and Diagnostics	139
6 Who Are We Assessing? Feedback and Motivation	166
7 When We Assess, How Can We Use Assessment to Move Forward?	188
<i>Appendix</i>	215
<i>Glossary</i>	223
<i>References</i>	231
<i>Index</i>	241

List of Figures and Tables

Figures

1.1	Dimensions of assessment	7
2.1	Alignment of learning goals, assessment and classroom activity	31
2.2	Specific curriculum expectations in listening and speaking	33
2.3	Aligning course activity and assessment tasks with intended learning outcomes through horizontal and vertical course planning and design	45
3.1	Assessment observations, conversations and products	76
4.1	Overview of a test development process	109
5.1	An example of an online diagnostic assignment	154
5.2	Follow-up diagnostic assignment	155
5.3	Diagnostic assessment: student profile of targeted needs	158

Tables

1.1	Purposes of assessment and evaluation	10
1.2	The commonplaces of language curriculum	21
1.3	Questionnaire: What is most important in teaching a language?	23
1.4	My philosophy of teaching and learning	27
2.1	Course example: ESLCO, Level 3 (Open), Listening and Speaking	46
3.1	Reading	78

3.2	Writing	79
3.3	Speaking and listening	80
3.4	Classification of Marta's portfolio requirements	89
4.1	History file (excerpt)	112
4.2	Table of specifications (sample specification)	113
4.3	Some commonly used item and task formats	117
4.4	Holistic scale – Section 1: summary writing	124
4.5	Analytic scale – Section 1: summary writing	126
4.6	Item analysis for Class 6B (Level 3)	130
5.1	Mapping assessment approaches onto philosophies of needs assessment	149
5.2	Diagnostic approaches	159
5.3	A sample diagnostic assessment tool	163
6.1	Assessment practices, teacher responses are in italics, students' work is underlined	169
7.1	Principles of test preparation practices	206
7.2	Overarching themes and code frequencies	210

Series Editor's Introduction

The purpose of this Applied Linguistics for the Language Classroom (ALLC) series is to help bridge what still appears to be a significant gap between the field of applied linguistics and the day-to-day classroom realities of many language teachers and learners. For example, Selivan recently wrote that: “Much applied linguistics research remains unapplied, is often misapplied, or is downright inapplicable” (2016, p.25). This gap appears to have existed for some time, and has yet to be bridged. For example, in 1954, Pulgram published *Applied Linguistics In Language Teaching*, which was followed a few years later by Robert Lado’s classic work, *Linguistics Across Cultures: Applied Linguistics for Language Teachers* (1957). However, we are still seeing articles 60 years later helping language teachers to apply linguistic theory to language lessons (Magrath, 2016).

Therefore, one of the features of this ALLC series that makes it distinctive is our focus on helping to bridge the on-going gap between applied linguistics and language classrooms. Our envisaged readership for these books is busy classroom language teachers, including those entering the profession and those who have been in it for some time already. We also gave a lot of thought to what teachers completing a first degree in Education, teachers doing MA TESOL courses, and language teachers completing other professional qualifications, would find most useful and helpful.

Bearing such readers in mind, one of the ambitious goals of this ALLC series is to present language teachers with clear, concise and up-to-date overviews and summaries of what they need to know in key areas: Assessment; Methods and Methodologies; Technology; Research Methods; and Phonetics, Phonology and Pronunciation. Attempting to do what

much larger and weightier volumes have attempted, but doing so in volumes that are slimmer and more accessible, has been a challenge, but we believe these books make an original and creative contribution to the literature for language teachers.

Another distinctive feature of this ALLC series has been our International Advisory Board, made up of Professor Kathleen Bailey and Professor David Nunan. These two outstanding figures in our field helped us to keep our target readers in mind and to stay focused on the classroom, while keeping the connections to applied linguistics, so we can advance the building of the bridges between applied linguistics and language classrooms.

In *Assessment in the Language Classroom: Teachers Supporting Student Learning* Liying Cheng and Janna Fox – two educators with more than 50 years of classroom experience between them – focus on how assessment can be used as a tool to help teachers help learners. Each chapter starts with a set of schema-activating questions, which encourage the readers to reflect on their own experiences before delving further into the text. Chapter One starts with definitions and the different purposes of some essential aspects of assessment. One of those aspects is the voices of test-takers, which are often conspicuous by their absence, but in this book, those voices are heard.

In Chapter Two, Cheng and Fox address the big and important but complex and complicated questions: What is worth teaching, and what is worth assessing? To help answer those questions, Cheng and Fox give concrete examples of specific curriculum expectations in relation to particular language modalities. The authors also help teachers align their classroom activities and assessments with the learning goals and outcomes. As in Chapter One, definitions and descriptions of key terms, such as 'learning outcomes' – which form the basic vocabulary items of the language of language assessment – are presented. The importance of contextual understanding is also highlighted in this chapter, which concludes with a very helpful template for planning a course.

Like the other books in this ALLC series, one of the main the goals of this book is to help language teachers in their day-to-day classroom practices. Therefore, Chapter Three starts by comparing and contrasting large-scale testing with classroom testing, as a way of explain the three key concepts of assessment of learning, assessment *for* learning, assessment *as* learning. To complement the voices of language test-takers in the first chapter, the voices of language teachers are presented here, and reiterating the importance of contextualization, detailed assessment plans – which are “an overall guide for how we will assess students’ achievement of the learning goals and outcomes relevant to instruction” (p. 66) – from two different contexts are given. Details of a wide range of assessment tools are given, including alternative assessment methods, such as use of portfolios, again, with detailed examples.

In more and more countries, classroom teachers are being asked to develop their own tests, to analyse tests, and to be able to evaluate the quality of tests. Unfortunately, in many of those countries, the initial teacher training programs do not include required courses on test design, analysis or evaluation. Chapter Four, therefore, helps teachers to build this knowledge and these skills, including the expanding of teachers’ language assessment vocabulary, with thorough explanations of ‘construct definition’, ‘criterion-’ and ‘norm-referenced assessment’, ‘Target Language Use’ and other terms which teachers need to know in order to be able to think, talk about and do assessment. A detailed overview of the test development process is also presented.

Chapter Five looks at *who* we are assessing, in relation to placement, needs analysis and diagnostics. In this chapter, Cheng and Fox start by emphasizing the importance of teachers and learners getting to know each other, which is not something that cannot happen with large-scale, standardized testing, in which all test-takers all over the world are, by definition, assumed to be the same. However, as we know, every learner and every teacher is unique, putting the teacher-learner relationship at the heart of classroom assessment. Again, the

importance of classroom context is shown, with a comparison of diagnostic assessment in a conversation class versus diagnostic assessment in an English for Academic Purposes (EAP) class.

In the penultimate chapter, the focus on the test-taker, in terms of who we are assessing, is expanded upon, this time in relation to feedback and motivation. The importance of the teacher-student connection, and the value of on-going feedback, in determining the quality of teaching and learning is discussed, as well as the use of assessment practices to motivate students to learn. And in the seventh and last chapter, Cheng and Fox address another essential question, but one that is not always considered in the traditional testing literature: "When we assess, how can we use assessment to move forward?" (p. 188). *Assessment in the Language Classroom: Teachers Supporting Student Learning* is filled with examples and activities, questions and answers, advice and guidance, all of which will help language teachers use assessment in their classrooms as something that is done *with* learners not *to* them.

Andy Curtis

Acknowledgements

Many people have contributed to the creation of this book. First, we would like to thank all the teachers and students whom we have worked with over the years in our endeavours to use assessment to support teaching and learning. It is all of you who have provided the inspiration to write this book. We thank the Series Editor Andy Curtis for encouraging us to write this book, and for his valuable feedback. Thanks also go to all the editorial staff at Palgrave, especially Paul Stevens and Cathy Scott. Most importantly, we are grateful to our family members for putting up with us working on this book over holidays, evenings, and weekends. We could not have written this book without your unconditional support.

Introduction

To the Readers of this Book: Personal Reflections

Like you, we are teachers. We both began our careers teaching English to students whose first languages were not English. We taught many of these students in Asia, the Middle East, Africa and North America, navigating our way through the teaching, learning and assessment of our students with little guidance from theory or resources. Over the years, we became increasingly sensitive to the negative influence and consequences of ill-considered assessment and testing practices. Although we could increasingly find resources on language teaching methods, strategies and techniques, very few of these resources provided systematic and coherent support for our assessment practices. There were no readily available answers to our questions in the research literature either – researchers tended to write for other researchers and their findings, although interesting, were not readily applicable in our classrooms.

Years later, our long-term interest in assessment led us to teaching courses to pre-service and in-service teachers: helping them to support their students' learning through sound assessment practices. Again, we searched for resources that could speak to the questions and issues arising in the classroom; we realized that the narrow scope of resources on classroom assessment rarely moved beyond test design and test analysis, and they were more appropriate for large-scale testing than for ongoing classroom assessment. That is why we decided to write this book – a book on assessment which is specifically for teachers. It is directly informed by our own teaching experience and by research that we have drawn on or engaged in over the years.

The book focuses on the integration of teaching and assessment in supporting student learning, with the notion of *alignment* as a guiding principle for high-quality assessment, that is, the degree of agreement among standards, curriculum, assessment (including tests) and instruction. Alignment, along with validity, reliability, fairness, consequences and practicality, is viewed as a central aspect of assessment practice which supports learning. We discuss these key assessment concepts throughout the book.

The book consists of seven chapters, which address fundamental assessment questions raised by teachers. The focus of each chapter is summarized below:

- **Chapter 1: Why do we assess?**

In our first chapter, we define assessment and the important role that teachers play in using assessment to support student learning. This chapter also unpacks key characteristics of high-quality assessment and allows teachers to explore their own philosophy of assessment and to consider how this impacts what they choose to do in the classroom. We introduce the important concept of *alignment*, namely the degree of agreement between standards, curriculum, tests and classroom instruction, and explore how our own teaching and assessment philosophies can influence the choices we make in our classrooms.

- **Chapter 2: What should we assess?**

In [Chapter 2](#) we discuss *what* should be assessed. We examine how to integrate learning goals or outcomes with assessment tasks and classroom activities so that they are aligned. We apply the *backward design* approach (see Wiggins and McTighe, 2005), defining our learning outcomes first, and then working backward from outcomes to classroom assessment tasks and activities. We discuss how such an approach would work in a range of language teaching contexts as we align our classroom assessment tasks with benchmarks, standards, curriculum and tests, which are defined externally; and/or with the needs and interests of the learners we are teaching, which are specific to our own classroom and arise internally.

- **Chapter 3: How do we assess?**

We devote two chapters for this discussion. [Chapter 3](#) focuses on formative assessment or assessment *for* and *while* learning and

how teachers develop assessment plans which support the learning outcomes for a course. We examine the range of assessment tools, methods, and procedures that support learning, including some so-called alternative methods, taking a close-up look at portfolio assessment. Throughout we consider the important notions of validity and reliability which are central to fair and ethical assessment practices.

- **Chapter 4: How do we develop a high-quality classroom test?**

In [Chapter 4](#) we focus on one of the assessments of learning: achievement or summative assessment. We examine how to develop high-quality tests, step by step from specifications to administration, and the role such test development can play in monitoring and supporting effective learning in our classrooms and promoting professional development and coherence in our programmes. We discuss the evidence we can use from test development and analysis to support claims of validity and reliability in this pervasive classroom assessment practice, and why such evidence is essential for high-quality testing.

- **Chapter 5: Who are we assessing? Placement, needs analysis and diagnostics**

[Chapter 5](#) and [Chapter 6](#) examine assessment practices that help us know more about our students' language learning experiences, capabilities, goals and motivation to learn. These practices can greatly enhance the effectiveness of our teaching. We focus on placement testing and needs analysis, and link our discussion of these assessment tools to the use of diagnostic assessment practices in our classrooms which relate individual results to specific (targeted) individual learning activities. We examine how computer-based learning management systems can support our use of diagnostic assessment within a classroom, across a programme and across a system.

- **Chapter 6: Who are we assessing? Feedback and motivation**

In [Chapter 6](#) we again focus on *who* we are assessing and the role that feedback and motivation can play in high-quality learning. We consider the parallel roles of teacher feedback, which serves to both coach learning-in-progress, as well as judge the quality and degree of efficiency that is attained as an outcome. We also discuss the complex interaction between assessment and a student's motivation to learn and consider strategies to support our students' use of feedback from assessment.

- **Chapter 7: When we assess, how can we use assessment to move forward?**

Chapter 7 is our final chapter and provides an overview of the key themes which recur throughout the book. We discuss how we can increase the positive impact of assessment on our students' learning. We examine grading and grading practices, test preparation (particularly in the context of high-stakes testing), and the importance of drawing on the testing experiences of our students to inform our teaching and assessment practices.

We, too, will use this book in our own classrooms – teaching teachers and graduate students about classroom assessment. We hope it will support your own learning and stimulate your further interest in assessment practices.

Background to the Field of Language Testing and Assessment: A Brief Overview

Language testing and assessment is a relatively new field within the broad context of applied linguistics. The terms *testing* and *assessment* are both used in this book to show the historical development of our field. Although, initially, language testing tended to focus on *testing* alone, the focus has moved in recent years to *assessment*, which includes testing. We thus discuss both aspects in this book, as both are day-to-day requirements for teachers like you.

We'd like to point out that the major theoretical frameworks that guide the field of language testing and assessment are derived from educational measurement based on classical and modern test theory (e.g., Bachman, 1990; Linn, 2010) and theories from applied linguistics based on communicative language teaching and learning (e.g., Canale and Swain, 1980; Savignon, 2003). Since then, the focus of the field has shifted its emphasis in two important ways: (1) to a consideration of what is at stake for learners learning the English language and their performance as test-takers; and (2) from an exclusive

focus on tests and testing to a broader consideration of assessment practices in language classrooms. Our book reflects these major changes and speaks to *language teachers who are working with students/test-takers in day-to-day classroom contexts*. Although the focus of the book is on language testing and assessment of the English language, the book also addresses broader issues that can apply to any language classroom.

Over the life span of the field, four types of resource have been available to those who are interested in language testing and assessment. The readers of these resources are: teachers who work with students and use testing and assessment in their classrooms as part of their ongoing practice; researchers (including Master's and PhD students) who conduct testing and assessment studies; and testing professionals who design and validate tests, although many of the scholars and practitioners in the field have overlapping roles as teachers, researchers and testing professionals.

The first type of resource is *scholarly*, for example, Bachman's (1990) *Fundamental Considerations in Language Testing* and Weir's (2005) *Language Testing and Validation*, which are read mostly by researchers and testing professionals (note that both have 'testing' in their titles). The second type is *textbook* in nature, for example, Bailey and Curtis's (2015) *Learning About Language Assessment*, Brown's (1996) *Testing in Language Programs*, Fulcher's (2010) *Practical Language Testing* and Douglas's (2010) *Understanding Language Testing*. These textbooks are read by in-service teachers and graduate students who are language teachers or starting out as language researchers. This type of textbook focuses on teaching the basics of language testing and assessment.

The third type of resource is a collection of *articles* on key aspects of language testing and assessment, for example, the *Encyclopedia of Language and Education: Language Testing and Assessment* (2008), the *Cambridge Guide to Second Language Assessment* (2012) and the most recent *Companion to Language Assessment* (2014). These books tend to have short, technical

texts on testing. Classroom assessment is typically relegated to one or more chapters in this selection. The fourth type of book features a *specialized focus* on topics of interest in language testing. There are two primary series of this type: *Studies in Language Testing* (e.g., Cheng, 2005 on washback) and the *Cambridge Language Assessment* series (e.g., Alderson, 2000 on *assessing reading*). While the third and fourth types are articles or books on specific topics in language testing and assessment, the first two types are general books on language testing and assessment.

Our book is best described as a *textbook* – but differs from what is currently available in this category by using teacher-friendly activities written in non-technical and more accessible language – combined with the more scholarly work associated with type-three resources (i.e., *collections of articles*). Recommended readings are identified in the book so that they may be further explored to more deeply inform our assessment practices. It introduces the core concepts of language testing and assessment *in context*, as they arise in practice. We guide our fellow teachers towards what to consider and what to do, given a learning context, and link our consideration of these concepts to empirical research articles that have investigated them.

We consider a central strength of the book is our formulation of assessment *of* learning, assessment *for* learning and assessment *as* learning. We argue that it is inaccurate to view assessment only as judgments on learner progress at the conclusion of a unit of teaching/learning. Rather, it should also be viewed as a way of obtaining evidence for where students are in reaching their learning goals, and what they need in order to progress towards these goals. Assessment *as* learning puts the focus on the students themselves, taking responsibility for their own learning through self- and peer-assessment, monitoring their own progress towards their goals, and employing strategies for achieving them. The principles of *alignment* and assessment *of*, *for* and *as* learning guide our discussion throughout the book and empower effective assessment practices in our classrooms.

Activate your learning

- What is assessment?
- Why do we assess as teachers? Why do our students carry out assessment as learners?
- What does high-quality assessment look like?
- What is your philosophy of assessment?
- Why is it important to reflect on your own views of: how best to assess, what to assess, when to assess, why to assess, and with whom?

1.1 Definitions and Dimensions of Assessment

Assessment plays an essential role in language teaching and learning. The day-to-day assessment of student learning is unquestionably one of the teacher's most important, complex and demanding tasks. As teachers, we are the principal agents of assessment, so we need to ensure the quality of classroom assessment practices and to use these practices in ways that best support our students' learning.

What is assessment? Assessment refers to 'all those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used as **feedback** to modify the teaching and learning activities in which they are engaged' (Black and Wiliam, 1998, p. 2). There are two important points about this definition. First, assessment involves teachers and students. An assessment activity can be carried out between a teacher and a student or a group of

students. It can also be carried out between a student and another student or among students themselves. Second, these activities include both the classroom tests and daily assessments that we teachers use in our classroom, and the large-scale tests that our students take such as the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS). In this sense, assessment has multiple dimensions, as do our instructional activities.

Assessment activities include *events*, *tools*, *processes* and *decisions* (Taylor and Nolen, 2008) – four fundamental aspects of classroom assessment:

- *Assessment events*, such as when a teacher organizes an oral presentation or a listening activity, can support students when the events occur with the right frequency, so that the teacher knows whether instruction is successful, which areas need more instruction, and which student or group of students may need additional support. As teachers, we need to consider how many assessment activities we use and how frequently we use them in our classroom. ***How do we judge the right amount and frequency of our assessment activities, and when should those activities take place?*** We examine responses to this question in [Chapters 3](#) and [4](#).
- *Assessment tools* can support student learning when the tools give students clear ideas about what is important to learn and the criteria or expectations for ‘good’ work, and when assessment goals are aligned with instructional goals. Assessment tools are important in that they are the instruments we use to collect evidence of students’ learning. Choosing an assessment tool is like choosing a scale to measure our body weight, a tape to measure the size of a desk, or making notes on the advantages (and disadvantages) of a new textbook we are choosing for a class. In our own classrooms, we may assess students’ ability to communicate by using performance tasks, such as making an oral presentation or writing a report. We may assess our students’ ability to read by using multiple-choice **items** on a test of reading comprehension or by asking them to read a passage aloud and explain what it means. The tool we use to assess has to fit with

both the type of learning that is being measured, and the type of learning that is the goal of our instruction. ***How do we ensure that we choose the right assessment tools to measure what we intend our students to learn?*** We address this question in [Chapter 2](#).

- *Assessment processes* can support students' views of their teachers as allies in their education; feedback can help students focus and better understand the requirements of a task. Assessment processes are about *how* we carry out our assessment activities. We need to remember that the central purpose of assessment is to support student learning, so in one way or another, our assessment has to centre on this purpose. ***How do we provide feedback to our students? How much feedback and what type of feedback will best support our students' learning?*** We respond to these questions in [Chapter 6](#).
- *Assessment decisions* can support students' learning when grades accurately reflect what students know and can do. We make a range of decisions based on the results of our assessment. These decisions range from micro-level course decisions, such as what we need to do more or less of in a follow-up lesson, to macro-level decisions, which have important (even life-changing) consequences for our students, such as deciding which class a student should be placed in or whether a student can be admitted into a university. ***What informs your assessment decisions? How can we ensure that our decisions are supporting our students' learning?*** [Chapters 3–7](#) address these questions.

As Taylor and Nolen (2008) rightly point out:

Assessment tools and processes have tremendous power in the lives of students – both in the ways that students come to judge their own abilities and in the ways that they represent what is important to know and be able to do. (p. 95)

As mentioned earlier, *assessment* is an umbrella term, which includes both our daily classroom assessment practices and large-scale testing, which is externally designed and administered to our students. In many cases, being successful in an

external, large-scale testing context is one of the learning goals of our students, which in turn influences our own instructional and assessment goals.

The following two terms best represent these dimensions of assessment:

- **Assessment *for* learning** refers to the process of seeking and interpreting evidence for use by students and their teachers to decide *where* students are in their learning process, *where* they need to go and *how* best to get there.
- **Assessment *of* learning** refers to assessments that happen *after* learning has occurred, to determine whether learning has happened. They are used to make statements about a student's learning status *at a particular point in time*.

As teachers, we engage in both dimensions of assessment practices. In fact, the synergy of assessment *for* learning punctuated with the use of assessment *of* learning is the best way to support our students' learning.

There are other ways of defining assessment, for example, formative assessment and **summative assessment**. According to Black and Wiliam (2009), formative assessment encompasses the following processes:

Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction. (p. 9)

Alternatively, Cizek (2010) refers to formative assessment as:

The collaborative processes engaged in by educators and students for the purpose of understanding the students' learning and conceptual organization, identification of strengths, diagnosis of weaknesses, areas of improvement, and as a source of information teachers can use in instructional planning and students can use in deepening their understanding and improving their achievement. (p. 6)

Formative assessment is in line with assessment *for* learning, where assessment processes are carried out collaboratively, and the assessment decisions are primarily about the direction in which teaching and learning should go. So typically there should not be a mark associated with formative assessment. Sometimes, as an incentive, teachers may award marks for participating in a process or completing a stage of an activity (e.g., participating in a writing conference, leading a group discussion). Such a practice may be problematic as it uses formative assessment for summative assessment purposes therefore changing the nature of the assessment. **Summative assessment** is used to evaluate student learning, skill acquisition and academic achievement at the conclusion of a defined instructional period – typically at the end of a project, unit, course, semester, programme, or school year. Summative assessment is in line with assessment *of* learning. Generally speaking, summative assessment is defined by the following three major criteria:

- Tests, assignments, or projects are used to determine whether students have learned what they were expected to learn. In other words, what makes an assessment ‘summative’ is not the design of the test, assignment, or self-evaluation, per se, but the way it is used, and the *decisions* made based on the assessment, that is, to determine whether and to what degree students have learned the material they have been taught. There is usually an evaluation mark given.
- Summative assessments are given at the conclusion of a specific instructional period, and therefore they are generally evaluative, rather than diagnostic, that is, they are more appropriately used to determine learning progress and achievement, evaluate the effectiveness of educational programmes, measure progress towards improvement goals, or make course-placement decisions, among other possible purposes.
- Summative-assessment results are often recorded as scores or grades that are then factored into a student’s permanent academic record, whether they end up as letter grades on a report card or test scores used in the university-admission process.

Summative assessments are typically a major component of the grading process in most courses and programmes.

In recent years, increasingly, a new term has been used based on the research literature on the relationship between assessment and motivation (Harlen and Deakin, 2003). **Assessment as learning** occurs when students reflect on and monitor their progress to inform their future learning goals. It is regularly occurring, formal or informal (e.g., peer feedback buddies, formal **self-assessment**), and helps students to take responsibility for their own past and future learning. It builds metacognition as it involves students in understanding the standards expected of them, in setting and monitoring their own learning goals, and in developing strategies for working towards achieving them.

ACTIVITY 1.1

This activity helps you to explore the meaning of classroom assessment, that is, what you do and what your students do (in peer/groups or individually). Look at [Figure 1.1](#) below and note examples of the three types of assessments that you use in your own teaching. Identify examples of assessment that you use in your own classroom (or have experienced) which help to clarify where the learner is going; where the learner is; and how to get there. Here are two examples elicited from teachers who responded to this activity:

Teacher 1: 'I provide oral feedback to students' first draft of writing to help them to refine their writing' – to provide feedback that moves learners forward.

Teacher 2: 'I have asked my students to work in groups to do research on world music, so each student has something to contribute to the learning, and so assessment *of* the learning is shared among learners on an on-going basis' – to activate students as learning resources for one another.

	Where the learner is going	Where the learner is	How to get there
Teacher	Clarifying, sharing, and understanding learning intentions	Engineering effective discussions, tasks and activities that elicit evidence of learning	Providing feedback that moves learners forward
Peer		Activating students as learning resources for one another	
Learner		Activating students as owners of their own learning	

Figure 1.1 Dimensions of assessment (adapted from Wiliam, 2015)

1.2 Purposes of Assessment

As we can see from [Figure 1.1](#) and from the above activity, assessment has multiple dimensions. Assessment also serves multiple purposes. Three main purposes of assessment are common in our classrooms.

- **Instructional:** This is the most essential purpose of our assessment practice, yet it is often neglected in the assessment literature. Assessment serves to gather information about students' understanding and skills. This purpose will require teachers to use *assessment tools* in collecting data on students' learning, understanding, and skills. Based on where students are in their learning process, as teachers we can plan and adjust our instruction. We engage in this type of assessment all the time in our classroom, for example, every time we ask our students if they have understood a point we

have presented, or when we observe a student working in a group. We need, however, to be more mindful in what assessment we do and how we do it. The *assessment events* and *processes* we conduct signal to our students what is important to learn, and how well their teacher supports their learning.

- *Student-Centred*: Diagnostic assessment refers to the process of identifying individual students' strengths and weaknesses (Alderson, 2005; Fox, Haggerty and Artemeva, 2016). The concept of diagnostic assessment has often been compared to assessments used for placement or remediation purposes. Both assess learning in order to establish which learning opportunities are appropriate for students' learning needs. In the case of **placement testing**, however, the purpose is to group students in a class based on, for example, similar skill development, learning needs, or proficiency levels, so that instruction is suitable for all of the students in the class. Diagnostic assessment, on the other hand, focuses on the individual student's strengths or weaknesses. Increasingly, external diagnostic assessment approaches are being used to identify students at risk of failing university courses in order to provide support for learning. For example, there are a large number of post-admission language assessment approaches (PELA) that connect external diagnostic assessment to learning opportunities (see, for example, Fox, 2009; Fox, Haggerty and Artemeva, 2016; Read, 2016). Within the classroom, teachers may draw on diagnostic assessment techniques to adapt course instruction and better meet their students' learning needs (Black and Wiliam, 1998). Both assessments for placement and for diagnostic purposes involve identifying students' needs in order to support their learning.
- *Administrative*: Grading is the process of summarizing student achievement using a numerical (e.g., 78) or ordinal (e.g., A, B, or C) scale. Grading is a complex evaluative practice that requires teachers to make judgments about student learning. Grades are used, most notably, to make public statements to students, parents and principals about student achievement. Thus, grading is one of the most high-stakes classroom assessment practices, with significant consequences for a student's self-perception, motivation for learning, for the prioritization of **curriculum** expectations, parental expectations and social relationships (Brookhart, 2013). Grading happens at the end of the instruction, and it is usually based on the final product or performance. This purpose is related to *summative assessment* or *assessment of learning* (see Chapter 6 for additional information).

ACTIVITY 1.2

Why do we assess our students? Take a look at the following purposes of assessment in [Table 1.1](#) and see if you can put them into the following three categories: student-centred purposes, instructional purposes and administrative purposes. You can add your own examples to the bullet points below. If you wish to find out how a group of 267 English as a Second Language (ESL)/English as a Foreign Language (EFL) teachers categorized these purposes, you can read Cheng, Rogers and Hu (2004, pp. 367–368) for more information.

1. Student-centred purposes

- Obtain information on my students' progress
- Provide feedback to my students
- ...
- ...
- ...

2. Instructional purposes

- Plan my teaching
- Diagnose strengths and weaknesses
- ...
- ...
- ...

3. Administration purposes

- Provide information to the central administration
- ...
- ...
- ...

Table 1.1 Purposes of assessment and evaluation***Purposes***

- To group my students for instruction purposes in my class
- To obtain information on my students' progress
- To plan my instruction
- To diagnose strengths and weaknesses in my own teaching and instruction
- To provide feedback to my students as they progress through the course
- To motivate my students to learn
- To 'make' my students work harder
- To prepare my students for standardized tests they will need to take in the future (e.g., the Test of English as a Foreign Language)
- To diagnose strengths and weaknesses in my students
- To formally document growth in my students' learning
- To determine the final grades for my students
- To provide information to the central administration (e.g., school, university)
- To provide information to an outside funding agency

Understanding the different purposes for assessment is the most important first step in making informed assessment decisions. The purposes of our assessment influence the frequency and timing of our assessment (*assessment events*), the methods we use (*assessment tools*) and how an assessment is carried out (*assessment processes*). The above assessment purposes are not mutually exclusive, nor are they inherently *good* or *bad*. The key is for us to contemplate assessment purposes, their impact and the many choices we have in conducting our assessments, choosing assessment tools and creating

assessment processes. Only by making a conscious and well-informed effort can we ensure high-quality classroom assessment practice.

1.3 High-Quality Assessment

High-quality assessment practices are those that provide results verifying and promoting targeted student learning. There are a number of fundamental aspects of such high-quality practices.

- **Alignment:** The degree of agreement among curriculum, instruction, standards and assessments (tests). In order to achieve alignment, we need to select appropriate assessment methods, which reflect or represent clear and appropriate learning outcomes or goals.
- **Validity:** The appropriateness of inferences, uses and consequences that result from the assessment. This means that a high-quality assessment process (i.e., the gathering, interpreting and using of the information elicited) is sound, trustworthy and legitimate based on the assessment results.
- **Reliability:** The consistency, stability and dependability of the assessment results are related to reliability. This quality criteria guards against the various errors of our assessments. For example, reliability is the indicator of the number of errors we are making in marking students' work and how consistent our marking is.
- **Fairness:** This is achieved when students are provided with an equal opportunity to demonstrate achievement and assessment yields scores that are comparably valid. This requires transparency, in that all students know the learning targets, criteria for success, and on what and how they will be assessed. Fairness also means that the students are given equal opportunity to learn. Fair assessment avoids student stereotyping and bias in assessment tasks and procedures. Appropriate accommodation is provided to students with special needs.

- **Consequences:** This term is associated with the results of the use or misuse of assessment results. Research into the consequences of large-scale testing tends to focus on the after-effects of test interpretations and use on various stakeholders including value implications and social consequences (Messick, 1989). The term **washback** – the influence of testing on teaching and learning – is now commonly employed in applied linguistics (Cheng, 2014). Some examples of consequences are that students can learn in ways that are consistent with the assessment **task**. Assessment can motivate (and, when it is of low quality, potentially demotivate) students to learn. The student–teacher relationship is influenced by the nature of assessment.
- **Practicality and Efficiency:** Considerations are given to the information that is gathered by assessment. A teacher's life is extremely busy, and this influences the choice of assessment events, tools and processes. Are the resources, effort and time required for the assessment worth the investment? For example, it is not realistic to mark all students' first written drafts. In fact, teachers should not mark the first draft. Rather, teachers need to guide students in self- and **peer-assessment** of their drafts, support students' learning, and encourage them to take responsibility for their own learning, that is, activate students' ownership of their learning as illustrated in [Figure 1.1](#).

ACTIVITY 1.3

1. Reflect on one experience in your life when you felt good about being assessed or taking a test.

Or

Reflect on one experience in your life when you felt bad as a result of being assessed or taking a test.

- What was the experience?
- Which aspects of testing and assessment were related to that experience?
- Was the experience related to test contents, test tasks, test types, test administration, or how the test was marked?

2. Write down the factors that you feel contributed to your positive experience in taking a language test. For example,
 - Clear test directions
 - Familiar test formats
 - ...
 - ...
 - ...
3. Write down the factors that you feel contributed to your negative experiences. For example,
 - Noisy testing environment
 - Too little time
 - ...
 - ...
 - ...
4. How do these positive or negative factors contribute to the overall quality of assessment?
5. Read the following two excerpts from test-takers about their test-taking experiences from Cheng and DeLuca (2011), and see how these experiences enhance or decrease the overall quality of assessment.

I couldn't hear the tape clearly so I immediately told the invigilator and was sent to the special room [an alternate listening room]. Already anxious about missing some listening materials, I was assigned to sit at the back of the room in front of windows that faced a basketball court. The door to my right was open and I could hear the noise of students playing downstairs and of those walking past. A portable stereo player was placed on a chair at the front of the room. The volume was really low and I found it extremely difficult to follow the content. (Cheng and DeLuca, 2011, p. 110)

Excerpt 1 Negative test-taking experience

Usually when taking listening tests, the time limit is so tight it makes me nervous. Some tasks or questions appear after a long audio text such that test-takers have to memorize what they have just heard; the tests require test-takers to have good memory skills. In contrast, this listening test was chunked into 4 parts, and the first three parts were divided into further sub-parts. Before and after each part and sub-part, sufficient time was given to test-takers to familiarize themselves with the questions, or check their answers. These practices allow test-takers to feel at ease and enabled assessment of test-takers' listening ability – which should be the major criterion. (Cheng and DeLuca, 2011, p. 111)

Excerpt 2 Positive test-taking experience

The first excerpt illustrates how test environmental factors contributed *negatively* to a test-taker's perceptions of assessment fairness. The second excerpt, on the other hand, describes a testing experience where time was more effectively managed so that this test-taker could complete tasks and respond to the test tasks in ways that were more consistent with authentic listening in non-test contexts. Furthermore, because the tasks were 'chunked' into well-timed parts and sub-parts, and breaks were allowed throughout the test, anxiety levels were reduced. As a result, this test-taker may well have performed at a higher level, which was more representative of his or her actual listening proficiency.

Understanding such testing and assessment experiences is an essential area for continued validation inquiry. Such inquiry can help teachers to learn experientially and reflect critically on core aspects of assessment, such as *construct representation*, that is, how well a test measures what it is designed to measure. Tapping into testing experiences gives us a window on test structure and content, scoring, administration/conditions and timing. All of these features of testing are

important in our consideration of how well the test is measuring the intended construct.

Investigating testing experiences also helps us to better understand sources of *construct-irrelevant variance*, that is, what interferes with the measurement of the construct. We need to ask, are we measuring something that we did not intend to measure? For example, if we are measuring listening comprehension on a **proficiency test**, and students at the front of the room can hear more clearly than students at the back of the room, then where students sit may negatively or positively influence their scores on the test. In this case, sound quality and seating position contribute to construct-irrelevant variance. We need to reflect on and address the many sources of construct-irrelevant variance that may interfere with good measurement, such as test coaching/preparation, emotions/self-efficacy and the misuse of test scores, in evaluating the usefulness of a test. In your own experience of taking tests, you may remember a time when you did not do well because of something that undermined your performance. Conversely, there may have been a time when you did better than expected, because of some advantage that was not related to the skill being measured. Eliciting testing experiences from stakeholders (e.g., students, parents and teachers) helps us to clarify the impact or social consequences of a test by considering their accounts of how valid and/or fair they consider the test. It helps teachers to better understand the interconnectedness of testing constructs, processes, uses and impacts on learning. We will revisit this point in [Chapter 7](#) of this book.

1.4 Beliefs about Assessment

What we assess and how we assess it depends to a large extent on our interconnected – typically implicit – system of knowledge, values and beliefs that inform our actions in the classroom. These constitute our *teaching philosophy*. As teachers, it is important for

us to be aware of our own philosophy of assessment, to explore the philosophies of others, and to acknowledge the overlaps and differences between the two. Reflecting on our own teaching philosophy is a useful first step towards building common ground on assessment in our classrooms, programmes, schools, universities and other language learning contexts. However, you may not have stopped to consider your own beliefs, assumptions and knowledge about assessment as you engage in the day-to-day teaching of your class. Teaching demands your full attention, and there is rarely time to step back and consider fundamental questions about what we are doing or why.

It is not that we do not spend a great deal of time planning, organizing and implementing our lessons, and a large part of this involves ongoing assessment such as marking our students' papers, maintaining records of student performance, providing feedback to students on work-in-progress, and developing quizzes, essay assignments, or projects. It is likely, however, that you may not have had the time to reflect on *why* you decide to assess one way instead of another, nor is it likely that you have had the time to discuss your **assessment plans** and experiences with other teachers. One unfortunate characteristic of most language teaching situations is that time for talking with colleagues is often limited either to meetings that are primarily focused on administrative concerns or to brief discussions during breaks or over lunch.

Even in educational settings where much or all of the assessment that counts is *external* to the teacher and classroom, for example, in contexts with **high-stakes** national examinations, proficiency tests, or professional certification tests, individual teachers will prepare their students in distinctively different ways. It may be that *all* of the students in a programme, school, or university are taking the *same* test, but research suggests that each teacher's approach to test preparation and support is unique (Fox and Cheng, 2007; Wang and Cheng, 2009). Over the years, much has been written about teacher decision-making in the classroom (e.g., Cheng, 1999; Woods, 1996) and although external factors such as tests, textbooks, students' proficiency, class size and workload may to

some extent influence a teacher's choice of assessment activities, what a teacher *believes* about language teaching and learning – based on attitudes, understanding, knowledge and experience – may be far more important in the choice of specific assessment activities.

Teachers' individual philosophies may align with or diverge from the prevailing educational or curricular expectations that shape the contexts within which they teach and assess (see Cortazzi and Jin, 1997; Gorsuch, 2000). As mentioned earlier in the chapter, alignment is the ideal in ensuring high-quality assessment. At times, however, there may be more divergence than alignment. Commenting in 2001 on the role of externally imposed tests and standards on classroom teachers in Australia, Davidson noted the 'tension between the philosophical base of the ESL field which emphasizes diversity and complexity, and the demands of the "mainstream" educational agenda for commonality, simplicity, and homogeneity' (Davidson, 2001, p. 30). When there is misalignment or divergence, teachers often adapt their assessment approaches so that they are consistent with their philosophies of teaching and learning. In the rest of this chapter, we will consider four prominent educational philosophies that have been influential in the history of language teaching. Then, we will help you to probe your own philosophy of teaching and learning. At the end of the chapter, we will ask you to consider how your philosophy of teaching and learning languages relates to the decisions you make about assessment.

1.4.1 Prominent Educational Philosophies in Language Teaching

White (1988) identified three different educational philosophies, which have characterized the mainstream educational agenda during different periods of language teaching history. He labelled these philosophies: *classical humanism*, *progressivism* and *reconstructionism*. White discussed the implicit beliefs, values and assumptions that are evident in these three different philosophies of language teaching curriculum. His labels are useful for our discussion of your own philosophy of assessment.

Using White's labels, let's briefly define each of the educational philosophies that inform language teaching curriculum. We will then look at how these curricular philosophies influence the way in which we undertake assessment in our classrooms.

- *Classical Humanism*

A classical humanist curriculum values tradition, culture, history, literature and knowledge about a language. The main curricular goal is to pass on to the most capable students the knowledge, values, literature and history of a language. Sometimes the emphasis on knowledge has meant that living languages – such as English, Spanish and French – have been taught as if they were dead ones, such as Latin. Students typically translate texts, memorize vocabulary, learn about the grammar of the language and practise grammatical exercises. Literature is of primary importance. Texts (usually short passages) are chosen because they have particular literary or historical significance. They may be memorized and recited, copied, analysed, or translated. Performance expectations are fixed, and typically, excellence is defined in terms of accuracy.

- *Progressivism*

A progressive curriculum focuses on the individual learner and defines language teaching activity as learner-centred or learner-led. Although informed by overall purpose and goals for language learning, progressive curriculums typically require teachers to define course activity and content in relation to **needs analysis** or other evidence of student interests, preferences, or performances, which set the direction, emphasis and activity of the course. The teacher's attention is focused on the ongoing processes of language learning and development in relation to the predefined purpose and/or goals of the course. Because each learner is unique, varying in background, aptitude, interest and motivation, learning is also unique and individual. Teachers may negotiate learners' individual work plans by probing learners' views on what they need or want from their language learning experience; and through examination of gaps between a student's current language capability and the goals of the student, teacher and course. Allwright (e.g., 1982) is often referred to as a pioneer in discussions of students' perceived *needs* and *wants* in relation to teachers' assessments of *lacks* – the difference between a current level and a desired level of language competence, proficiency, or performance. The language teacher's role, according to Allwright, and in keeping with a

progressive curriculum, is to: (1) help students to identify with increasing clarity their language skill areas in need of improvement, and (2) elaborate strategies, which would help students to attain the desired increase in their achievement of those skills. Ongoing activity is structured in relation to the process and progress in the course. What happens in a class depends directly on the teacher's day-to-day assessment of what an individual or a group of learners need to move closer to the purpose or goals that have been negotiated.

- *Reconstructionism*

A reconstructionist curriculum identifies specific learning outcomes based on overall course purposes and goals. These learning outcomes guide the selection of activity and experiences in the course, all of which are designed in relation to the predefined criteria, which identify what students should know and be able to do at the end of the course. Teachers may work with specific materials, textbooks, or tasks as they support their students in achieving the predefined course goals, which are often referred to as intended learning outcomes (e.g., Biggs and Tang, 2011). Some reconstructionist curriculums are mastery-based (pass or fail), others are related to successfully demonstrating a degree of achievement and attainment in relation to benchmark criteria (e.g., the **Common European Framework** of Reference; the **Canadian Language Benchmarks**). Criteria are often related to tasks with different levels of difficulty. Where tasks are not defined, teachers develop their own tasks in relation to the benchmark criteria, taking into account the students they are working with. In [Chapter 2](#) we take a closer look at outcomes-based or reconstructionist curriculum in language teaching.

To White's three curricular philosophies, we will add one more:

- *Post-Modernism or Eclecticism*

This educational philosophy values the spontaneous, unique, unplanned and emergent learning that takes place each day in a classroom setting. Because we are all unique individuals, every classroom, every interaction between a student and a teacher, the experiences of a language activity, and what may be learned from it are also unique, individual and unpredictable. Developmental and individual learning is valued; outcomes will differ because learners differ. Typically learners are divided into groups by proficiency levels, interests, ages and so on. The teacher supports their

development on an ongoing basis by identifying whatever activity seems to provide the next best step in supporting the individual student's learning processes.

The four curricular philosophies defined above have very different expectations for teachers, students, content, context and assessment. According to Connelly and Clandinin (1988) we can analyse any language curriculum by examining the role of the content, the teacher, the student and the context or milieu. To this list of 'commonplaces of curriculum' (pp. 84–85), we will add the role of *assessment*. We will now examine the implicit roles which each of the four curricular philosophies suggests in the following activity.

ACTIVITY 1.4

Each of the four language curriculum philosophies defined in this chapter is listed in [Table 1.2](#). Working alone, with a partner, or in a small group complete the table with what you understand to be the role of teacher, student, content, context and assessment. An example is provided for you in [Table 1.2](#).

After identifying the role of assessment, list one or two assessment practices that you think would be used by teachers who subscribe to this teaching philosophy. If you fill in the table on your own, discuss it with a partner or a group when you have finished.

Also answer these questions below if there is time.

- Have you ever taught a class in which the curricular philosophy was similar to one of the four we have defined in this chapter?
- Have you ever studied/learned a language in a class which was similar to one of the four educational philosophies?

Table 1.2 The commonplaces of language curriculum

Educational philosophy of curriculum	Role of the teacher	Role of the student	Role of content	Role of context	Role of assessment: Assessment practices	My experience
Classical humanism	To provide a model; to teach reading of valued texts; to monitor careful replication of texts and their meanings. Focus is on reading and writing.	To reproduce, memorize, or translate valued texts.	Students learn about a language (often its history and culture) through exposure to highly valued texts. They do not learn how to use the language.	The texts provide the context.	Focus on correct form, accuracy: e.g., grammar tests, translation, dictation, recitation, vocabulary tests.	
Progressivism						
Reconstructionism						
Post-Modernism						

1.4.2 What is Your Own Philosophy of Assessment?

Now that we have had a chance to discuss different curricular philosophies, consider your personal philosophy of teaching and learning and its relationship to what you assess and how you assess it in your own language classroom. It is useful to reflect on our own values and beliefs about teaching and learning languages and to consider how these influence assessment decisions that we make in our classrooms. In examining our teaching philosophy it is important to keep in mind Prabhu's (1990) advice that there is no one best method for language teaching. This is equally true of assessment.

ACTIVITY 1.5

You can use your responses to the statements in the questionnaire in [Table 1.3](#) to reflect on your teaching philosophy and consider how it influences decisions you make about assessment in your classroom.

Respond to each of the following statements by circling the number on the right-hand side of the table which best reflects your agreement (or disagreement). This questionnaire uses what is referred to as a Likert scale (named after the man who invented it). If you circle zero, you indicate that you completely disagree with the statement; if you circle five, you indicate that you completely agree. In many cases you may be somewhere in the middle – at times, leaning in the direction of disagree; at other times, leaning in the direction of agree. Circle the number that best reflects your response to each of the statements.

Please respond to all of the statements! After you have finished, you may want to compare your responses with one of your peers or colleagues and discuss where your responses were similar and where they differed.

Table 1.3 Questionnaire: What is most important in teaching a language?

ITEM	STATEMENTS	DISAGREE -----AGREE					
1.	When I teach a language, my primary focus is on the rules of grammar.	0	1	2	3	4	5
2.	Memorizing vocabulary is essential in learning a new language.	0	1	2	3	4	5
3.	First and foremost, the needs of my students determine what I teach.	0	1	2	3	4	5
4.	Day-to-day interaction with my students guides what we will learn next, and how we will learn it.	0	1	2	3	4	5
5.	Understanding of the literature, culture, and history of a language is the most important reason for teaching a language.	0	1	2	3	4	5
6.	I plan all of my lessons in advance in relation to predetermined goals for learning.	0	1	2	3	4	5
7.	It is important to teach what the syllabus or textbook tells me to teach, to ensure the aims of the course are systematically met.	0	1	2	3	4	5
8.	Before teaching a new course, I start by defining what my students should know and be able to do by the end of the course.	0	1	2	3	4	5
9.	I prefer to negotiate the content we will cover during a course directly with my students.	0	1	2	3	4	5
10.	The quality of learning increases when it is meaningful to a learner.	0	1	2	3	4	5
11.	Students learn the most when they are actively engaged.	0	1	2	3	4	5

- | | | | | | | | |
|-----|---|---|---|---|---|---|---|
| 12. | It is impossible to predict what an individual student will learn, but I will help each student to learn as much as possible. | 0 | 1 | 2 | 3 | 4 | 5 |
| 13. | My overall plan is to let language learning just happen freely and naturally in my classroom. | 0 | 1 | 2 | 3 | 4 | 5 |
| 14. | How and what I teach depends directly on what seems to work with the greatest number of students. | 0 | 1 | 2 | 3 | 4 | 5 |
| 15. | I always begin a course by identifying my students' needs and interests. | 0 | 1 | 2 | 3 | 4 | 5 |
| 16. | The ongoing <i>process</i> of learning is most important – not the content we cover, the products, or the outcomes. | 0 | 1 | 2 | 3 | 4 | 5 |
| 17. | I'll try any approach that I think will work to support my students' learning. | 0 | 1 | 2 | 3 | 4 | 5 |
| 18. | It is important to correct any and all student mistakes. | 0 | 1 | 2 | 3 | 4 | 5 |
| 19. | Unplanned and spontaneous student interaction promotes the most meaningful language learning. | 0 | 1 | 2 | 3 | 4 | 5 |
| 20. | My main focus is on the long-term goals of my course. | 0 | 1 | 2 | 3 | 4 | 5 |
| 21. | <i>Who</i> I am working with determines what and how I teach a language. | 0 | 1 | 2 | 3 | 4 | 5 |
| 22. | I need to constantly reflect on my teaching with a view to modifying my goals in relation to my students' day-to-day development. | 0 | 1 | 2 | 3 | 4 | 5 |
| 23. | I do not expect all of my students to achieve the same outcomes, because every student is different. | 0 | 1 | 2 | 3 | 4 | 5 |
| 24. | The best and brightest students in my class deserve the most attention. | 0 | 1 | 2 | 3 | 4 | 5 |

- | | | | | | | | |
|-----|--|---|---|---|---|---|---|
| 25. | Who is in my class is not as important as what I am teaching. | 0 | 1 | 2 | 3 | 4 | 5 |
| 26. | I continuously evaluate my students' learning in relation to the learning outcomes defined for my course. | 0 | 1 | 2 | 3 | 4 | 5 |
| 27. | Students only value what is graded and marked. | 0 | 1 | 2 | 3 | 4 | 5 |
| 28. | Students' achievement increases when classrooms are highly competitive. | 0 | 1 | 2 | 3 | 4 | 5 |
| 29. | All language in the class must be monitored and corrected by the teacher, because if students are exposed to errors, they will learn and reproduce them. | 0 | 1 | 2 | 3 | 4 | 5 |
| 30. | Teachers must plan and then follow their plans in teaching a language class, because digressions waste everyone's time. | 0 | 1 | 2 | 3 | 4 | 5 |
| 31. | Incidental or unexpected learning is not as important as predetermined course outcomes. | 0 | 1 | 2 | 3 | 4 | 5 |
| 32. | Spontaneous and free-flowing interaction is essential in learning to use a new language. | 0 | 1 | 2 | 3 | 4 | 5 |
| 33. | I don't know exactly what I'm going to teach until I'm actually teaching it. | 0 | 1 | 2 | 3 | 4 | 5 |
| 34. | Every language task, activity, or experience should be undertaken with a clear purpose or goal in mind. | 0 | 1 | 2 | 3 | 4 | 5 |
| 35. | I define the learning outcomes of my course first, and then I design the tasks and activities that will help my students achieve them. | 0 | 1 | 2 | 3 | 4 | 5 |
| 36. | Students learn the most when something is personally interesting or useful. | 0 | 1 | 2 | 3 | 4 | 5 |

1.4.3 Your Own Assessment Profile

You may have found some of the statements in the questionnaire above were difficult to answer. You may also have been surprised to find, if you discussed your responses with others, that some statements prompted considerable disagreement. The areas of disagreement are the most informative because they suggest deep values, beliefs and assumptions about learning and teaching, which are unique to individuals. These differences are key when we are working with our colleagues within a programme. It is important to acknowledge and discuss all of our areas of agreement, but it is equally important to air our differences and work together towards the co-creation of common ground. This is particularly important when we are deciding how best to assess multiple classes in a programme, to evaluate programme quality and monitor the programme's success in meeting its aims and purposes.

ACTIVITY 1.6

We can now relate your answers to the questionnaire in [Table 1.3](#) to our discussion of curricular and personal teaching philosophies. Each of the questionnaire items reflects one of the four curricular philosophies introduced above, namely, classical humanism, progressivism, reconstructionism, or postmodernism/eclecticism.

Directions: Write each number you circled on the Likert scale in the questionnaire above ([Table 1.3](#)) next to the questionnaire items listed in [Table 1.4](#) below. When you have finished entering the numbers in [Table 1.4](#), add up each column in order to create your own teaching, learning and assessment profile. The totals provide you with an overall profile of your philosophy of teaching and learning across the four curricular philosophies.

- Which curricular philosophy is dominant in your profile? Which is least relevant to you?

- How does your personal profile relate to the curricular philosophy of the context in which you are teaching (or plan to teach)? Is your personal profile aligned with the prevailing philosophy of education in your context of teaching?
- Is there a gap between your personal philosophy and that of your programme?
- How does your profile compare with those of your peers or colleagues?

Table 1.4 My philosophy of teaching and learning

Classical Humanism	Progressivism	Reconstructionism	Post-Modernism or Eclecticism
Item 1:	Item 3:	Item 6:	Item 4:
Item 2:	Item 9:	Item 7:	Item 13:
Item 5:	Item 10:	Item 8:	Item 14:
Item 18:	Item 11:	Item 20:	Item 16:
Item 24:	Item 12:	Item 26:	Item 17:
Item 25:	Item 15:	Item 30:	Item 19:
Item 27:	Item 21:	Item 31:	Item 32:
Item 28:	Item 22:	Item 34:	Item 33:
Item 29:	Item 23:	Item 35:	Item 36:
Total:	Total:	Total:	Total:

1.5 Looking Back at Chapter 1

At the beginning of this chapter, you read about assessment. Next you considered beliefs, values and assumptions about teaching and learning languages, which, along with your

knowledge, come together to form your general philosophy. This is how you can relate your philosophy of teaching and learning to assessment.

Now we are ready to focus on assessment – your philosophy of assessment.

- Given your profile, which assessment practices would best align with your personal philosophy of teaching and learning languages?
- Which assessment practices would seem not to?
- Which types of assessment dominate the context in which you are teaching (or planning to teach)?
- Can you identify ways in which you might adapt or work with a dominant assessment strategy so that it supports your philosophy of teaching and learning?

You may want to read more about the issues discussed in this chapter. The following are suggested resources for further reading and reflection.

Suggested Readings

Cheng, L., Rogers, T. & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods and procedures. *Language Testing*, 21(3), 360–89.

This is one of the first studies published in language testing on teacher assessment literacy. This comparative survey study examined a range of teaching contexts represented by Canadian ESL, Hong Kong ESL/EFL, and Chinese EFL in which 267 instructors participated. This study documented the purposes, methods and procedures of assessment in these three contexts. The findings provide insights into the nature of assessment practices in relation to ESL/EFL classroom teaching and learning at the tertiary level.

Clandinin, D. J. & Connelly, F. M. (1992). Teacher as curriculum maker. In P. Jackson (ed.), *Handbook of Curriculum Research* (pp. 363–41). New York: Macmillan.

Clandinin and Connelly's seminal discussion of the centrality of the teacher in curriculum making (planning, implementation and evaluation). In their long-term collaboration, Clandinin and Connelly have published extensively on *narrative inquiry*. This research approach gathers teachers' personal narratives of their classroom experience in order to increase our understanding of teaching and learning in practice.

Davidson, C. (2001). Current policies, programs and practice in school ESL. In B. Mohan, C. Leung & C. Davison (eds.), *English as a second language in the mainstream: Teaching, learning and identity* (pp. 30–50). London: Longman.

Davidson examines the tensions that exist between external policies and language teachers' perspectives on teaching and learning at the classroom level. This chapter demonstrates the important interactions that take place when policies, programmes and practices are being discussed by language teachers who have very different beliefs about teaching.

Fox, J. (2009). Moderating top-down policy impact and supporting EAP curricular renewal: Exploring the potential of diagnostic assessment. *Journal of English for Academic Purposes*, 8(1), 26–42.

Fox provides an in-depth view of the role that a diagnostic assessment approach played in addressing issues arising from a policy decision with negative impact on language teaching in an English for Academic Purposes (EAP) programme. Her research investigates the differing responses of a group of EAP teachers to diagnostic information provided to them through assessment. These differences are evidence of teachers' varying philosophies of teaching.

Activate your learning

- What should we assess?
- What is worth teaching and assessing?
- How can assessment be integrated into the learning tasks?
- How can we align assessment in our classrooms to benchmarks and standards, curriculum, external tests and textbooks?

2.1 What is Worth Teaching and Assessing?

This chapter addresses some of the most challenging aspects in teaching: deciding what to teach, what to assess and how to align assessment in the classroom to the learning goals and outcomes for our students. Such goals and outcomes may be *explicitly* defined for a whole programme by benchmarks and standards, curriculum and external tests, or *implicitly* defined by teachers through textbooks at the classroom level. Teachers may also define outcomes by eliciting information from their students through needs analysis, student–student and student–teacher interaction, student self-assessment, and so forth.

Look at [Figure 2.1](#) and think about what it means to us as teachers to achieve instructional goals or outcomes through teaching and assessment. In the centre of this triangle is our students' learning. The first question we need to ask relates to the learning goals or outcomes we have for our students: *What do I want my students to learn? What do I want my students to be able to do?* Moving clockwise in [Figure 2.1](#), we need to ask how we will monitor and evaluate learning, or what information is

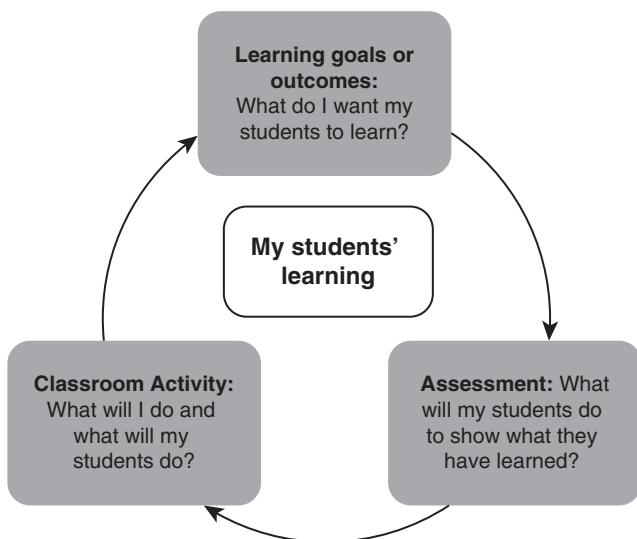


Figure 2.1 Alignment of learning goals, assessment and classroom activity

essential in order to determine whether my students have met or exceeded the required expectations: *What will my students do to show that they have learned? What will I do as a teacher and what will my students do as learners?* Given the evidence that we plan to collect during our course, we then need to identify the actual classroom activities that will support our students' learning and development.

Assessment serves as the key to check on learning and provide essential information to teachers. This process is cyclical, as instruction is a process of providing support to learners through assessment and teaching. In this sense, teaching and assessment are an integral process in learning. Teachers need to constantly ask themselves: *Have my students learned? How well have they progressed through assessment practices?* The assessment component of [Figure 2.1](#) illustrates the *why*, *what* and *how* of your assessment.

It is important to emphasize that the starting point in teaching a course and assessing learner progress (as [Figure 2.1](#) illustrates) is to clearly define our goals or learning outcomes. Our goals need to be shared and discussed with our students. We know that when our students understand the goals or intended learning outcomes of a course, and are aware of the evidence we plan to collect in order to support their learning and evaluate their progress, our students are more likely to meet or exceed our expectations (Biggs and Tang, 2011; Wiggins and McTighe, 2005).

Let us take a detailed look at a context where learning goals are defined by a curriculum. If you are teaching language in a public school, your goals, learning outcomes and decisions about activities may be shaped by curricular guidelines. For example, read the following excerpt taken from the English as a Second Language (ESL) curriculum for Level 3 (ESLCO), which guides English language teaching in the public schools in Ontario Canada:

This course further extends students' skills in listening, speaking, reading, and writing in English for a variety of everyday and academic purposes. Students will make short classroom oral presentations; read a variety of adapted and original texts in English; and write using a variety of text forms. As well, students will expand their academic vocabulary and their study skills to facilitate their transition to the mainstream school program. This course also introduces students to the rights and responsibilities inherent in Canadian citizenship, and to a variety of current Canadian issues. (Ontario Ministry of Education, 2007, p. 83)

A curriculum like the ESLCO provides a considerable amount of guidance as to what you can do as a teacher and what your students can do as learners at a particular level of ESL, but these guidelines do not specifically define your assessment activities by stating what your students should *do*, to show what they have learned. In other words, what links the learning goals or outcomes and assessment is the teaching process. [Figure 2.2](#) is a

page from the ESLCO curriculum, which provides specific details on overall expectations in listening and speaking:

ESL Level 3, Open

LISTENING AND SPEAKING

OVERALL EXPECTATIONS

By the end of this course, students will:

1. demonstrate the ability to understand, interpret, and evaluate spoken English for a variety of purposes;
2. use speaking skills and strategies to communicate in English for a variety of classroom and social purposes;
3. use correctly the language structures appropriate for this level to communicate orally in English.

SPECIFIC EXPECTATIONS

1. Developing Listening Comprehension

By the end of this course, students will:

Listening for Specific Information

1.1 demonstrate comprehension of specific information in more detailed directions, instructions, and classroom presentations, with reduced contextual and visual support (e.g., construct or draw a model of an item based on a partner's oral instructions; identify main ideas from news broadcasts; extract key concepts from audio webcasts and library dial-a-story services; take point-form notes on main ideas from classroom oral presentations using an outline or graphic organizer)

Teacher prompt: "How did the use of a graphic organizer help you to listen for and understand information from the presentation?"

Listening to Interact

1.2 demonstrate understanding of spoken English on familiar and content-area topics in a variety of interactive situations (e.g., conduct a survey of classmates about reading preferences in English and in their first language; participate in a small-group place-mat activity to reach agreement about the accomplishments of Alexander Graham Bell; show understanding during discussions in a literature circle by contributing relevant questions)

2. Developing Fluency in Speaking

By the end of this course, students will:

Speaking to Interact

2.1 engage in spoken interactions on personal and content-area topics (e.g., contribute information in a jigsaw group discussion on current events; share ideas in a literature circle; give feedback to a classmate in a peer-assessment activity)

Teacher prompt: "When you are going to participate in a discussion, what kind of preparation do you find most helpful?"

Using Conversational Strategies

2.2 use a number of conversational expressions to negotiate spoken interactions (e.g., take turns speaking by using expressions such as "What do you think about that?", "What's your opinion?", "It's _____'s turn now", "I'd like to add..."; indicate understanding and sympathy with expressions such as "Oh no!", "That's too bad", "I'm sorry to hear that"; ask for clarification with expressions such as "I'm not sure I understand", "Would you please repeat that?", "Pardon?")

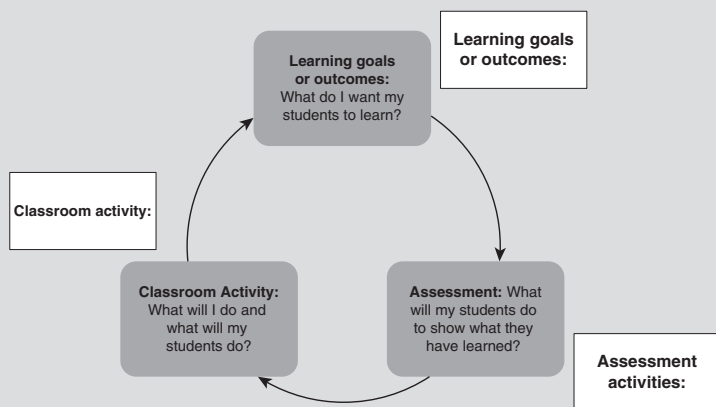
Speaking for Academic Purposes

2.3 present ideas and information orally for academic purposes in supported situations (e.g., make short oral presentations on familiar topics using appropriate elements of a classroom presentation format such as an introduction, question-and-answer exchange, and conclusion; explain the points of view of different characters in a novel using a graphic organizer as a guide)

Figure 2.2 Specific curriculum expectations in listening and speaking

ACTIVITY 2.1

Examine [Figure 2.2](#) and think of one learning goal, one strategy for assessing your students' progress or attainment of that goal (or learning outcome), and one classroom activity that would support your students' development of competencies related to the goal or learning outcome. Write your ideas in the boxes provided or on a separate piece of paper. When you have finished, if possible, discuss your ideas with your colleagues.



2.2 How Do We Align Our Assessment and Classroom Activity with Our Learning Goals and Outcomes?

In [Chapters 1 and 2](#), we refer to *alignment* as a critical requirement for high-quality assessment. In [Chapter 1](#), we briefly defined alignment as, ‘the degree of agreement among curriculum, instruction, standards and tests’. We noted that agreement is achieved when our careful and systematic selection of assessment methods, reflects or represents clear and appropriate learning outcomes or goals. Also in [Chapter 1](#), we explored your own individual teaching and assessment philosophy. We

examined the relationship between our beliefs, knowledge and assumptions about language teaching and learning, and our choices of or preferences for assessment methods in our classrooms. These are all important aspects when we align our assessment and classroom activity with our learning outcomes.

2.2.1 Defining Learning

Reflect for a moment on your own philosophy of teaching and assessment as discussed in [Chapter 1](#). What is learning to you? In *The Ontario Curriculum Grades 9 to 12 English as a Second Language and English Literacy Development* (Ontario Ministry of Education, 2007), learning is defined as the acquisition of knowledge, skills, attitudes, values and experiences. The curriculum definition focuses on the acquiring of *defined* knowledge, skills, attitudes, values and experiences. Do you agree?

An alternative, more cognitive definition of learning would describe learning as a process of formulating new and more complex understandings of the world, and as a process of revising and refining mental constructs, that is, the understandings that guide how we think, speak and behave. This difference in definitions will determine the nature of our assessment events, tools, processes and decisions.

Consider the three examples below. Question 1 is an assessment of a fact. There is a right and wrong answer, which your students either know or do not know. If they do not know, they can find the correct answer relatively easily online. The learning assessed by this type of question is memory-based and limited to one item or unit of information. Question 2 requires a different process of learning. Students can deal with $2 + 2$ or $2 - 1$ as a first step and arrive at the same answer in the end. The learning process is a bit more complex than the first question. Now look at question 3, the learning outcomes are varied and the learning process far more complex. Although there is one evident answer (i.e., the British Broadcasting Corporation) there is also the need to provide supporting evidence that illustrates its importance in British history.

1. What is the capital of the United Kingdom (UK)?
2. What does $2 + 2 - 1$ equal?
3. What is the importance of the BBC in British media history?

Our teaching and assessment should reflect such multiple and complex processes.

2.2.2 Defining Learning Outcomes

Learning outcomes state what our students should know and be able to do at the end of a course, as a result of all of our activity – lessons, assignments, feedback and tests. Learning outcomes are explicit statements of expectations (or criteria) that describe the skills, knowledge, attitudes and capabilities that our students should achieve as a result of our work with them during a course. They may also be referred to as learning targets, which require teachers to specify: (1) what a student should know and/or do as a result of instruction; and (2) the criteria for evaluating mastery or achievement of knowledge and performance (McMillan, 2014).

Learning outcomes set the agenda for a course and hold it together. They give us a clear sense of what our course is aiming for, of what we expect our students to achieve as a result of our teaching activity during a course.

- As teachers, writing down and reflecting on 4–6 learning outcomes for a course, this process makes explicit the overall expectations that we have of our students and helps us to develop a better sense of our course as a whole.
- Sharing our overall expectations with students will help them to better understand the course, and to develop a clearer sense of the standards they will be working to meet or exceed.
- Learning outcomes provide a measure against which we can evaluate our students' progress, development and achievement in a course.
- Explicit learning outcomes help us to evaluate the alignment of our content, activities and assessment – to ensure that our

course plans and goals for our students' learning are in sync and coherent.

It is important to remember in our discussion of learning outcomes that these are statements of what we *intend* for our students to learn, know and be able to do, as a result of a course. It does not mean that other important learning, which is unanticipated, will not be valued or valuable – what some educators (e.g., Biggs and Tang, 2011) have referred to as *unintended* course outcomes.

As teachers we know that unanticipated learning takes place on an ongoing basis, it is often unique to an individual, and may differ from one class to the next. Our intended learning outcomes, however, define the overall goals for all of our students in a class and our assessment will be based on the degree to which our students meet (or exceed) those expectations. At the end of a course, our grades will reflect how well each of our students met those expectations (see [Chapter 7](#)). The evidence we collect that supports our grades comes directly from the assessment tasks and activities we designed or drew on during our course.

Keep in mind, intended learning outcomes are written from the student's perspective. Learning outcomes tend to follow a formula which:

connects a phrase such as...

At the end of this course, students will/should be able to ...

By the end of the course, students must show they can ...

In order to pass this course, students must demonstrate that they can ...

to an action verb such as...

*explain,
identify,
analyse,
organize,
evaluate,
produce,
distinguish*

within a certain content domain.

Here are a few examples of learning outcomes taken from an advanced course in writing:

- By the end of the course, students must show they can apply the in-class defined criteria to the evaluation of a televised news report.
- At the end of this course, students will be able to write a short news story about a current event.
- In order to pass this course, students must demonstrate that they can identify the difference between factual information and opinion in a story about current events.

2.2.3 Connecting Learning Outcomes to Tasks

Learning outcomes provide overall goal statements that frame activities in a course. They are the starting points for activities, but we also need to understand what specific learning is involved in a task and what an assessment task is actually measuring. For example, is our task related to one (or more) of the following aspects of learning? If yes, which aspect of the learning process is involved?

1. Knowledge and simple understanding: declarative (know what) and procedural (know how)
2. Deep understanding and reasoning
3. Skill (development, proficiency)
4. Product (performance)
5. Affect (personal development, for example, self-efficacy, self-assessment, goal setting)

We will provide a sample task analysis to demonstrate how we align learning outcomes with tasks through task analysis. We have used a simple example of baking a cake. Here is what a teacher can do in analysing the learning outcome of baking a cake by dissecting the learning process and identifying the micro-level activities that are required in order to meet or exceed the course requirement or outcome of baking a cake. The activity is performance-based, and provides an example of how you can conduct a task analysis of your own of a required performance.

The first step is to define the learning outcome or outcomes. The sample learning outcome (below) might have been written in a number of ways, depending upon the teacher's emphasis, but here is an example, drawing on the formula provided above for writing learning outcomes:

By the end of the course, students will be able to independently, safely, and successfully bake a cake with a minimum of guidance.

The second step is to undertake a task analysis. Begin by listing what students must know and/or be able to do in order to achieve the learning outcome. In listing each micro-level activity that is required, we also identify which aspects of the learning process are involved:

- How to set and use an oven (procedural knowledge)
- The meaning of terms such as stir, whip, beat, blend and so on (declarative knowledge)
- How to prepare the baking pans (procedural knowledge)
- How to operate a mixer, food processor, or other kitchen equipment safely and correctly (procedural knowledge and skill)
- What ingredients can be omitted or substituted (reasoning-analysis)
- How to read (skill)
- How to follow directions (skill)
- How to troubleshoot and correct errors in following the directions (reasoning)
- How to assemble the parts of the cake (skill)

The third step is to define higher levels of performance where a degree of achievement can be reached by different students to a greater or lesser extent and at different times:

- Envision what the final product will look or taste like by reading the ingredients and directions (deep understanding and reasoning)
- Exercise judgment in selecting a recipe which is do-able, given constraints of time and limited ingredients (deep understanding and reasoning)
- Enjoy baking or cooking (or eating) in order to want to bake a cake (affect)

- Appreciate the aesthetic qualities of a cake such as appearance, aroma, taste, texture (affect)
- Evaluate and explain what separates an exceptional cake from a more ordinary one (deep understanding and reasoning)

You can use the following table as a self-assessment tool or a template to initiate your own task analysis.

Learning Outcome: In order to pass this course, students must be able to analyse evidence (such as diaries, leaflets, letters, artefacts and photographs) to determine the causes of an historical or contemporary event.

Initial task analysis

Selected time period

Knowledge	Thinking Skills	Possible Concepts Required
Background information about a given time in history • Key figures • Economic conditions • Political ideologies • Social issues	• How to select relevant evidence from a given time period • How to extract a theme or issue from among a set of documents • How to look for common claims across documents • How to look for diverse perspectives on an issue • How to use knowledge of the source to interpret the trustworthiness of the information • How to examine artefacts (e.g., tools used for print, writing and other forms of written communication) for how they might affect human events and communication • How to link information from evidence to historical events	• Cause–effect • Supply–demand • Social mores • Imperialism • Colonialism • Representative government • Federalism • States’ rights • Democracy

2.2.4 Taking Stock: Learning Outcomes, Backward Design and Assessment Tasks

As the above examples illustrate, once we have spelled out the learning outcomes for our courses, we can design ongoing classroom assessment activities that reflect and represent them. The learning outcomes identify what we want our students to know, be able to do and/or value by the end of the course. Wiggins and McTighe (2005) have popularized the idea of *backward design* as a useful way to plan activity in a course, by working backward from the intended learning outcomes to the assessment tasks and instructional activities – which we identify in advance.

Alignment can occur in relation to different goals for learning within different language learning contexts and across a range of purposes. Alignment is not a new concept in assessment. Basically, it is a fundamental principal of **criterion-referenced assessment** – that assessment tasks should align with, reflect and represent what we expect our students to learn and provide the benchmarks by which we will judge their achievement or mastery.

Activity 2.2 offers additional practice in developing assessment tasks in relation to learning outcomes. Activity 2.3 (see page 43) provides an approach for evaluating the quality of a learning outcome.

ACTIVITY 2.2

In the table below are three of the intended learning outcomes for an advanced course in writing. Working alone or in pairs, list one or more assessment tasks that might be used to evaluate the degree to which one of your students met or exceeded each learning outcome. Refer back to the task analysis above in defining the task or tasks you would use to evaluate each of the outcomes. When you have finished, compare your tasks with those that others have identified.

Intended Learning Outcomes	Assessment Task(s)
1. By the end of the course, students must show they can apply our in-class defined criteria to the evaluation of a televised news report.	
2. At the end of this course, students will be able to write a short news story about a current event.	
3. In order to pass this course, students must demonstrate that they can identify the difference between factual information and opinion in a story about current events.	

ACTIVITY 2.3

How can we evaluate the quality of a learning outcome, which we have written for a course? When we write learning outcomes for a course, there are number of criteria we can apply to evaluate their quality. For example,

1. Is the outcome attainable? Are my students at a level that will allow them to meet or exceed my expectations given the purpose and amount of time allowed for the course?
2. Is the outcome specific and clear? Is it fundamentally important?
3. How exactly will my students demonstrate that they have met my expectations? What specific evidence will I collect to show they have met or exceeded the learning outcome?
4. Could a student pass in my class without meeting this learning outcome? (If you answered yes to this question, it would be a good idea to reconsider and revise.)

Part 1

Directions: To complete this activity, look back at the curriculum guidelines for listening and speaking taken from the ESLCO curriculum. Examine how the overall learning outcomes or expectations are aligned with recommended assessment activities. Using the four criteria listed below, evaluate the quality of the ESLCO learning outcomes (see Activity 2.3, p. 43). Discuss your evaluation with a colleague. Would you modify the outcomes in any way? How would you use this curriculum if you were teaching a course in listening and speaking at this level? Which assessment tasks would you use? Why?

Part 2

Directions: Identify a course you are currently teaching or planning to teach. Consider its purpose, the level of the students enrolled in the course, and how much time you have for the course. Jot down a response for each of the following in the spaces provided below:

Name of course: _____

Purpose: _____

Level of students: _____

Amount of time (number of hours/week and duration): _____

Keeping in mind that there are generally 4–6 learning outcomes for a course, try your hand at writing one or two learning outcomes for the course you listed above. Keep in mind the guidelines introduced above and the criteria for evaluating a learning outcome. If possible, after writing the learning outcomes, ask someone else to apply the criteria list above in evaluating their quality.

Learning Outcomes

1. _____

2. _____

What assessment tasks would allow you to evaluate whether your students have met or exceeded your expectations during the course? List a few of them below. Provide sufficient detail – what micro-level activities comprise the assessment task? What aspect of learning does each activity represent?

Assessment Tasks

1. _____

 2. _____

2.3 Designing Assessment for a Course: Horizontal and Vertical Perspectives

Chapter 2 emphasizes that assessment planning is an integral part of course planning. Assessment (formative, summative, or diagnostic) is embedded in every activity and event within a course. Planning requires us to answer some key questions about a course. Later in Section 2.4, we illustrate an approach to course planning using a template for overall course design. The template identifies the information that is essential in planning a course. Once we have clarified the purpose, participants, resources and restrictions that apply, we have the background necessary to design a course that will stimulate and engage our students and increase their understanding and application of learning. Assessment during a course should be viewed from both horizontal and vertical perspectives.

- *The horizontal plan* begins with the intended learning outcomes and the assessment tasks (e.g., quizzes, reports, presentations,

portfolios, projects, group interactions), which we will use to collect evidence of our students' learning. Through defined learning outcomes and backward design, we move from the outcomes of the course and their assessment to the day-by-day activity in the course.

- *The vertical plan* consists of the day-by-day activity in a course. It maps assessment tasks, activity, events and resources in time (and over time) on to the intended learning outcomes of a course, as shown in [Figure 2.3](#).

Drawing on the ESLCO Curriculum for Listening and Speaking, below is an example of how the assessment tasks we identify allow us to integrate learning outcomes with course activity in the process of course design.

[Table 2.1](#) provides an organizational model for planning, which links the horizontal or overall design of the course with the vertical, or day-by-day. You may wish to use the headings here as a means of developing your own course design, taking into account how much time you have to support your students' mastery of the intended learning outcomes.

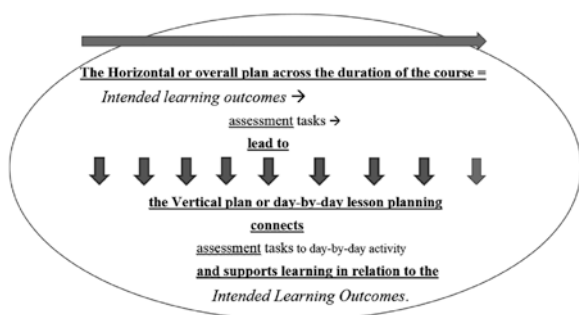


Figure 2.3 Aligning course activity and assessment tasks with intended learning outcomes through horizontal and vertical course planning and design

Table 2.1 Course example: ESLCO, Level 3 (Open), Listening and Speaking

Time allocation: week/date/time	Intended learning outcomes	Assessment: evidence collected and tasks	Activities: purposes, events, resources
Week 1 of 10 2 Sept/ 9:00–10:30	Use speaking skills and strategies to communicate for a variety of classroom and social purposes.		Purpose: Developing rapport in the new class; initial introduction to group interaction.
Event 1 (9:00–9:45)		Formative feedback to support interaction Observational notebook: Record reflections on class interaction, levels of proficiency, fit, and focus.	Event 1: List five things we share in common. 1) Problem solve in pairs; (2) Pairs re-combine in small groups; (3) Groups report to class (write lists on blackboard). Discuss similarities and differences across groups. Resources: None
Event 2 (9:45–10:30)		Self-assessment based on can-do responses; needs analysis to determine individual preferences, goals, and interests.	Purpose: Collecting information about individuals to inform activity. Event 2: Needs analysis:(1) self-assessment of speaking and listening proficiency; (2) goal and interest inventory. Resources: Can-do statements. Interest and goal inventory. Personal Statement of goals for this course.

Having defined the learning outcomes, develop the assessment tasks, which you will use to collect evidence of mastery or achievement, you can then identify the activities (by purposes, events and resources) that will provide the content and context for your students' learning.

Writing down your plan as in [Table 2.1](#) will allow you to reflect on the relationships that you have developed between the course goals or outcomes and the day-to-day learning taking place in your classroom. Is the relationship aligned or not? If not, what changes will you make? These are the questions we might ask in evaluating how well our plan is working in meeting the goals identified for our course.

In the sections below we reconsider the dual influences of our own teaching philosophy and the teaching context in considerations of course planning and design.

2.3.1 Reflecting on Your Teaching Philosophy

In [Chapter 1](#), we explored our own philosophies of teaching and learning and examined some curricular philosophies (e.g., classical humanism, progressivism, reconstructionism and post-modernism or eclecticism), which have tended to dominate educational thinking and language teaching during different periods of history. We noted that the role of the teacher, the student, and the content differs in relation to these philosophies. For example, classical humanism tends to be content-led or content-centred curriculum. Progressive curriculum tends to be learner-led or learner-centred curriculum. In the current educational climate, outcomes-based or reconstructionist curriculum tends to dominate the educational agenda. Reconstructionist curriculum is outcomes-led, evidence-driven and assessment-centred.

What role do teachers play, then, with their individual and varying teaching philosophies given the current emphasis on learning outcomes, quality assurance, accountability and assessment? In some periods of educational history, particularly during past reconstructionist periods, curriculum designers attempted to

teacher proof the curriculum. They viewed teachers as technicians, robots, or means to an educational end. They did not allow for the unique, organic and dynamic character of teachers and students interacting in the processes of learning. However, throughout, attempts to control teachers, reduce their influence on learning processes, or restrict the power that they exercise through decision-making in the day-to-day activities in their classrooms failed.

Although the current model of outcomes-based curriculum is reconstructionist in that it is outcomes-led, it emphasizes the pivotal role that teachers play in planning assessment tasks, activities and events to support their students in meeting or exceeding intended course outcomes. Whatever our philosophy of teaching and learning, we remain the primary decision-maker in defining how to best support our students' learning. It is important to acknowledge, however, that our decision-making will be influenced by the context within which we are teaching as we seek to align our classroom activity through assessment with the learning outcomes that define our course goals or expectations. In section 2.3.2, we discuss some recurring contexts for alignment.

2.3.2 Exploring Various Contexts for Alignment

There are many different contexts for alignment of assessment tasks and classroom activity with intended learning outcomes. For example, outcomes may be defined by:

- benchmarks and standards;
- curricular guidelines (like those in [Figure 2.2](#), from the ESLCO curriculum);
- external tests;
- textbooks, which identify goals for learning chapter by chapter;
or
- needs analysis.

Each of these contexts (and others, as this is not an all-inclusive list) will influence how we define the intended

learning outcomes in our course, and which assessment tasks we will use in order to collect evidence of our students' development and achievement. In the following section, we will explore each of these contexts and examine strategies that will support our identification and use of assessment methods that agree or align with our goals for learning.

Benchmarks and standards

Whether they are defined as *benchmarks* or *standards*, these are statements, which describe expected developmental processes and stages in learning a language. They provide signpost descriptors along a developmental continuum through increasing levels of language capability in listening, speaking, reading and writing. As such, they are essentially criteria that define what students should know and be able to do, given the specific purposes for which a course is being taught and the target learners that are enrolled in the course. As mentioned earlier in the chapter, standards may be fixed requirements imposed on learning.

In some instances these criteria are non-negotiable, reflecting specific expectations, which must be met by the end of a specific year of study. In this case, the standards are the intended learning outcomes, which teachers must address through their design of assessment tasks and activity in their classrooms. For example, the *No Child Left Behind* policy in the United States mandated the definition of standards for English language proficiency from kindergarten through Grade 12. This definition was enforced through external standardized testing at the end of a period of study (i.e., each grade level). These powerful standards have had considerable impact on funding across school systems in the United States; testing is high-stakes, because much depends on the performance of students on external tests, including, in some cases, teachers' retention and/or employment. Thus, the standards guided all decision-making – from curriculum development, to instruction, to ongoing assessment at the classroom level.

In other instances, minimum competency standards may have less impact, but are nonetheless important in terms of their influence on the definition of learning outcomes. For example, in order to graduate from a Canadian high school in Ontario, students must pass the Ontario Secondary School Literacy Test (OSSLT),¹ an external, standardized test administered at the end of Grade 9. The OSSLT measures whether or not students are meeting the minimum standard for literacy across all subjects up to the end of Grade 9. Successful completion of the literacy test is one of the requirements to earn an Ontario Secondary School Diploma (see Cheng, Klinger and Zheng, 2007; Fox and Cheng, 2007). This standard for literacy generates the most impact at the Grade 9 level. At other levels, curriculum such as the ESLCO (see Figure 2.2 for an example) defines (and allows teachers to define) learning outcomes in a more flexible manner.

The developmental continuum, which is elaborated in standards, typically has milestones or markers to define key developmental changes or stages. These milestones or *benchmarks* are criterion referenced to set off one developmental stage from another. The benchmark levels of the Common European Framework of Reference (CEFR) are perhaps the most notable examples of such benchmarks.² CEFR provides a comprehensive and detailed description of what learners need to be able to do in order to communicate effectively in a language at given levels of mastery. As its name suggests, CEFR was developed with the intent of providing a commonly shared understanding across the 47 member countries of the Council of Europe of the criteria that define language development, and was intended as a reference for what learners can do when they use language at increasing levels of proficiency. The CEFR describes foreign language proficiency at six levels: A1 and A2, B1 and B2, C1 and C2. It also defines three 'plus' levels (A2+, B1+, B2+). Arguably the most frequent references to CEFR are the benchmark criteria, which define six levels of proficiency, from A1 (the

lowest) to C2 (the highest). The criteria at each of the six levels describe in positive terms how learners use language and have become the reference points for many proficiency scales and proficiency tests. Benchmarks have been developed by many other national groups (e.g., Australia, Canada), and are often used as reference points in the assessment of proficiency and language development over time. They provide teachers with a resource in defining learning outcomes for groups of learners at different levels of proficiency.

Curricular guidelines

Curricular guidelines are an important resource for learning outcomes, as we have seen in the case of the ESLCO guidelines for listening and speaking. Such guidelines may be flexible and simply a point of reference for teachers in defining their own learning outcomes for an individual class. In other contexts, they may be more prescriptive and codify the learning outcomes for all teachers across a system. When they are prescriptive, they are often accompanied by system-wide external assessment.

External tests

There are many language teachers who are involved in preparing their students to write and pass external high-stakes language proficiency tests that determine, in whole or in part, whether a student can enter university, practice medicine, become a citizen of a new country, or apply for a job. In the context of test preparation, language teachers may feel conflicted because their students are driven by the need to pass the test, but their teachers understand that developing a student's language proficiency, which, after all, the external tests are measuring, is the most useful outcome of a test preparation course. The elaboration of intended learning outcomes for a test preparation course can help to resolve

some of the tension that teachers may feel in this context. Learning outcomes can respond to both the student's short-term goal to pass the test as well as the teacher's recognition that supporting language development will serve the student's long-term interests.

Textbooks

In many parts of the world, textbooks define the expectations for learning in a course. In some jurisdictions, ministries identify and approve one or more textbooks. Subsequently, school administrators, programme coordinators, or teachers select the textbook they find most useful for their students and programme. The textbook provides a resource for the development of intended learning outcomes for a course. Through backward design (from textbook to learning outcomes) teachers can elaborate assessment tasks and activities that will collect evidence that a student has met or exceeded the intended learning outcomes for a course, which are nonetheless related to the expectations set out in the textbook and draw on the content the textbook incorporates.

Needs analysis

Yet another resource for the development of learning outcomes is needs analysis. Needs analysis is often the starting point for language teachers at the beginning of a new course and an essential assessment tool for gathering information that will help teachers in their decision-making as a course unfolds. In [Chapter 5](#) we delve more into the development and use of different types of needs analysis. As is the case with textbooks, curricular guidelines, or benchmarks, the information teachers elicit from their students about, for example, their purposes for taking the course, their goals, levels of proficiency, or interests can help refine the learning outcomes and assessment tasks that a teacher decides to use.

ACTIVITY 2.4

Directions: Stop for a moment and reflect on your own teaching context (or one that you are familiar with) and answer the following questions.

- At the present time, are you teaching (or are you planning to teach) in a context which has explicit learning criteria or expectations such as *benchmarks* or *standards*?
- Or, are you teaching in relation to *curricular guidelines* (like the ones in [Figure 2.2](#))?
- Or, are you teaching to a *textbook*, which spells out goals for learning chapter by chapter, and often across different volumes in a series that is geared to levels of language performance or proficiency?
- Are you defining *your own course goals*, based on your students' individual needs, interests, purposes for studying and levels?
- Are you teaching students who have a narrow but compelling goal of, for example, passing a test like the Test of English as a Foreign Language (TOEFL) Internet based test (iBT)? Or the International English Language Testing System (IELTS)? If so, you may find at times that your students' goals are in conflict with your own judgment of how best to support their language development.

Write a short response to the following questions: How does the context in which you are teaching influence what you set as course goals or learning outcomes? What are the constraints? What are the opportunities? You may prefer to make a list in point form.

My current or intended context:

Opportunities

- _____
- _____
- _____

Constraints

- _____
- _____
- _____

How important are the opportunities and constraints in influencing what you actually do in your classroom? When we plan a course, we evaluate both the opportunities and the constraints of a teaching (and learning) context.

2.4 A Template for Course Planning/Syllabus Design

A template is provided here for planning a course. It is offered as a useful model which might be used at a later time. It might also serve as a homework assignment or as an in-class activity.

If we use this template to design one of our own courses, our responses to each of the questions in the template will be informal (perhaps we will simply jot down a response in point form based on what we know, suspect and can later find out). Responding to these questions will help us, as individual teachers, to clarify our understanding of the course and to develop a systematic plan for course activities that will respond to the particular learning and teaching context of the course. However, if we are using the template to prepare a *syllabus* (i.e., a course plan which will be followed by other teachers in your programme or others), our responses will be more formal and may be written up as a background report for discussion and reference as a foundational curriculum document.

In order to clearly define the course, we need to answer each of the questions in the template. We can use a range of approaches or strategies to elicit the information required (e.g., conversations with colleagues, reviewing materials, interviews,

questionnaires, focus groups). It may be useful to examine other or previous programme descriptions, syllabuses, course documents, calendar or curricular descriptions.

Talking to (or interviewing) our colleagues who have taught the course before, may teach similar courses (at the same, higher, or lower levels), or who plan to teach the course in future is also very helpful. By the same token, talking with or interviewing students who have taken the course before (or similar courses) provides useful information. Other stakeholders may also have important information about the course (e.g., programme directors, coordinators, administrators). Setting up meetings or taking the opportunity for a casual chat about the course can often fill in missing information, clarify our understandings, or extend our thinking in important ways.

In reviewing the template below, note the way in which the assessment tasks are directly informed by the information collected through the use of the template. There are three steps identified in this template. The first step explores critical background information that situates the course within key features of the teaching and learning context. Informed by the context information generated in Step 1, Step 2 provides a guide for planning the overall or horizontal direction (scope) of the course (see questions 5–6); and the activities and events that define what happens in the day-by-day (sequence) of the course (see questions 7–9). Step 3 connects the horizontal to the vertical planning using the model in [Table 2.1](#).

Step 1: Questions 1–4 help to identify the specific characteristics of the course, which will influence the decisions we make in course planning. Answers to these questions provide teachers with essential background.

1. Why is this course necessary?

Type of course: ESL (English as a Second Language – language instruction in a context where English is spoken extensively outside the classroom context) – but also FSL (French as a Second Language), SSL (Spanish as a Second Language) and so on

EFL (English as a Foreign Language – language instruction in a context where English is used in the classroom, but another or other languages are used extensively outside the classroom context) – but also, JFL (Japanese as a Foreign Language) or CFL (Chinese as a Foreign Language)

ESP (English for Specific Purposes, such as engineering, business, nursing)

EAP (English for Academic Purposes)

Settlement (language instruction for newly arrived immigrants or refugees)

Test Preparation (language instruction to support test performance)

Immersion (content-based instruction in the target language)

Other _____

- Having specified the type of course, what precisely is its purpose?
 - Are there externally defined expectations (e.g., benchmarks and standards, curricular guidelines, textbooks, or external tests)?
 - How restrictive are the expectations? To what degree will they impact the way in which the course will be taught?
 - Is the curriculum or educational framework within which the course is embedded *explicit* (i.e., communicated through documentation) or *implicit* (an unwritten understanding of a community of teachers)?
2. Who are the stakeholders?

In order to answer this question, begin by consulting approaches to needs analysis (see, for example, J. D. Brown's recommendations (1995, pp. 35–65) regarding needs analysis in his book *Elements of language curriculum*).

Students:

Age?

Background relevant to the course/programme (e.g., academic, linguistic, cultural)?

Needs (relevant to the purpose of the course)?

Lacks (relevant to the purpose of the course/programme)?

Wants (relevant to the purpose of the course)?

Teacher(s): Training/educational background?
 Experience relevant to course/programme?
 Philosophy of teaching and learning (i.e., cultural predisposition to methods, teaching, learning, textbooks, students, change)?
 Needs (relevant to the purpose of the course)?
 Lacks (relevant to the purpose of the course/programme)?
 Wants (relevant to the purpose of the course)?

Other key stakeholders: (e.g., principals, coordinators, owners, parents, deans, directors)
 Expectations?
 Predispositions?
 Impact/influence (power)?
 History?

3. Where will the course take place? What are the key features of the learning space? (Or what is the typical context of instruction in the case of syllabus design?)

Classroom (e.g., physical space in a school, online space in a computer lab, 3D virtual learning spaces)?

Key features? _____

Workplace (e.g., engineering firm, business office, hotel meeting room)?

Key features? _____

Other _____

Key features? _____

Space/resources (constraints and potential)?

Materials (photocopy facilities, paper, chalk and so on)?

Resources (blackboards, interactive boards, computers/ projectors and so on)?

Class size (range in the case of syllabus design)?

Room configuration (movable chairs and tables, fixed desks and so on)?

Other (libraries, volunteers, guest speakers and so on)?

4. When will the course take place?

Total time allowed? ____ months
 ____ contact days
 ____ contact hours

How is contact time subdivided?

____ hours/week
 ____ classes/week of ____ hours

Step 2: Having considered the key background, we are now ready to plan our course in specific detail. In responding to questions 5–9, you will want to use the table presented in Step 3. Note: questions 5–7 help to frame the macro-level or overall direction of the course (which might be considered the horizontal perspective in course planning); questions 7–9 relate to the day-to-day planning of activity in a course (the vertical perspective in course planning).

5. What are the intended learning outcomes for the course? (identifying outcomes)
6. How will I evaluate the overall effectiveness, impact and usefulness of the course? (mapping assessment tasks onto learning outcomes)
 - What evidence will I collect over time to evaluate the quality of the course in achieving the purposes for which it was designed?
7. How will I know my students have learned what I have taught? How will I assess their learning in relation to the intended outcomes of the course? (aligning assessment with intended learning outcomes; mapping assessment onto activity over time within the course)
 - What evidence will I collect over time to evaluate my students' achievement in meeting or exceeding the intended learning outcomes?
 - What will be the most useful assessment tasks given the context and purpose of the course?

Question 7 relates to assessment tasks which map onto the overall intended learning outcomes of the course on the one hand, and day-to-day course activity on the other. Assessment tasks are at the nexus of both horizontal (long-term) and vertical (day-to-day) planning in a course.

8. What will be taught? (mapping content onto outcomes and time)
9. How will the course(s) be taught? (mapping content onto activity over time)

Step 3: Now use the table below to connect the day-to-day assessment with overall course purpose and day-to-day activity. You can use [Table 2.1](#) as the example to follow.

My course:

Time allocation week/date/time	Intended learning outcomes	Assessment: evidence collected and tasks	Activities (purpose, events, resources)
-----------------------------------	----------------------------------	--	---

2.5 Looking Back at Chapter 2

In Chapter 2 we considered *what* to assess in relation to goals, expectations, and our identification and definition of intended learning outcomes for a course. We examined how to write and evaluate learning outcomes, their relationship with assessment tasks and class activity, and how they are influenced by the contexts in which we teach. Our focus in this chapter was on alignment and backward design from learning goals and outcomes to assessment tasks and activity. Now that you have read this chapter, how would you respond to the following questions?

- How does the alignment of intended learning outcomes and assessment tasks improve the quality of a course?
- Why do we view teachers as course designers? Why do we use the verb ‘design’?
- In what ways do explicit learning outcomes and assessment tasks support grading and evaluation at the end of a course?
- Why is it important to share intended learning outcomes with our students?
- What role can our fellow teachers play in helping to refine our learning outcomes?

You may want to read more about the issues discussed in this chapter. The following are suggested resources for further reading and discussion.

Suggested Readings

Biggs, J. & Tang, C. (2011). *Teaching for quality learning at university*, 4th edition. Maidenhead: McGraw Hill.

Biggs and Tang provide detailed information about the use of *intended learning outcomes* in course planning, design and implementation. Their seminal review provides extended examples drawn from their use of this approach at the post-secondary level in Hong Kong.

Graves, K. (2000). *Designing language courses: A guide for teachers*. Boston, MA: Heinle & Heinle.

Drawing on models and approaches from the literature on course planning and curriculum design, Graves provides a particularly comprehensive and accessible look at the process of course planning and design. She adeptly integrates language teachers' accounts, reflections and practices in this systematic and detailed examination of key steps in language course design.

Wiggins, G. & McTighe, J. (2005). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Wiggins and McTighe offer a well-detailed guide to course planning using *backward design* for teachers, assessors and curriculum planners. They argue against other planning approaches, which emphasize coverage of content coverage or activity, and in favour of an approach that focuses on students' meaningful *understanding*. Their book abounds with practical advice, examples and strategies. It is accompanied by a website, <http://ubdexchange.org>, which features templates for design, curricular units and assessment tasks. (Retrieved 20 October 2016.)

End Notes

- 1 For the purpose of the OSSLI, literacy comprises the reading and writing skills required to understand reading selections and to communicate through a variety of written forms as expected in the Ontario Curriculum across all subjects up to the end of Grade 9. In the reading component of the test, students use reading strategies to interact with a variety of narrative, informational and graphic selections to

construct an understanding of the meaning of the texts. Students are asked to demonstrate their understanding of explicit (directly stated) and implicit (indirectly stated) meanings as well as to connect their understanding of the text to their personal experience and knowledge. The reading selections that students are asked to read are representative of those expected across subject areas in the Ontario Curriculum up to the end of Grade 9. In the writing component, students are prompted to write two short responses, a series of paragraphs expressing an opinion, and a news report. Through their responses, students demonstrate their ability to communicate ideas and information clearly and coherently. Since writing on large-scale assessments does not allow for a complete revision and refinement process, written work on the OSSLT is scored as first-draft (unpolished) writing. The written forms in which students are asked to write are representative of those expected across subject areas in the Ontario Curriculum up to the end of Grade 9. Additional information can be found at: <http://www.eqao.com/en/assessments/OSSLT/educators/Pages/About.aspx>. (Retrieved 20 October 2016.)

- 2 The uses of the Framework include: The planning of language learning programmes in terms of their assumptions regarding prior knowledge, and their articulation with earlier learning, particularly at interfaces between primary, lower secondary, upper secondary and higher/further education; their objectives; and their content. The planning of language certification in terms of the content syllabus of examinations; and assessment criteria, in terms of positive achievement rather than negative deficiencies. The planning of self-directed learning, including raising the learner's awareness of his or her present state of knowledge; self-setting of feasible and worthwhile objectives; selection of materials; and self-assessment. Learning programmes and certification can be: global, bringing a learner forward in all dimensions of language proficiency and communicative competence; modular, improving the learner's proficiency in a restricted area for a particular purpose; weighted, emphasizing learning in certain directions and producing a 'profile' in which a higher level is attained in some areas of knowledge and skill than others; and partial, taking responsibility only for certain activities and skills (e.g., reception) and leaving others aside. Additional information is available at: <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680459f97> (Retrieved 20 October 2016.)

CHAPTER 3 How Do We Assess?

Activate your learning

- What are the differences between large-scale testing and classroom assessment practices?
- What is an assessment plan?
- What assessment tools and procedures do we use?
- What is a portfolio? Why is portfolio assessment a good example of continuous assessment (i.e., combining assessment *for*, *as* and *of* learning)?
- How can we use assessment to support our students' learning?

3.1 Delineating Large-Scale Testing and Classroom Assessment Practice

As we discussed in [Chapter 1](#), teachers routinely deal with *large-scale* testing which is external to their classrooms, often with more at stake (or higher stakes). They routinely engage in *small-scale* testing, which is internal to their classrooms and measures achievement at the end of a unit or course with less at stake or (with lower stakes). Such testing, often referred to as assessment *of* learning, tends to be a special event, a signpost or marker in the flow of activity within a course. On the other hand, assessment *for* and *as* learning is part of ongoing classroom assessment practices. In [Chapter 2](#), we examined how defining learning goals and outcomes, and designing our learning activities and assessment tasks in relation to those goals and outcomes, can both support our students' learning and inform and focus our teaching.

Before discussing the processes and procedures of classroom assessment planning and practices, we will highlight some of the key differences between large-scale testing and classroom assessment practices. We will then walk you through classroom test development in [Chapter 4](#).

Brookhart (2003) argues that the key difference in these two practices is the *context* (or lack thereof) where the testing and assessment take place. She describes the context dependence of classroom assessment, its inextricable relationship with instruction, and its simultaneous formative and summative functions. The difference between decontextualized measurement, one-time measurement undertaken by outsiders as in large-scale testing, and the ongoing contextualized measurement undertaken by a teacher in his/her classroom changes the nature of the assessment, that is, what, how and why we assess. For example, in large-scale assessment content does not come from a teacher, with whom students have a relationship, and the students are not the primary consumers of its results. High-quality classroom assessments are not only measures of learning, but genuine episodes of learning in and of themselves. In classroom assessment, there is a psychosocial context wherein classroom assessment and instruction are integrated, and classroom assessment is primarily formative, that is, with the purpose of supporting learning as defined by a course's learning goals and outcomes.

The intention behind such classroom practices is generally to help students to develop the ability to:

- define their own learning goals;
- monitor their current level of work in relation to their desired performance; and
- act in such a way as to close the gap between their current performance and their goals.

In the classroom, it is the students who are 'measured'. It is also students who need to be encouraged to draw inferences

from information arising as a result of the measurement, and then to take actions with the support and guidance of their teachers. This process is the central part of the formative assessment process. Therefore, students' awareness of and benefits from assessment information are of critical importance in evaluating the quality of classroom assessment practice. In addition, teachers' feedback varies in degree (positive or negative, focusing on strengths or shortcomings) and purpose (evaluative or descriptive), and therefore feedback varies in formative usefulness to students. Usefulness of feedback, in turn, is part of the consequential evidence for the validity of classroom assessment information (Messick, 1989).

In classroom assessment, we ask the following questions:

- What does a particular speaking or writing performance mean to the student?
- Does the student accurately interpret his or her performance?
- Is the communication from the teacher to the student clear; that is, does the student understand teacher feedback and teacher grades in the way the teacher intended?
- Is that feedback helpful? Does it support a student's development? If so, how?
- Does the student pay attention to the feedback?
- Does the student trust the information?
- How helpful is the information in fostering future learning or continued progress toward the student's individual goals and the course's learning outcomes?

In fact, students are in the best position to respond to the above questions if, as teachers, we teach them how to reflect on their learning (Sadler, 1989; Stiggins, 2001) and provide them with time and resources to engage in self-reflection. This is the central and foundational principle of assessment as learning. These above questions are related to the assessment events and assessment processes discussed in [Chapters 1 and 2](#).

In the case of large-scale testing, test design follows processes and procedures, which can be very useful for us to understand

and apply in our own small-scale classroom testing (see Davidson and Lynch, 2002, on writing and using test specifications, and Grabowski and Dakin, 2014, on test development literacy). Whereas our own classroom tests tend to reflect what we have emphasized in our teaching, the amount of time spent on learning, and our increasing knowledge of the students we are working with, large-scale tests are not influenced by an individual classroom context. They are external to the ongoing learning taking place day-to-day. Such external, large-scale tests tend to have high stakes – the more important the decisions made by test score users, the more impact they may have on our students’ learning and our own teaching. When an external test is aligned with the learning outcomes that we are working toward in our own classrooms, there is a potential for positive impact on our students’ learning.

It is often the case that what we do and how we do it may be directly influenced by a large-scale standardized test that exerts external influence on our students’ attitudes towards learning a language, their goals (which may be short-term, and simply reflect their expressed need to ‘pass the test’), and our own concerns about how best to help our students to achieve their goals. When the interpretation of a test has consequences for us and for our students, when the inferences we draw from a test exert influence (positive or negative), we typically refer to this as *washback* – the influence of testing on teaching and learning (see, for example, Cheng, 2008, 2014; Fox and Cheng, 2007).

The validity goal in classroom assessment is to meaningfully and accurately interpret assessment information (e.g., grades, scores, teachers’ oral and written comments, student observations of errors, and students’ recognitions about their learning – those remarkable eureka moments when students become aware of and/or acknowledge their learning). It is important to point out that large-scale testing and classroom assessment follow similar processes and procedures despite fundamental differences when it comes to the interpretations

and uses of test scores and assessment results. Both have washback potential. Our intention as teachers is to support positive washback from tests (both external and internal to our classrooms) and through assessment to increase meaningful learning.

As teachers, we are sensitive to the influence that tests can have on what we do with our students, and how we do it. Look at the anonymous comments of some ESL teachers who wrote about the impact of tests on their teaching at a recent workshop on classroom assessment:

- 'I know how important it is for my students to pass the test, but all they want to do is take practice tests. This is a waste of time as far as I'm concerned. It's much more important for them to actually increase their language proficiency – not their test taking skills.'
- 'We had to stop working on the class newspaper because they had to get ready for the test. It's such a shame really. They were learning so much and enjoying the challenge.'
- 'It's so frustrating because I've been working with my students on their speaking and listening ... because the new curriculum emphasizes speaking and listening. But there's no test of speaking, so what's the point? The only learning that counts is what the test measures.'
- 'My whole class has lost their motivation. They were so discouraged when they took the test. It's like the test let all the air out of our balloons!'
- 'We seem to spend all our time on testing, testing, testing!'

Each of the comments above is an example of negative washback.

3.2 Creating a Classroom Assessment Plan

An assessment plan is an overall guide for how we will assess students' achievement of the learning goals and outcomes relevant to instruction. Cheng, Rogers and Wang (2008)

conducted a comparative interview study of teachers who taught in a range of ESL and EFL contexts. Their study found that whereas in Canada instructors were mainly teaching English courses to pre-university students, in Hong Kong and China instructors were most often teaching English courses to undergraduate and graduate students. The differences between the courses can be explained in terms of the purposes the students had for studying English. The Canadian ESL courses were mainly designed to prepare students who came from different parts of the world with insufficient English proficiency to enter an English-medium university. The Hong Kong ESL courses were mainly designed to train university students so that they were competent in using English in their specialized areas of study. The Chinese EFL courses were mainly designed to prepare university students to further enhance their language proficiency in English and pass the required College English Test.

When the instructors were asked whether or not they had an assessment plan for their courses, 21 of the 24 instructors in Canada, all of the instructors in Hong Kong, and half of the instructors in China indicated that they did. A review of their assessment plans revealed that the plans of the Canadian and Hong Kong instructors tended to focus on class presentations, portfolios, journals, projects and individual presentations in class with due dates. Further, their plans included accounting for student effort. In contrast, the Chinese instructors tended to use quizzes and tests, and they did not consider awarding credit for student effort.

Before considering the implications of the Cheng and colleagues' (2008) results, what does an assessment plan look like? It is important to note that there is no one way to set out the assessment plan for a course you are teaching. In some contexts what counts toward the final mark in a course may be predetermined at a department, school, or programme level. The flexibility you have in defining an assessment plan for your class will depend upon your instructional context. However, the starting point will be to identify all of the

performances that you will consider in calculating the final mark and their relationship to the overall learning outcomes for your course.

Take a look at the outline of assessment activity in the Assessment Plans of Teacher A and Teacher B below. Consider how context influences their assessment planning.

Context: Teacher A is teaching an intermediate class in an English as a Second Language (ESL) school for adults. She is free to choose which activities will count toward the final pass/fail grade in her class. If students do not pass, they are simply required to repeat the same level. Most students in this programme are taking English out of personal interest.

Teacher A's Assessment Plan

In order to successfully complete this course, students must receive a *pass* on each of the *four course requirements*:

Course Requirement	Evaluation
1. Personal reflective journal on learning. One entry per week which responds to these questions: • What did I learn this week? • What is still a problem? • What would I like to learn more about?	To receive a pass on this requirement a minimum of eight entries must be received. Entries should be collected in one folder and handed in on each of the following dates: • 10 January • 17 January • 24 January • 31 January • 7 February • 14 February • 21 February • 28 February • 7 March • 14 March

- | | |
|--|---|
| 2. Group poster and presentation | To receive a pass on this requirement students must contribute to the development of a poster and presentation on a topic of interest to the group (additional details will be provided in class). |
| 3. In-class tests (there will be four tests during the course) | To receive a pass on this requirement students must average 70% or higher on the four tests. We will have a short test after we finish each unit in the textbook. These will be announced in advance and support will be provided. |
| 4. Final in-class reflection on learning | To receive a pass on this requirement students must bring their personal reflective journals to class. Instructions will be provided for the in-class reflection at the beginning of class on 21 March. |
-

Context: Teacher B is teaching an intermediate class for credit toward a high school diploma. Letter grades are issued to the students at the end of the course. Students must pass this course if they wish to graduate. If they fail to meet the course requirements (i.e. fail the course), they will be required to reregister and take the course again. A textbook is assigned to all of the classes at this level and the final test is based on the textbook readings and exercises.

Teacher B's Assessment Plan

In order to pass this course, students must achieve an overall *average of 70%* across the following course assignments. All assignments are important, but some contribute more heavily to course outcomes than others.

Course Requirement	Evaluation	Weight
Summary writing	Students will be asked to summarize textbook readings that have been assigned in class at four points during the term. Exercises assigned from the textbook will also be due on the following dates: • 25 September • 15 October • 5 November • 5 December	40% (50% summary; 50% exercise)
Learning Log	Students should collect information related to their research reports in a learning log (see below). While collecting information, reflect in writing on what the information will contribute to your research report and presentation. Learning logs should be brought to class on the following dates for review and discussion. • 18 September • 6 October • 22 October • 15 November We will work with the learning logs in class on these dates.	20%

Research Report	Drawing on the information collected in the learning log, the final research report on a topic related to personal interest is due in class on 19 December. Further details will be discussed in class.	20%
Final Test	The final test will be based on the textbook only. Details regarding test format will be provided in class. The test will be administered to all students in the programme on 12 December.	20%

In the two assessment plans above, Teacher A chose to work on a pass or fail basis. Students were required to meet the requirements she set out, but the relative quality of their work was not important in passing the course. The stakes were lower for her students. If they did not meet the minimum requirements to pass the course, they could simply repeat it.

Teacher A used a great deal of student self-reflection. Questions such as ‘What did I learn this week?’, ‘What is still a problem?’ and ‘What would I like to learn more about?’ are very helpful in focusing students’ attention on their own learning process and helping students to self-regulate their learning. In Teacher A’s plan, this assessment *as learning* practice is balanced by classroom tasks such as the poster presentation (assessment *for learning*) and classroom tests (assessment *of learning*). Such an assessment plan focuses more on the mastery of student learning when the assessment system adopted is pass/fail, that is, the assessment plan aims to map out what students have learned to a certain level (pass) without further categorizing the learning performance levels.

On the other hand, Teacher B was working in a context with less flexibility (e.g., the textbook was required; the final test must be based on the textbook). The stakes were higher for students in Teacher B's classes – they need to pass this course in order to graduate from high school. Teacher B has a more traditional approach to assessment where a number of assessment tasks (e.g., summary writing tasks, learning logs and a research report) and tests (a final test) are combined to obtain students' final marks. It is important to note, however, that Teacher B also emphasized the importance of ongoing processes of learning (e.g., summaries of textbook information, collections of information for the research report, learning logs) by allocating 60% of the mark to ongoing processes of learning, and 40% to outcomes or products (e.g., the research report, the final test).

These Assessment Plans demonstrate the role of context and constraint in planning course assessment and how much each requirement will contribute to the final mark. They also illustrate that teachers have considerable latitude, even in courses with defined requirements and higher stakes, in defining the importance of ongoing activities that support the learning process in relation to activities that are products or outcomes of a course (see, for example, Fox, 2009).

You may want to stop reading here and do Activity 3.1. It asks you to reflect on your own experiences with washback.

ACTIVITY 3.1

Reflect on the washback potential suggested by each of the following. If possible, compare your own answers with those of a colleague or a group.

1. What are the strategies in assessment *of* learning, assessment *for* learning and assessment *as* learning?
2. Have you ever taught a class to students who were taking your course because they needed to pass a proficiency test in order to enter a university degree programme? If yes, did this external

testing context impact your teaching? Did it appear to influence your students' willingness to learn? Explain.

3. Have you ever taught a course where your students needed to pass an external test (or tests) in order to receive their high school diploma? If yes, did this influence how you taught and/or how your students responded to your teaching? Explain.
4. Have you ever taught in a programme where all of the teachers collaborated on an external test to measure students' development across a level? It may even have been the case that your own philosophy of teaching was not aligned with the test which your colleagues negotiated. If you have experienced this, how did the negotiated test influence your teaching? If you have not experienced this, how do you think it would influence your teaching?
5. Think of a context where an external, large-scale test was a requirement for you or for your students. Did the test support learning? Did you or your students benefit from the test? If yes, how? If not, why not? Did the system within which the test was used benefit from the test?
6. What would your own Assessment Plan look like if you were teaching in a context like that in question four above? Working alone or with a colleague, take a few minutes and sketch out a possible assessment plan using the models for Teachers A and B above.

3.3 Classroom Assessment Tools and Procedures

We began this chapter with a discussion of internal, classroom assessment (i.e., small-scale) and external assessment (i.e., large-scale). We then examined the overall assessment plans teachers employed with their students in the classroom. Clearly, tests are only one of the many assessment tools we use as teachers on an ongoing basis (as Teacher A's and B's assessment plans illustrate). In [Chapter 4](#) we will examine tests and test development in depth. It is important to note, however, that any

number of assessment tools can support student learning when the tools give our students clear ideas about what is important to learn, provide students with criteria, or clearly communicate expectations for good work, and when assessment matches instruction. In the Appendix, at the end of this book, you will find a list of frequently used assessment tools, along with examples and short explanations of how they are used.

3.3.1 Classroom Assessment Tools

Assessment tools, such as those mentioned above, can support student learning when the tools give the students clear ideas about what is important to learn and the criteria or expectations for good work, and when assessment is aligned with or matches instruction. While you explore the assessment tools you use, you can reflect on which methods best help your students to achieve their learning goals and how to use various assessment methods to support all students. A combined use of assessment tools of receptive and productive language skills, language components (e.g., grammar, vocabulary and pronunciation), and direct and indirect assessment of skills can help to enhance the validity and reliability of our assessment practices. You may find the list of assessment tools, provided in the Appendix, helpful in completing Activity 3.2.

3.3.2 Categorizing Assessment Methods

There are a number of ways we can categorize assessment methods. For example, we can categorize them in terms of question or item types. For instance, *open-ended or supply questions* require students to provide extended responses; *selection questions* require students to select an answer from a number of different options; and *student-centred assessments* require student participation and decision-making in ongoing assessment practice (see Cheng, Rogers and Hu, 2004):

- Open-ended or Supply-type questions
 - Short-answer questions
 - Oral presentation

- ...
- ...
- ...
- Selection-type questions
 - Multiple-choice questions
 - True or false questions
 - ...
 - ...
 - ...
- Student-centred assessment
 - Self-assessment
 - Portfolio
 - ...
 - ...
 - ...

ACTIVITY 3.2

Add to the list below, different ways, methods and tools with which you have been assessed as students, or which you have used to assess your own students, or the methods and tools you are aware of, but have not used before.

- Essay questions
- Multiple-choice questions
- True or false questions
- Oral presentation
- Writing Portfolio
- Self-assessment
- ...
- ...
- ...

Alternatively, we can categorize assessment methods as observations, conversations, or products, as in [Figure 3.1](#).

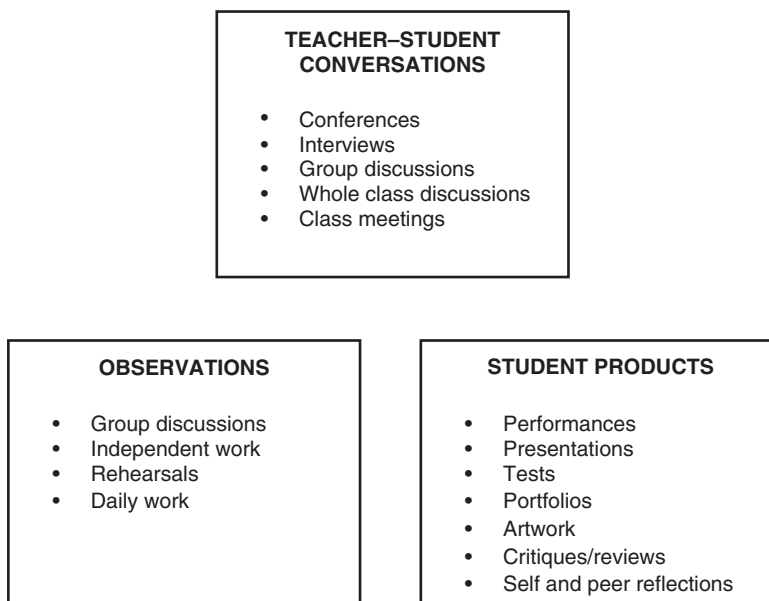


Figure 3.1 Assessment observations, conversations and products

In the education literature on assessment, the following three categories have also been identified:

- Teacher-made assessment methods
- Student-conducted assessment methods
- Standardized testing in reading, writing and speaking/listening

These three categories also represent major assessment constructs that generally guide teachers' assessment practices no matter where language is taught or who is learning it.

Instructor-made assessment methods refer to those assessment methods designed and administered by teachers *to* their students, whereas student-conducted assessment methods are those that *directly involve* students' participation in the

assessment process. These three categories best summarize the assessment tools in an assessment *for* learning context, and can support you in rethinking your own assessment practice.

It is important to recognize that the more our assessment practice directly involves our students, the greater the probability that our students will develop increased awareness of their learning progress and take responsibility for their learning. When students take ownership of their learning, when they want to learn, they are able to learn more effectively. Our assessment practices are there to engage and motivate our students and support our students' learning.

At the end of this chapter there is information about websites where more information is provided on codes of practice that have been developed to guide ethical, fair and valid assessment practice.

ACTIVITY 3.3

This two-part activity will help you to reflect on your own assessment practices. Keep in mind that there are no *right* or *wrong* answers here. As a teacher, you use various assessment methods to achieve instructional and assessment goals, as well as goals in terms of classroom management. After you have completed the activity, you may want to know what ESL/EFL teachers in other contexts reported about their assessment practices, reported in Cheng, Rogers and Hu (2004) or Cheng, Rogers and Wang (2008).

Part 1

Begin by indicating which assessment methods you actually use by putting a check mark (✓) in the column to the left in Tables 3.1 to 3.3. Notice that we begin with reading, followed by writing, speaking and listening. Spaces have been provided at

the end of the list for methods not on the list. If you use other methods, please write these down as they are equally relevant to your assessment practice.

Second, rank the methods you put check marks (✓) next to in each of the tables. For the method you use the most, place a '1' in the column to the right ('Rank'); a '2' in the 'Rank' column next to the method you use the second most; a '3' next to the third most used method; and continue until you have ranked all the methods you identified. Please rank only the methods that you indicated you use.

Table 3.1 Reading

1. Methods I use to assess reading (✓)	Assessment Methods	2. Rank
	1. Read aloud/dictation 2. Oral interviews/questioning 3. Teacher-made tests containing a. cloze items (e.g., words or phrases are systematically removed from a passage and students are required to fill in or identify what's missing) b. sentence-completion items c. true/false items d. matching items e. multiple-choice items f. interpretative items (e.g., reading passage; or a map or a set of directions)	

- g. forms (such as an application form or an order form of some kind)
- h. short-answer items
- i. Student summaries of what is read
- 4. Student reading response journals
- 5. Student portfolio
- 6. Peer-assessment
- 7. Self-assessment
- 8. Standardized reading tests
- 9. Other:
- 10. Other:

Table 3.2 Writing

1. Methods I use to assess writing (✓)	Assessment Methods	2. Rank
	1. Teacher-made tests containing a. true/false items b. matching items c. multiple-choice items to identify grammatical error(s) in a sentence d. editing a piece of writing such as a sentence or a paragraph e. short essay f. long essay 2. Student journal 3. Peer-assessment 4. Self-assessment	

5. Student portfolio
6. Standardized writing tests
7. Other:
8. Other:

Table 3.3 Speaking and listening

1. Methods I use to assess oral skills (✓)	Assessment Methods	2. Rank
	1. Oral reading/dictation 2. Oral interviews/dialogues 3. Oral discussion with each student 4. Oral presentations 5. Public speaking 6. Teacher-made tests asking students to a. give oral directions b. follow directions given orally c. provide an oral description of an event or object d. prepare summaries of what is heard e. answer multiple-choice test items following a listening passage f. take notes g. retell a story after listening to a recorded passage 7. Student portfolio 8. Peer-assessment	

9. Self-assessment
 10. Standardized speaking test
 11. Standardized listening tests
 12. Other:
 13. Other:
-

Part 2

Continue to reflect on the assessment tools you use in your own classroom by answering the following questions. The questions below ask you to re-examine your choices of assessment tools.

Points to Remember

- The better you understand the benefits and limitations of your assessment tools, the more likely you are to increase the validity and reliability of your assessment practices.

For example, you could choose to use the assessment tools that provide your students with the most useful feedback. Or, you could use those assessment tools that seem to best motivate your students.

- Remember your students are going to be motivated in different ways based on their own learning characteristics. This will require you as a teacher to use different assessment tools.
- Given your classroom context, you will also need to consider the feasibility of your assessment practices.

For example, the design of multiple-choice items takes a long time, but they are relatively easier to mark. Essay questions are relatively easy to design; yet they take a long time to mark and provide feedback. So you will need to balance the use of these tools at different stages of your instruction.

- In the end, you will want to use as many direct assessment tools as you can – for example, oral presentations to assess students' oral ability. However, such assessment takes a great deal of class time. You will need to balance performance time with instructional time.

Keeping the above in mind of recognizing the need to make some tough assessment choices in your own classroom, look at your top two choices in Tables 3.1 to 3.3. Answer the following questions. If possible, discuss your responses with other colleagues. In your view,

1. Which assessment tools yield useful feedback to students?
2. Which assessment tools are most likely to be motivating to students?
3. Which tools are easier to design and/or to score (feasibility of assessment)?
4. Which assessment tools provide the most *direct* information about a student's performance without interference by confounding factors?

3.3.3 Alternative Assessment Methods: A Closer Look at Portfolio Assessment

You may have noticed that in Tables 3.1 to 3.3 *student portfolios* are listed as an assessment alternative. Portfolio assessment has become pervasive in language teaching contexts and is often associated with educational reform or renewal (e.g., Fox, 2014; Hargreaves, Earl and Schmidt, 2002). Although there are many different definitions of portfolios, arguably one of the most widely used is Paulson, Paulson and Meyer's definition (1991). It is not the most up to date, but it reflects commonly shared purposes and goals in language teaching and learning:

A portfolio is a purposeful collection of student work that exhibits the student's efforts, progress, and achievement in one or more areas. The collection must include student participation in selecting contents, the criteria for selection, the criteria for judging merit, and evidence of student reflection. (p. 60)

The portfolio is widely used as a classroom assessment procedure in language arts, music, maths and many other disciplines. In the teaching of English as a second language (ESL), a portfolio is one of the best assessment tools for documenting a student's language development over time (see Fox, 2014; Fox and Hartwick, 2011; Little, 2009). Portfolios 'provide a way to collect and present a variety of performance data, creating a rich and comprehensive portrayal of each student's accomplishments' (Carpenter and Ray, 1995, p. 34). Specifically, a portfolio provides a place (whether a folder, a notebook, a binder, or a file) where students, in collaboration with their teachers and peers, can place evidence of their work in a language course or courses over time. The portfolio provides a site for collecting evidence of a student's learning.

We have also witnessed a trend in the use of e-portfolios – digital spaces made accessible through advances in technology. Given the prevalence of portfolios in language teaching and learning, in this section we'll take a closer look at this assessment alternative and discuss why it has become so frequently used in language teaching contexts.

The starting point for our discussion of portfolio assessment is *purpose*: 'Without a clear understanding of purpose, portfolios are likely to be indistinguishable from unorganized collections of materials' (Linn and Gronlund, 2000, p. 292). In general, two purposes have been identified for creating portfolios of student work: (1) for keeping track of what a student knows and can do (i.e., for evidence of achievement at the end of a unit, project, or course); and (2) for evidence of ongoing learning, growth, or development over time. Some (e.g., Fox, 2014) use the terms *showcase* portfolio (which features finished

products or outcomes) and *working* portfolio (which collects ongoing evidence of work undertaken over time) as labels for these two purposes. However, a number of other classifications have been identified for types of portfolios. For example, Herman, Gearhart and Ashbacher (1996), use the term, *progress portfolio* to refer to collections of work-in-progress (i.e., successive drafts, checklists, conference records, reflective journals), which taken together demonstrate growth and achievement during a course. We argue that because ongoing learning is an integral part of achievement, these two purposes for portfolio assessment should not be mutually exclusive; rather they are of greatest benefit when they work together.

Portfolios (and portfolio assessment) may be used to document and support development in language skills – listening, reading, speaking and writing – but also to develop students' self-awareness, goal-setting, responsibility for personal learning and autonomy (e.g., Little, 2005). Other outcomes, such as increased intercultural awareness, may also be developed through portfolio use – see, for example, Little's research (2009) on the use of the English Language Portfolio (ELP) which is explicitly linked to the values and goals of the European Union (EU) and the Common European Framework of Reference (CEFR).

In general, proponents of portfolio assessment suggest they are particularly useful for:

- monitoring how students manage tasks over time;
- reviewing student development and performance;
- examining the nature of different tasks and/or distinguishing situations in which students are most or least successful;
- assessing performance; and
- developing students' and teachers' insights into second or foreign language learning and the activities that are the most effective in promoting learning.

In sum, proponents argue that portfolios provide students who are learning a new language a physical record of their increasing language development over time, a stimulus for increasing

self-reflection and self-awareness and a means of encouraging personal goal-setting and autonomy. At the same time, portfolios provide clear evidence of a student's achievement in a course.

When portfolios are used in teaching, it is important to note that their strength depends on students' involvement in the selection of portfolio contents. This means that teachers need to provide specific guidelines, based on the purposes of the portfolio and the learning goals or outcomes they have identified for the language class. Once the guidelines are defined, and with the support and guidance of the teacher, students can begin to select, collect and reflect on the work they choose for inclusion in each section of the portfolio.

What should portfolio guidelines include?

Guidelines for a portfolio begin with the intended learning outcomes for a course. As we have discussed earlier in this book, the guidelines for a portfolio should be aligned with and reflect the learning outcomes. Guidelines are typically communicated to students by defining the sections that are required for portfolio development. For example, look at the requirements spelled out by Marta, an EAP teacher of academic writing. Marta's portfolio guidelines were developed for a group of intermediate-level students, who plan on studying in an English-medium university as soon as they can pass a required proficiency test. She begins by defining her learning outcomes. Notice how the portfolio requirements that she lists here are aligned with the learning outcomes.

Portfolio assessment can be an attractive option for teaching and learning language. Marta's guidelines provide an example of how portfolio assessment can work for both formative and summative purposes. Although Marta spells out the required sections for the portfolio, it is her students' responsibility to collect and choose the work that will be included as evidence of their learning during the course. Their self-reflections, logs and the selection process itself provide exceptional teaching opportunities.

Instructor: Marta Ruiz

Class 4B: Writing for Academic Purposes

Focus: Improving academic writing; preparing for writing sections of high-stakes proficiency tests

Level: Intermediate to advanced

Duration: 12 weeks

Learning outcomes

- By the end of this 12-week course, students will be able to write a short academic essay under pressure of time, similar to essays on high-stakes proficiency tests.
- Because students will reflect on their writing development, as evidenced by work collected in each section of the portfolio over time (see below), students will be able to define appropriate personal goals for their writing and self-assess their progress.
- As an outcome of the course, students will be able to review, synthesize and critically evaluate information in writing both personal and argumentative essays, like those required on high-stakes proficiency tests.
- As a result of our work during the course, students will be able to identify and cite academic sources using academic writing conventions.

In order to successfully complete this course, ALL STUDENTS MUST DEVELOP A PORTFOLIO OF THEIR WORK.

Please bring a three-hole binder to class; section dividers will be provided at the beginning of the course. Label each section of your binder/portfolio where you will collect samples of your work.

These are the required sections

1. *Table of contents*: Include a title for each section and a list of each piece of writing included in the portfolio. Please date

your writing and include page numbers. (The Table of Contents will be updated at intervals during the course – just before handing in the portfolio for evaluation at mid-term and at the end of the course). (10% of your final mark.)

2. *Initial diagnostic and self-evaluation:* Include your initial results on the diagnostic test of writing, which identified strengths and weaknesses in your writing at the beginning of the course. Also included in this section should be your own written reflection on what the diagnostic results suggest about your writing. Do you agree with the results? What do you want to work on? What priorities will you set for your writing? What specific issues in writing do you hope to improve as an outcome of the course? (NOTE: At the mid-point and end of the course we will repeat the diagnostic test to see how you are doing. These results should be added to this section along with your written reflections about what is improving and what is not.) (15% of your final mark.)
3. *Reading logs:* Write and include one log each month (a total of three), which summarizes what you have read from newspapers, journal articles, short stories, novels, etc. Use academic citation conventions in your summaries. The topics will be negotiated in class based on your academic interests and the logs will reflect the types of responses required by high-stakes proficiency tests. We will discuss this further in class. (10% of your final mark.)
4. *Personal expressive writing:* Over the 12-week course, class time will be provided for the writing of: (1) a personal narrative, focusing on events in your own life; and (2) a personal essay, focusing on a central idea supported by your life experience. These will be handed in for marking and follow-up feedback on your writing development. Both will be written in class during a strict 30-minute time limit. (20% of your final mark.)

5. *Academic writing on tests:* At the end of the ninth week of our course (6 March), 60 minutes of class time will be provided for the writing of one composition of about 250 words on a set topic. Reading resources will be provided in class. Evidence drawn from these resources, including correct academic citations, is required. You will hand in your composition for marking and follow-up feedback will be provided on your writing development and included in your portfolio. (20% of your final mark.)
6. *Practice tests and reflections:* Throughout the course, we will have practice tests, which simulate the writing requirements of high-stakes proficiency tests. The practice tests will duplicate (insofar as it is possible), the prompts, timing requirements and conditions imposed on test-takers in live test situations. Your practice tests will be marked, and feedback will be provided and discussed. You need to include all practice tests in your portfolio. In addition, you should write a short reflection on your performance on each practice test by answering the following questions in writing: What do you notice? What is changing in your writing? What are your strengths and weaknesses? What do you need to work on? (20% of your final mark.)
7. *My best writing during the course:* Select the best piece of writing you produced during the course. In a short reflection, explain why you selected it. (5% of your final mark.)

NOTE: In order to get full credit for your portfolio (which constitutes 100% of your mark in this course), all sections must be completed as required. We will discuss these requirements further in class.

When students reflect on their learning or explain why they have chosen specific work for inclusion in the portfolio as evidence of their learning, they are articulating their

Recognizing both the benefits and challenges of portfolio assessment

It is important to note that while the potential of portfolio assessment is well recognized, the challenges are also a matter of record.

- *Paper and binder- or folder-based portfolios can be bulky and difficult to manage*; there are issues of who should keep them (i.e., should they be stored in the school for safe keeping or should students be responsible for them). If they are to be stored in the school, what space is available? How secure is it?
- Some of these challenges may be addressed by the use of *e-portfolios* if we are teaching in contexts that have rich technological support. It is now possible to store scanned digital documents in a student's e-portfolio for the duration of a course or, in some cases, over years of students' development in a programme. E-portfolios are more convenient to use, increase access, can use systematic and attractive formats, and store a vast amount of data that can be easily updated and reviewed.
- E-portfolios are not without their disadvantages, however (Savin-Baden, 2008). Key among these is *'student buy-in' and motivation*. Teachers must train their students on how to use the technological resource and develop their students' understanding of its potential. Further, there must be technical support for teachers and adequate controls in place to insure confidentiality and the security of online work.
- *Teachers may be challenged in controlling or managing the communication and activity that is required for a portfolio approach*. Teachers have often reported that portfolios require too much time – and lamented, in some cases, that they can *trap* a teacher in a *portfolio prison* (Hargreaves, Earle and Schmidt, 2002).
- *Teachers may feel they need to work harder to use the approach*. Because this assessment strategy may be new to many students, teachers need to actively support their students' understanding of the approach by being explicit as to why it is being used, the benefits to students, and the necessity of engaging in the work required for the portfolio *over the entire period of the course*. Guidelines for the portfolio must be spelled out clearly (as Marta's guidelines demonstrate).

- *When students misunderstand the purpose of a portfolio approach, they may undermine its impact.* Nothing is more frustrating to students and teachers alike than preparing a portfolio just before it is to be handed in for marking; or, subverting intended learning by writing a finished essay and then inventing two drafts of the essay, after the fact, in order to meet a requirement. When students leave the work required for a portfolio to the last minute, the potential of a portfolio approach is completely undercut. Therefore, it is essential that teachers build in ongoing checks (as Marta has) to monitor portfolio development over time.
- *One of the most challenging issues in portfolio assessment is how to mark them.* They provide an opportunity for students to be involved in both self-evaluation and peer assessment (as Marta's guidelines suggest). Rolheiser, Bower and Stevahn (2000) claimed that 'self-evaluation is key to ownership of learning and the continuous improvement of professional practice. In particular, positive self-evaluation can encourage you to set higher goals and to continue to devote personal effort toward achieving those goals' (p. 124). By reflecting on their own learning, students begin to identify the strengths and weaknesses in their work. These weaknesses then become improvement goals. Furthermore, self-evaluation could serve as a summary reflection, which aims to review each student's goals, identify how students grow as a result of the portfolio practice, and articulate future goals (Rolheiser, Bower and Stevahn, 2000).

Here is an example of a self-evaluation form:

Self-evaluation of writing

1. What is the strength of this piece?
 2. What is the weakness of this piece?
 3. What did I learn while writing this piece?
 4. What would I do differently if I were to write this again?
 5. What was the most difficult aspect of writing this piece?
 6. How would I rank this piece on a scale of 1 to 5 (5 is the highest)?
 7. Who did I ask to read this writing?
 8. What suggestions did he or she make? Were they helpful or not? Why?
-

The portfolio also provides a context for the development of peer assessment. Peer assessment, as a form of getting feedback from classmates, may serve as another benefit to develop students' critical thinking – an essential part of academic studies at all levels of education. Here is an example of a peer-assessment form, which elicits a student's responses to another student's writing.

Responding to peer writing

Student: _____ Title of writing: _____

Reader: _____ Date: _____

1. What is the strength of this piece?
 2. Does the beginning attract your attention?
 3. Does the writer provide evidence to support what is claimed?
 4. Is the supporting detail effective to support the writer's point?
 5. Are there any parts you had difficulty understanding?
 6. What would you suggest that would improve the writing?
-

Finally, there is the teacher's evaluation, which, together with students' self-assessment and peer-assessment, would be effective in promoting improvements in students' work. In the case of Marta's guidelines, she developed and used a checklist for evaluation of the portfolio requirements. For essay marking, she applied the following **rating scale** or scoring rubric, referred to it and discussed it throughout the course with her students. This, as we have noted in earlier Chapters, is essential to high-quality assessment. For example, with the permission of some of her former students, Marta brought in samples of their writing and asked her current students to use the rating scale for evaluation. Their ratings of the samples were discussed in the class and supported students' understanding of the criteria, which also helped them reach their learning goals.

Criteria for evaluating your writing

Excellent (5):

- Focus:** ♦ Says something new about the topic (insight)
 ♦ Remains on topic throughout
 ♦ States main idea and three supporting ideas in the introduction
 ♦ Relates conclusion directly to the main idea
- Support:** Examples, reasons and explanations are relevant, accurate, convincing, sufficient (but concise), and specific
- Organization:** ♦ Has effective introduction, body and conclusion
 ♦ Has unified paragraphs with topic, supporting, concluding sentences
 ♦ Paragraphs flow from one to the next and sentences are linked within the paragraph
- Style:** ♦ Excellent sentence variety
 ♦ Excellent vocabulary: varied, accurate
 ♦ Formal level of language
- Mechanics:** ♦ No major errors
 ♦ Two or three minor errors

Well developed (4):

- Focus:** ♦ Says something about the topic (insight)
 ♦ Remains on topic throughout
 ♦ States main idea and three supporting ideas in the introduction
 ♦ Relates conclusion only vaguely to the main idea
- Support:** Examples, reasons and explanations are relevant (throughout), accurate, reasonably convincing and reasonably sufficient
- Organization:** ♦ Has introduction, body and conclusion
 ♦ Paragraphs are unified with topic, supporting, concluding sentences

- ◆ Some body paragraphs do not flow into the next paragraph or do not have linked sentences within
- Style:**
 - ◆ Good sentence variety
 - ◆ Good vocabulary: varied, mostly accurate
 - ◆ Formal level of language
- Mechanics:**
 - ◆ One or two major errors
 - ◆ No more than three minor errors

Acceptable (3):

- Focus:**
 - ◆ Remains on topic
 - ◆ States main idea only indirectly
 - ◆ Relates three supporting ideas only adequately
- Support:** Examples, reasons and explanations are partially relevant, appropriate, primarily accurate and developed unsatisfactorily
- Organization:**
 - ◆ Has introduction, body and conclusion
 - ◆ Has topic sentence and some supporting sentences
 - ◆ Some attempt to connect paragraphs and to make connections within the paragraph
- Style:**
 - ◆ Attempted sentence variety
 - ◆ Attempted variety and accuracy in vocabulary
 - ◆ Formal level of language generally
- Mechanics:**
 - ◆ No more than three major errors
 - ◆ Excessive minor errors

Partially developed (2):

- Focus:**
 - ◆ Remains partially focused on topic
 - ◆ Development is inadequate or ineffective; ideas are incomplete
- Support:** Examples, reasons and explanations are somewhat relevant, repetitious, generally inaccurate and undeveloped

- Organization:** ♦ Attempts introduction
 ♦ Weak body paragraph
 ♦ Attempts conclusion
- Style:** ♦ Inappropriate (i.e., colloquial or slang mixed with formal language)
 ♦ Weak vocabulary
- Mechanics:** ♦ Four or five major errors
 ♦ Excessive minor errors

Undeveloped, unclear (1):

- Focus:** ♦ Unable to relate ideas to topic; superficial
 ♦ Develops ideas randomly, disjointedly
- Support:** Examples, reasons and explanations are vaguely relevant, very repetitious, mainly inaccurate, unconvincing and illogical
- Organization:** ♦ No introduction
 ♦ Minimal evidence of or no paragraphing
 ♦ No conclusion
- Style:** ♦ Level of language not formal
 ♦ Vague vocabulary
- Mechanics:** ♦ More than five major errors
 ♦ Numerous minor errors

Communicating results

At the end of the term, teachers can take advantage of the completed portfolios to communicate among students in the classroom and to parents or guardians outside the classroom. In a student-led conference or at a class meeting, all the portfolios can be displayed to provide an opportunity for students to share their portfolios with their classmates and to learn from each other. In a parent-teacher conference, portfolios offer an excellent means for parents to enter the classroom experience by reviewing and reading their children's portfolios. This gives them a more intimate basis for seeing aspects of their children's experiences in school; provides a framework for meaningful discussions of the students' achievements,

progress, and areas to work on next (Linn and Gronlund, 2000, p. 311); and invites stakeholders be more actively involved in enhancing their education.

Planning for portfolio assessment

If you are planning to use portfolio assessment in one of your language classes, you may want to begin by considering each of the following questions. You can use these questions as a checklist for your preparation for portfolio assessment. Notice how Marta's guidelines answer each of these questions.

What is the purpose of the course?

Who will be taking it?

Why are students taking this course?

Are there any constraints? (e.g., curriculum, standards, textbook, external tests)

What are my learning outcomes?

What should I assess as evidence of learning?

When and how often should I assess learning?

When and how often should my students self-assess their learning?

What are my guidelines or requirements for the portfolio (sections, contents, timing and so on)?

How much value will I place on each requirement?

If you wish to read about the journey of a teacher using a portfolio in her writing course, see Christine Armstrong's account (2006). Summarizing her action research on the use of the portfolio, she writes in her thesis abstract:

The writing portfolio involved multiple processes, including writing numerous articles with multiple drafts which were all corrected, first by the student, then by a peer, and finally by me, in conjunction with completing a questionnaire eliciting students' self-reflection about their writing process. Students who participated fully in the combined use of these elements improved their written French, and showed evidence of an increase in their learner responsibility. Those who did not participate did not show evidence of notable improvement.

3.3.4 The Complexity of Assessment Practices

As the questions above illustrate, teachers' assessment decisions, events, tools and procedures are extraordinarily complex. These complex practices are shaped by considerations of:

- *context* (adults or children; EFL or ESL; and so on);
- *purpose* (student-centred, instruction-centred and administration-related);
- *method* (instructor-made, student-made and standardized/external testing of reading, writing, speaking and listening); and
- *procedure* (sources of assessment, feedback and reporting, the time spent on assessment, its value or weight and so on).

We take each of the above into account in developing our assessment plans. Instruction and assessment are interwoven and influence each other in the day-to-day decisions we teachers make in classroom teaching. This is an ongoing process and a complex endeavour involving complex decision-making.

There are guidelines that we can consult to inform valid, fair and ethical assessment practices. See, for example:

- *The Code for Fair Testing Practices for Education* (<http://www.apa.org/science/programs/testing/fair-testing.pdf>);
- *Principles for Fair Student Assessment Practices for Education in Canada* (http://www2.education.ualberta.ca/educ/psych/crame/files/eng_prin.pdf);
- *Standards for Teacher Competence in Educational Assessment of Students* (<http://buros.org/standards-teacher-competence-educational-assessment-students>).

For more information on standards in language test development and use, see the website of the International Language Testing Association (ILTA) at www.iltaonline.com or, better yet, join ILTA and participate in their efforts to improve the quality of language testing around the world. ILTA has published a *Code of Practice*, which defines the requirements for ethical, fair and valid language testing practices. It is a useful document to consult when you have questions about the ethics and fairness of testing.

ACTIVITY 3.5

In reviewing this chapter, challenge yourself further by exploring the following questions. These questions will help you to rethink the above complex relationship between assessment purposes, assessment methods and assessment procedures. These questions also push you to revisit the fundamental aspects of assessment – for example, the *use* and *interpretation* of our assessment information. It is the use and interpretation of assessment that have the greatest impact on our students.

1. Who are our assessment users? Who are the primary users of the data or information from our assessment methods?
 - Teachers
 - Students
 - Parents
 - Schools, universities, or colleges
 - Funding agencies
 - Government
 - ...
 - ...
 - ...
2. Who, in the above categories, makes decisions based on the information?
3. Are they users, decision-makers, or the object/subject of the decision?
4. What will happen when the ‘subjects’ of an assessment decision, our students, have no say or do not use any assessment tools? Why should we involve our students as much as possible in assessment practices?

There are also regional and national associations, which are concerned with fair and ethical testing practices, such as the European Association of Language Testing and Assessment (EALTA), or the Canadian Association of Language Assessment (Association canadienne pour l'évaluation des langues).

Is there a local or regional association in your own area? You may want to do some research to identify whether a professional assessment and testing association exists. Such associations bring together teachers, researchers and testers to consider issues in testing that impact teaching and learning and improve the consequences of such tests through positive dialogue and the exchange of information and ideas.

Whether we are teaching adults or young English Language Learners (ELLs), assessment is the cornerstone of our teaching practice. As Gottlieb (2006) puts it:

As educators, we are constantly challenged to make informed decisions about our students; to do so, we plan, gather, and analyze information from multiple sources over time so that the results are meaningful to teaching and learning. That's the core of the assessment process and the centrepiece in the education of linguistically and culturally diverse students. If reliable, valid, and fair for our students, assessment can be the bridge to educational equity. (p. 1)

3.4 Looking Back at Chapter 3

In our discussion of assessment plans in Chapter 3, it would have been helpful to understand the processes and practices of test development and how such tests work in our classrooms. Knowing more about test development would enable us to judge the quality of the tests we create in our classrooms as well as the external tests that impact our students. In [Chapter 4](#) we examine how a high-quality classroom test is developed – step by step – and how engaging in such test development processes in our own courses and programmes

can improve the quality of teaching and learning through assessment.

Suggested Readings

Cheng, L., Rogers, T. & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education*, 33(1), 9–32.

This comparative interview study was conducted in a range of three ESL/EFL university contexts in Canada, Hong Kong and China. Six major aspects of ESL/EFL classroom assessment practices were explored: instructors' assessment planning for the courses they taught; the relative weight given to course work and tests in their instruction; the type of assessment methods (selection vs. supply methods) that they used; the purposes each assessment was used for; the source of each method used; and when they used each method.

Fox, J. (2014). Portfolio based language assessment (PBLA) in Canadian immigrant language training: Have we got it wrong? *Contact, Special Research Symposium Issue*, 40(2), 68–83.

Fox examines the implementation of a portfolio assessment approach in the context of a national language training programme for newly arrived immigrants and refugees in Canada. She argues in favour of the formative purposes for portfolio assessment and suggests that the government's emphasis on the use of portfolios for summative purposes may undermine learning potential.

Gottlieb, M. (2006). *Assessing English language learners: Bridges from language proficiency to academic achievement*. Thousand Oaks, CA: Corwin Publishing.

A useful overview of assessment approaches and techniques. Teachers will appreciate the many suggestions, examples and materials that are provided by Gottlieb to support teachers as they monitor, provide feedback and document learner development. Although the book is directed at ELLs in schools, the information is appropriate for language teachers at any level.

Savin-Badin, M. (2008). *Learning spaces: Creating opportunities for knowledge creation in academic life*. New York: Open University Press.

Savin-Badin redefines the notion of learning spaces in her consideration of beyond-the-classroom approaches to teaching and learning. She stimulates our thinking about the boundaries of educational time and space in her discussion of, for example, reflective spaces, writing spaces and digital spaces – all of which extend learning potential.

How Do We Develop a High-Quality Classroom Test?

Activate your learning

- How do we develop a test?
- How do we analyse a test?
- What should we look for in evaluating the quality of a test?
- Why is it important to understand how to develop a high-quality test?

4.1 Developing a Test

It should be noted, that there is no one way to design a test, but there are commonly agreed standards that should apply to all testing activities, whether they occur within your classroom as a result of your own test development, across a programme, a system, a nation, or around the world. In the sections below, we will ‘walk through’ a process of designing a test.

First, we will consider some of the key steps in developing a test. The higher the stakes of the test, the more time and effort will go into work on each of these steps. However, as teachers, we also need to clearly understand the importance of each step and do our best, working alone or with our colleagues, to insure that we have designed a test that

- measures what we intend it to measure;
- adequately represents or samples the outcomes, content, skills, abilities, or knowledge we are measuring; and
- elicits information that is useful in informing our teaching and in supporting the learning of our students.

The above are three of the key standards we apply in judging the quality of a test.

A lot of advance planning goes into the development of a test that will meet the standards listed above. From the outset, it is helpful to consider a test as an operational definition of what we intend to measure. In other words, when we **operationalize** an outcome, concept, competence, skill, or ability we translate it into what we can actually measure. For example, an airline could announce that ‘passengers with heavy bags will not be allowed to bring them on board the airplane’. But, what is heavy to one passenger is not heavy to another. In order to operationalize the concept or idea of *heavy*, the airline will advise passengers that ‘bags weighing more than 15 kilograms/33 pounds, will not be allowed on board’. The operational definition of ‘heavy’ is 15 kg/33 lbs. A test makes operational the learning that we are working towards in our classrooms by translating what we define as learning into what we can measure in a test.

Consider another example. In a course that we are teaching we have the goal or learning outcome of supporting our students’ reading comprehension by developing their ability to define new words through the use of contextual cues. If this is our goal we need to consider not only the learning activities that will support the development of this skill, but also how we will measure their learning in operational terms. Although we can exercise many options in this regard (as we noted in [Chapter 3](#)), one of the principal means of measuring learning is through testing. Further, it is important to understand that the more we know about what each item and task is measuring in a test, the more useful, meaningful and interpretable our test will be.

4.2 Key Terms and Concepts in Test Development

Before examining the steps in a test development process that will improve the usefulness and meaningfulness of our tests, there are a few key terms we need to define. These terms are central to the discussion which follows below.

4.2.1 Construct Definition

Once the mandate for a test has been established (or we have decided that we want to use a test as our assessment tool) and the purpose of the test is clearly understood and stated (e.g., to place students at the appropriate level of an ESL programme; to admit students to study in a university; to award a certificate of mastery in a trade or skill; or to determine the degree of achievement attained by individual students in a class), we need to define precisely what we intend to measure.

Construct definition is the process of defining what it is we intend to measure. A construct is most often theoretically or empirically informed (i.e., by research of the context in which language is used); it may be identified in a curricular document, which spells out goals for learning; or defined by the learning outcomes that have been identified for a course (as we discussed in [Chapter 2](#)).

For example: we are teaching English to a group of adult learners who are studying in order to pass a high-stakes test of proficiency and enter a university degree programme. What in your view are the language abilities, performances and skills that should be tested in order to determine whether your students have sufficient English to enrol in a university academic programme – whether in engineering, business, history, or science? Think about this question in responding to Activity 4.1 below.

4.2.2 Criterion- and Norm-Referenced Assessment

How we define the construct of our test will influence all other aspects of our test. Are we guided by criterion-referenced definitions, such as those available to us in the Common European Framework of Reference (CEFR) or the Canadian Language Benchmarks (CLB)? Here is an example of a criterion-referenced statement:

Skill: Reading

Sample Criterion Descriptor: Students at level 3 will be able to guess the meaning of an unfamiliar word by using cues provided by surrounding text.

ACTIVITY 4.1

Practice in Construct Definition:

Take a minute and write a statement of *one* key skill, ability, or performance that you think is an important component of what should be measured in this context:

It is important to note that *what* we intend to measure (the construct), does not typically emerge only on the basis of self-reflection.

If the above criterion is a defined learning outcome of a course we are teaching, we need to develop items and tasks that will measure whether our students have reached (or exceeded) level 3. The main purpose of criterion-referenced assessment is to *describe* what students know and can do.

At times the purpose of a test or assessment procedure is to compare a student's performance with other students, to rank a student's performance on the basis of the test performance of their peers – from high to low. Such a test or assessment procedure is called **norm-referenced assessment**. Its main purpose is to *discriminate* between students.

4.2.3 Target Language Use (TLU) Domain

In many settings we are interested in developing our students' proficiency, performance, or skill without specific reference to a context of use. Nonetheless, we know that language changes in relation to the context in which we use it. The words we might guess from the context while reading the sports page in a newspaper, an item on a restaurant menu, or a form at a

bank or hospital will differ in relation to their contexts. So, when we interpret a criterion we may want to add some context to an item or task in our test.

If, however, we want to limit our test to a specific context, we will be engaging in domain-referenced assessment or the assessment of language for specific purposes (Douglas, 2000), and only the language that is typically used, and tasks that routinely occur within that domain, will be selected for use in our test. So, for example, if a proficiency test is developed in order to measure the language necessary for an undergraduate student to engage with the demands of a first-year engineering programme, all of our criteria will be domain-referenced. In this case, the **target language use (TLU) domain** – the domain of interest – is first-year engineering. We would not, for example, use a poem to test reading comprehension, but we might ask our test-takers to follow instructions from a lab assignment or interpret a graph.

ACTIVITY 4.2

Practice in Refining a Construct to reflect a TLU Domain

Part One

Look back at the construct you defined above in Activity 4.1. If you can, compare what you have written with that of a colleague.

Identify your own ‘personal philosophy’ or your colleague’s in the statement of construct? Adjust the statement of construct that you wrote above (and that of your colleague’s, if relevant) to reflect a specific context by adding a TLU domain, such as English for travel guides, English for business, English for teachers of English as a Second or Foreign Language (TESL/TEFL). Revise the construct to reflect the TLU domain:

Part Two

A typical issue that arises when we are in the process of defining a construct is the issue of specificity (Davidson and Lynch, 2002). In the activity below, we look at how some professional tests define their constructs.

Search the Internet for the website that describes a high-stakes language proficiency test (e.g., the Test of English as a Foreign Language Internet-based test (TOEFL iBT); the International English Language Testing System (IELTS), or another test of your choosing).

1. What does the website say the test is measuring? This is a key statement (or series of statements) which we can use to judge the quality of the test.
2. Given the purpose of the test, how adequate is the construct definition in your view?
3. How does the construct that you wrote above compare with that of the high-stakes proficiency test? Remember, the construct that you wrote was limited to a single key skill, ability, or performance. Proficiency tests such as TOEFL iBT or IELTS have much broader constructs. Do you see the difference? Would you add to your construct based on what you found on the Internet? If yes, what would you add? If no, why not?

4.3 The Important Role of Test Specifications

An essential question we must answer when developing a test is whether or not our statement of construct is too broad, too narrow, or just right. This leads us to the definition of another key term in high-quality test development: **test specifications**.

When we write out information about what our test is measuring we are engaged in test specification writing. This is a map, a blueprint, or a recipe for our test. Our specifications should be

detailed enough to allow another teacher or group of teachers to develop a test that is similar to ours. Test specifications let teachers and test developers develop multiple versions of a test, which are arguably parallel or as similar as possible. Writing test specifications for a test you are developing will allow you to:

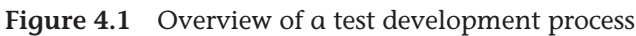
- generate other tests for other classes at equivalent levels of difficulty;
- accumulate systematic information about what you are testing and what your students are learning;
- document over time the relative performance of groups in your programme; and
- increase the clarity of your purpose and directly support the link between testing and teaching.

Test specifications are not fixed. They evolve and over time they tend to stabilize. Developing test specifications can be a very helpful means of encouraging collaboration and cohesion within a programme as participating teachers map out and define test specifications as part of a test development project. In other words, the process of negotiating specifications for a test is an excellent professional development opportunity.

As we discussed in [Chapter 3](#), tests are only one of the many assessment tools that we, as teachers, use in our classrooms to evaluate learning. When we do decide to test, however, we need to be systematic in our test development process. This will take time and advance planning. Otherwise, there are many other more useful tools that will provide us with better information about how much our students have learned at any particular point in time.

4.4 An Overview of a Test Development Process

As discussed, tests are one of the many assessment tools we have at our disposal. How do we build quality into our tests? We need to look at test development as a process ([Figure 4.1](#)).



Throughout a test development process, at every step, we collect evidence of how each element of the new test (i.e., items, tasks, scales) is working, evaluate that evidence and revise accordingly. This is not only true during initial development, but is an ongoing requirement during the life of the test. This ongoing collection of evidence of test function and use is the essential requirement for *validity*. The process itself is referred to as *test validation*.

4.4.1 The Mandate

As Figure 4.1 illustrates, the *mandate* ‘to test or not to test’ is the starting point for the development process. The mandate

for a new test is the determination that it is the most appropriate or meaningful assessment alternative given the purpose of the assessment. The mandate may arise internally – you may want to administer a test because you feel it is the best way to evaluate what your students have learned at a particular point in the programme. Or, your colleagues in your programme – your coordinator or programme director – may feel that a test is essential in order to compare the learning across classes, groups, or levels in a programme in order to gain evidence that the programme is working, or to more effectively place students in appropriate levels. Alternatively, the mandate may arise *externally*. It may be prompted by interested stakeholders who are not working within your programme – by ministries or departments of education, by government agencies who fund your programme, or by other groups external to your programme (e.g., parents who want evidence of how their children are doing; or owners of schools who want tangible evidence that the programme works, so that they can market it to new customers). The mandate is important because it shapes in fundamental ways how the test will be designed.

4.4.2 The Purpose

The mandate motivates the purpose of the test and provides parameters for the definition of useful constructs in the test. In defining the constructs of interest in the test, we need to answer the question: what do we need to measure in order to provide evidence of learning that addresses the mandate and purpose of the test?

4.4.3 The Construct(s)

As we discussed above, the next step is key and complex. We need to define the construct or constructs that we want to measure in our test. What specific learning do we intend to measure? How is this learning construct informed by theory and research?

We need to begin by stating in clear terms what we intend to measure. Look back at the statement that you wrote in Activity 4.1 that defined one skill, ability or performance that an English for Academic Purposes (EAP) student entering university should demonstrate in order to engage in a first-year undergraduate degree programme. This is a statement of construct. The sample criterion descriptor is also a statement of construct. Our focus in test development is to translate the constructs we intend to measure into operational terms we can measure, namely items and tasks on the test.

4.4.4 The Table of Specifications

Tables 4.1 and 4.2 detail the overall features of a test to provide you with an example of how specifications are used in practice. You may want to use these tables when you engage in developing a test for the next class you teach. They are offered as a model only, as there are many different ways in which we can present overall information about a test. It is fair to say, however, that developing a Table of Specifications for a test is the key feature of high-quality testing. Engaging in the process of planning your next test by using a Table of Specifications will improve the quality of your test, increase the clarity of the information the test provides, and enhance the validity of the inferences you draw on the basis of your students' performances in the test.

Once we have identified the construct or constructs of interest, we begin to develop the Table of Specifications, which is an overall blueprint, map, or summary of the test. [Table 4.1](#) (History file) and [Table 4.2](#) (Table of specifications) describe an internally mandated test of reading that was developed by a group of teachers in an EAP programme. Their students were adults who wanted to increase their English proficiency in order to enter a university degree programme. The test's purpose was to evaluate learning at the mid-point of the course.

A Table of Specifications helps us to define in precise terms what we intend to measure. It provides an overview of our operational definition of the construct: what we will elicit as

Table 4.1 History file (excerpt)

Date	Modifications/Comments/Reflections
26/01/2016	<i>(L. Jenkins; M. Khan; R. Roessingh)</i> Eliminated the reading passage at the beginning of the test because it took too much time and did not add to the overall quality of the writing.
10/05/2016	<i>(Janet Martin, test development coordinator)</i> Decided to add another reading section to increase test length. We will add another reading text and 10 multiple-choice questions to test comprehension. Teachers felt the test didn't sufficiently reflect the textbook emphasis on comprehension and their work with students on the exercises at the end of each textbook unit. We'll try out the new reading section at mid-term in 10 weeks. Specifications are under development now. This means we will have to adjust the point totals and increase the amount of time we allow for the test. We are discussing the implications now.

evidence of learning (in performances, tasks, items); how we will sample this learning in relation to time and effort; and how we will value performances, tasks and items. It is a blueprint for the test we are administering now and for future tests.

Because a Table of Specifications will be used not only for a current test, but potentially for future tests, we begin by spelling out the test's history (see [Table 4.1](#)). The **history file** provides an ongoing record of the evolution of the test. It answers such questions as, When was it developed? Who was involved? What have we learned over time about the test? What revisions have been made? Why were these revisions made? It is important to record the revisions and reflections that have influenced the Table of Specifications over time (its history), file them for future use, and add to them over time. Remember, Tables of Specifications are not written in stone. They

Table 4.2 Table of specifications (sample specification)

TEST NAME: Integrated skills: Reading-to-Writing for Academic Purposes (Level 3, Mid-Term)					
TIME: 60 minutes (1 hour)					
TIMING: Mid-term (6 weeks after the beginning of the course)					
VALUE: Accounts for 25% of the final mark					
Skill by section order	Time in minutes	Weight	Number of items/tasks	Item/task type	Resources
1. Summary writing of reading texts (comprehension of main ideas and supporting details)	20	20% (20 points total)	2 tasks	Extended written precis or summary	Two extended texts (400–500 words each). Same topic: one text advocates for a position (is positive); the other argues against it (is negative)
2. Genre Awareness	10	10% (1 point per item)	10 items	Multiple choice	Source texts
3. Guessing vocabulary from context	10	10% (2 points per item)	5 items	Constructed response	Source texts
4. Personal reflection on the topic	10	10%	1 task	Written response to topic	Source texts Prompt: What do you think? Who is right?
	50	Total = 50% ÷ 2 or 25% of the final course mark			

evolve and improve over time and reflect changes in purpose, perspective, curricula and student needs. Keeping a record of these changes will support effective teaching over time if teachers use the Table of Specifications to generate dialogue about teaching and learning within their programmes.

In the Table of Specifications provided in [Table 4.2](#), note the headings: skill, time in minutes, weight, number of items/tasks, item or task type and resources. Tests must always operate within constraints of time and so in our planning there is a trade-off between what we would like to measure and how much time we have. An important rule of thumb, however, is to look for a relationship between the amount of time given for a particular task and the amount of weight it is accorded. In apportioning time and weight, we are operationalizing an important aspect of the construct. Examine the relative amount of time and weight apportioned for this 60-minute test of reading-to-writing in an academic context. Would you divide the time for the test in the same way?

Another rule of thumb to consider is the relationship between tasks or items and points. Keep it simple. Notice in the sample Table of Specifications ([Table 4.2](#)) the relationship between points and the number of items/tasks in the test. For example, 20 minutes are apportioned to summary writing at the beginning of the test, and this section is valued at 20% of the total. The summary writing section is comprised of two tasks each valued at 10 points. So the 20% weighting of the summary writing section equals 20 points on the test. Look at *genre awareness*. Ten minutes are allowed for this section of the test. There are 10 multiple-choice items, accorded a 10% weighting or 1 point per item. This result is 10 points for this section of the test out of the 50 possible total points on the test. Think about what this is saying about the construct. Does the value placed on genre awareness reflect your understanding of the role that genre awareness plays in comprehending a reading passage?

There are three different item or task types on this test's Table of Specifications: extended written response (for the summary writing section); multiple-choice items (for the genre

awareness section); constructed response or short answers (for the guessing vocabulary from the context section); and extended written response (for the personal reflection section). Each time we alter item or task types for a test, we increase the demands on our students. The formats we select for our tests should reflect the types of routinely occurring tasks that we use day-to-day in our classroom learning activities. Again, to get the most useful information, keep the testing activity as similar as possible to classroom activity.

The most challenging element in this Table of Specifications is the identification or development of the source texts (see the 'Resources' column in [Table 4.2](#)). Because the resources for this test have very specific requirements (i.e., one topic, pro and con positions, 400–500 words for each position), it may be necessary to write these texts or to adapt existing texts found through research on a topic.

There are advantages to developing texts for testing purposes rather than attempting to use naturally occurring texts (so-called *authentic texts*) that one might find in textbooks, newspapers, or on the Internet. If test development is viewed as professional development, in the case of the present example, it would be very helpful to have teachers in a programme write responses to pro and con positions on a topic.

For example, as a starting point, we might ask our colleagues to discuss and list advantages and disadvantages of organic farming. Working in small groups, we could then write out a response to the pro and con positions on the topic. After we have written our own responses to the topic, we could meet as a group and negotiate a group response. Since in the present example this is an academic context, we should have access to the Internet at this point to seek out information that further supports the pro and con positions.

Engaging in writing a response to the topic helps us to better understand the demands of the task and supports the later step we will be required to take in coming to terms with what we will value in marking the responses of our students. It also provides a context for discussion of the relative importance of

features of reading and writing and supports self-reflection and awareness. Active participation in test development can improve the overall cohesiveness of a programme as a whole and our students are the end beneficiaries of such cohesiveness. It also lays the foundation for training markers, or raters, who evaluate performances on the test.

Once the Table of Specifications has been developed, we turn our attention to the specification of tasks and items. Some of this specification work is already evident in the overall test blueprint (see ‘Resources’ in [Table 4.2](#)). But we also need to specify the nature of each task and item in our test.

4.4.5 Item and Task-Writing

A Table of Specifications allows us to map out our intended test in relation to time and overall organization. The next critical step is to actually write the items and tasks that will comprise the test itself.

First, it is important to distinguish between an *item* like that found in a multiple-choice test and a *task* like that found in a summary or extended written responses in a test ([Table 4.3](#)). An *item* is often referred to as a discrete measure. Discrete-point items tend to have right or wrong answers (although in some cases partial credit may be awarded), and tend to measure one feature of a construct at a time (e.g., main ideas, vocabulary, supportive details). A *task*, on the other hand, is more complex. It typically involves a performance, such as writing an essay, responding to questions in an interview, role-playing, or reporting on key information in a lecture. When we rate or mark a task, there are a number of different criteria at play, which account for different features of a construct simultaneously within a single task performance. When we rate each item on our test, we mark each item separately as a discrete feature of a construct.

When we choose an item or task format, it is important to understand how each of our decisions will impact the students’ responses to the test. Our choices will also reflect the ongoing

Table 4.3 Some commonly used item and task formats

Item formats	Task formats
<i>Multiple-choice</i> (selecting the right answer from several choices or distractors)	<i>Essays</i>
<i>True/false</i> (identifying the right/wrong answers)	<i>Summaries</i>
<i>Matching</i> (e.g., a word with a picture; synonyms)	<i>Written reports</i>
<i>Ordering</i> (e.g., identifying what happened first, second, third, and so forth in a sequence from beginning to end)	<i>Information gaps</i>
<i>Information transfer</i> (e.g., labelling a graph or picture, based on information provided in a text)	<i>Oral interviews (one-on-one)</i>
	<i>Presentations</i>
	<i>Role plays</i>
	<i>Interactional transfer (gap-filling)</i>
	<i>Group-based interviews</i>
	<i>Integrated tasks (Listening-to-writing; Reading-to-writing)</i>

instructional methods used in our class. Each item or task format has an impact on how a student or test-taker engages with the test. Test developers are aware of this and refer to the impact of the item or the task format on responses as a *method effect*.

ACTIVITY 4.3

The Impact of Test Methods on Teaching

Reflect for a moment on how you would prepare for and respond to a multiple-choice test. How would your preparation and your response differ if you were writing an essay test? Write your response below and then, if possible, compare it with a colleague's.

There are some item formats that have a particularly strong method effect. One of these is the cloze-test format, where words are omitted from a text on a systematic basis (e.g., every fifth, sixth, or seventh word) and students or test-takers are asked to fill in the missing words based on their comprehension of the text as a whole. The method effect of the cloze-test format is discussed by Alderson, Clapham and Wall (2001), who provide a comprehensive overview of item and task formats. We include a list of item and task formats in the Appendix at the end of this book.

A particularly useful guide to writing and using language test specifications is the book *Testcraft: A Teacher's Guide to Writing and Using Language Test Specifications* by Fred Davidson and Brian K. Lynch (2002). In this book, Davidson and Lynch make the point that the quality of a test depends directly on the richness of its items, tasks and test specifications, and demonstrate the many advantages of specification-driven test development. We have adapted the specification components they identify to provide an example of what an *initial specification* might look like for one of the sections in the Integrated Reading-to-Writing Test in Tables 4.1 and 4.2.

This is an initial specification for Section 2, the genre awareness section of the test, because it precedes its administration – hopefully as part of the pilot or trial step in the development process (discussed next). The specification will evolve over time as we collect evidence, revise, and review the test's usefulness over successive test administrations (see [Figure 4.1](#)). Below is a sample specification at the section and item level.

Title: Section 2: Genre Awareness

General Purpose/Description: At the end of Level 3, students should be able to *read between the lines*; to identify where a text(s) would most likely originate; to assess the tone of the author(s) in relation to the views being expressed; and to relate tone to words or phrases in the text(s) which suggest a particular belief, assumption, or attitude.

Prompt: Students will be asked to respond to ten multiple-choice items with four choices each. The item **stem** will pose

a question, which is answerable by only one of the choices. The other three choices are **distractors** – choices which are not correct, but provide plausible alternatives and help to determine the quality of the test-taker's understanding. All questions will be based on the texts provided in Section 1 of the test (see 'Resources' below). There will be four questions related to the first text (which presents a positive view of the topic), and the same questions related to the second text (which presents an opposing or negative view).

The following stems will be used for *each* of the two texts (questions 1–8) in random order:

Text A:

1. Which of the following best describes the author's tone or attitude?
2. Which of the following words or phrases taken from the text most clearly suggest the author's tone?
3. Which one of the following actions or responses would the author most likely agree with?
4. Although it is not stated in the text, what is most likely the author's background?

Text B:

5. Which of the following best describes the author's tone or attitude?
6. Which of the following words or phrases taken from the text most clearly suggest the author's tone?
7. Which one of the following actions or responses would the author most likely agree with?
8. Although it is not stated in the text, what is most likely the author's background?

Both texts:

The following two stems will be used for items 9 and 10 and relate to both texts:

9. Where would you expect to find texts like these?
10. Who is most likely the intended audience?

Of the four distractors, one should be clearly the correct answer; one should be clearly incorrect; the other two distractors should have elements that are correct, but also elements that are not.

Resources: For this section of the test, students need to use the texts provided in Section 1, which express positive (pro) and negative (con) positions on the same topic. Texts may be drawn from a range of contexts (i.e., magazines, newspapers, textbooks, reviews, lab assignments or other reports).

Student Responses: Students will respond by selecting the best answer from four options. Students will be provided with a space next to each item where they can comment if they find the item confusing, unfair, ambiguous and so on. See sample instructions and item below:

Instructions: There are 10 items on this section of the test. Choose one answer for each item by circling the letter that is next to it. If you would like to comment on the item, space has been provided for you in a box on the right-hand margin of the test.

Text A:

1. Which of the following best describes the author's tone or attitude?

- a. angry (*clearly right*)
- b. enthusiastic (*clearly wrong*)
- c. unhappy (*somewhat right, but the overall tone is angry*)
- d. impatient (*somewhat right, but the overall tone is angry*)

Comments:

4.4.6 Scoring and Rating

In the example above, one point was awarded to each of the correct answers. However, teachers note the comments of students and pay particular attention to their concerns regarding

ambiguity, fairness, and so on. If you find these concerns are reasonable, remove the item from consideration in calculating the final mark and make a note on the test specification record so that the item can be revised.

In the sample specification above, the focus is on the specification of items in Section 2 of the test. All of these items use a multiple-choice format with a correct answer identified for each item. How would the specification for scoring differ if we were working on a specification for Section 3? This section tests *guessing vocabulary in context* and calls for *constructed responses* or short written answers generated by the test-takers or students themselves. So, for example, in this section of the test we might find the following text and item:

TEXT. What is a *phobia*?

Every human being feels fear at times. Young children are often afraid of the dark. Adults may feel uneasy and fearful in a thunder and lightning storm. For some, however, fears are excessive, irrational and out of control. When we feel intense fear or have what has been referred to as a *panic attack*, in spite of the fact that there may be little or no reason for such feelings, we may have a *phobia*. A *phobia* is a type of treatable mental illness. For example, people who suffer from the **condition** known as arachnophobia, have an intense and uncontrollable fear of spiders. In extreme cases, even seeing a web or a picture of a spider can cause panic.

ITEM 1.

In Text A, line 9, what does the word *condition* mean?

Note that the Table of Specifications (Table 4.2) indicates that there will be five items like this and that each will be awarded a maximum of two points. This will make the scoring of the answers that our students provide much more complicated, because we will need to consider *partially correct answers*. In the specifications for this section of the test, we would need to help markers or raters (in most cases other teachers) interpret what we mean by partially correct answers so that we all award points in a consistent manner. The specifications can help to insure consistency in our scoring. Insuring consistency increases the *reliability* of our measurement.

Look at the following explanation for scoring responses to Section 3, Item 1 with relation to the definition of the word *condition*.

- Two points are awarded to completely correct answers. For example: Condition refers to a specific phobia, an extreme fear of spiders, or arachnophobia, as a mental illness that is treatable.
- One point is awarded if the information for *at least one* of the above underlined phrases is included in the response.
- Zero points are awarded if none of the information in the underlined phrases is included in the response.
- Do not award half (0.5) marks in this section of the test.

If all of the raters or teachers marking the test systematically follow the guidelines, and the guidelines are specific enough to cover typical responses to the questions on the test, the specifications will support overall reliability. However, the specifications for scoring will evolve in relation to the responses produced by test-takers and are subject to review and revision on an ongoing basis. The process of negotiating the scoring (like all other parts of the test) is an excellent professional development resource. However, short constructed response items like those in Section 3 of the sample test will not yield as rich a discussion as the marking of the extended writing in the summary produced by the students or test-takers in response to Section 1 of the sample test. How to score this section will

ACTIVITY 4.4

Examine the description of Section 1 (summary writing) in the Table of Specifications (Table 4.2). How would you mark this section of the test? What would be key in your evaluation of this reading-to-writing (integrated) task? If you have a colleague nearby, discuss your approaches to evaluating the task. Jot down a few key criteria that would be important in evaluating task completion.

generate a great deal of discussion among teachers – even though they are teaching at the same level (Level 3, in the case of the sample test), with the same overall learning outcomes, purpose and mandate.

When it comes to evaluating complex performances (such as the extended writing elicited in Sections 1 and 4 of the sample test considered here), we need to provide specific guidance to teachers (raters or markers). Such specific information would be detailed in the scoring section of a task specification. Typically, this information is transferred to rating scales or marking keys. In the section which follows, we look more closely at types of rating scales.

4.5 Rating Scales

A *rating scale* is essentially a description of the behaviours, features, or characteristics of a performance, which are referenced to point values that form a *scale*, that is, typically, a numerical range from low to high.

A rating scale allows us to operationalize increasing levels of proficiency, ability, knowledge and so on in a language – that is, to relate increasing levels to a numerical measure, in much the same way that we put a numerical value on a measurement of *heat* (temperature) by using a thermometer or thermostat. By using a rating scale we can situate an individual performance in relation to the continuum of learning we have defined.

Some scales simply define *mastery* in relation to a description of the amount of skill, ability, or competency that is evident in the performance we are evaluating. Mastery scales

Table 4.4 Holistic scale – Section 1: summary writing

Points	Description
0	Barely attempts (or does not attempt) to address the task.
5	Although there is some evidence that the writer/test-taker understood the task, the summary is very limited and/or largely incomprehensible. Most of the information in the two text(s) is missed, miscommunicated, or misunderstood.
10	Some important information is missing from the summary (i.e., all of one text is summarized, but the other text is not; only the main ideas are summarized without mentioning supporting details). The answer is generally difficult to understand because of systematic problems in the writing.
15	Overall, there is an adequate summary of the overall positions taken by the authors of both texts, but the test-taker does not include supporting details. It may be difficult to understand some sections of the summary.
20	Fulfills all task requirements: summarizes the authors' respective positions in both texts comprehensively (i.e., provides a statement of the main ideas and the supporting details). Fully comprehensible, although there are errors in expression.

result in a pass/fail evaluation. Other scales are concerned with the *degree* of skill, ability, or competency that is evidenced in the performance. There are two types of such scales, namely **holistic scales** (Table 4.4) and **analytic scales** (Table 4.5).

- **Holistic scales** rely on the general overall impression of a rater. Raters consider a performance (i.e., in writing or in speaking) as a whole and relate their impression to a scale that provides a criterion-referenced description that is linked to points. The rater does not consider specific features of the writing or speaking.

Above is an example of a holistic rating scale which might be applied to evaluate a summary of texts in Section 1 of the Sample Integrated Reading-to-Writing Test in Figure 4.1.

- **Analytic scales** identify specific features of a performance, separated into categories on the rating scale. Analytic scales are often preferable in assessment contexts because they provide so much more information to teachers and students about specific aspects of a performance that need attention or show development. For example, in an analytic scale that was designed to evaluate a student's performance in an oral interview, we might have categories for accuracy (i.e., grammar, vocabulary, comprehensibility and content).

Compare the holistic scale above with an analytic rating scale (below) for the same section of the sample test.

Regardless of the type of scale we are using, we need to provide teachers or raters with practice runs using the scales to evaluate test performances. Rater training promotes consistency in rater judgments. Research on the rating of speaking and writing has demonstrated that it is possible to obtain high levels of agreement across raters – provided they receive training in the interpretation and use of rating scales. When there is *high inter-rater reliability*, we contribute to the overall quality of inferences we draw from test scores and this is a requirement for validity.

Table 4.5 Analytic scale – Section 1: summary writing

Meaningfulness of Content	Points	Accuracy of expression	Points
No attempt or very limited attempt.	0	Too little to judge.	0
Some attempt to summarize, but the summary is largely inaccurate or too limited to be meaningful. There may be evidence that the writer misunderstood, or the writer copied verbatim from, the source texts.	1–2	Inaccurate; extensive error in the simplest of phrases; very challenging to read and understand; very limited control; repeated use of simple vocabulary. Or, copied directly from the source texts.	1–2
Key relevant components of the summary are missing or misstated (e.g., main ideas are not identified or no/few supporting details are provided). The summary lacks completeness. The demands of the task are not met.	3–4	Errors are scattered throughout; expression is uneven and at times incomprehensible. Vocabulary is limited. The writer is clearly struggling to express his/her ideas. There may be evidence of patch writing (weaving phrases lifted from the texts into the summary).	3–4
Although overall the writer addresses the demands of the task, the summary is minimal; important information is missing or under-explained. Content is reduced, simplistic, or inaccurate.	5–6	Sporadic errors are offset at times by facility in expression. Some sophisticated use of language and/or attempts to use longer, complex sentences; variety of vocabulary; idiomatic expression. Respects academic conventions in citing directly from the source texts.	5–6

Adequately meets the demands of the task. Somewhat uneven control of content summaries of the two texts. Although the overall summary of the main ideas is clear, some important supporting details may be missed.	7–8	Some consistent, systematic errors but these do not interfere with overall comprehension. Generally controls syntax. Expression may be somewhat redundant; overall the writing is comprehensible and sophisticated.	7–8
Meets or exceeds the task requirements; full summary provided of main ideas and supporting details of the two texts. Academic conventions are observed in citing from the source texts.	9–10	Somewhat limited or challenged at times; a few errors, but fully and easily comprehensible.	9–10
Totals:	/10		/10

ACTIVITY 4.5

Respond to each of the following questions. Compare your answers to those of a colleague if this is possible. Do you agree or disagree?

- Which type of rating scale would you prefer to use as a teacher marking test writing? What are the advantages and disadvantages of each?
- When would it be most appropriate to use a holistic scale? When would it be most appropriate to use an analytic scale?
- What is the feedback potential of the two scale formats – which one would allow for the most learning?

Rater training provided to teachers in a language programme is another excellent means of supporting their professional development and increasing the coherence of the programme.

4.6 Test Administration and Analysis

If we are developing tests for use within our own classrooms or programmes, we may not always have a chance to pilot or try out our tests in advance. It is useful to administer a new test to a group of students at another level (below or above the level the test is designed for). It is also useful to administer the test to other colleagues (fellow teachers who are familiar with the content of the course). Testing the test before it is administered in a live test-taking situation is a very important step. It will reveal issues with the test that were not evident at the planning and writing stages.

Once we begin administering the test, we will start to collect evidence of what is working and what is not. This evidence will be used to revise the test (and test specifications if the evidence

suggests that this is warranted). If you look at the item specification for Section 2 of the sample test, you'll note that in this multiple-choice section, a *comment* box was provided for test-takers to provide feedback on items. In many cases, test-takers will comment on items that they found confusing or ambiguous. In live test situations this kind of feedback is very valuable. It helps us to see the test from the test-taker's perspective. It will contribute directly to revisions of the test and the test specifications. We can also collect evidence in the form of test-taker questionnaires, which ask our students to provide feedback on the test as a whole when they have finished their work. Again, their feedback is particularly useful in revising and improving the overall quality of the test.

In addition to collecting responses directly from test-takers in order to evaluate how our test is working and to identify what to revise, we can also undertake a simple, straightforward analysis of how our items and/or tasks are functioning. We can use the information provided by *the overall test scores* to analyse **item difficulty**. We can also analyse how well an item discriminates between high-performing students or test-takers and low-performing ones: **item discrimination**. An example of how we can determine item difficulty and item discrimination is provided below.

Take a look at the test results from one class of 15 students (Table 4.6). There were 50 points (100%) possible on the whole test. If you look at Table 4.6, you'll see that Ari received 100% – a perfect score; but Johanne got only 40% on the test as a whole. Let's analyse the 10 items in the multiple-choice section of the test in relation to these students' overall test scores. What do you notice? Note: If '1' is entered, the student got the answer right. If there is a '0' and a letter, such as '0/A', the student got the item wrong and the letter indicates which incorrect distractor they chose – so 0/A indicates that they chose answer A, which was incorrect. By simply reviewing the results provided in the example of Table 4.6, we can answer all of the following questions:

Table 4.6 Item analysis for Class 6B (Level 3)

Integrated skills: Reading-to-Writing for Academic Purposes (Level 3, Mid-Term) Section 2: Genre Awareness												
Name	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	OVERALL Score (%)	All Sections
1. Ari	1	1	1	1	1	1	1	1	1	1	100	
2. Maryam	1	1	1	1	1	1	O/B	1	1	1	90	
3. Lily	1	1	1	1	1	1	O/B	1	1	1	90	
4. Paul	1	1	1	1	1	1	O/D	1	1	1	88	
5. Lu	1	1	1	1	O/A	1	O/D	1	1	1	80	
6. Ying	1	1	1	1	1	O/B	1	O/C	1	1	78	
7. Ali	1	1	O/B	O/C	O/B	1	1	1	1	1	70	
8. Kim	1	1	O/A	1	1	1	O/B	1	1	O/A	70	
9. Minnie	1	1	1	1	O/B	1	1	O/B	O/C	O/C	65	
10. Emma	1	1	1	1	1	O/A	1	O/B	O/C	O/C	60	
11. Don	O/D	1	O/C	O/A	O/C	1	1	1	1	O/C	56	
12. Shin	1	1	1	1	O/B	O/B	1	O/B	1	O/B	50	
13. Shaheen	O/A	1	O/C	1	O/C	O/B	1	1	O/C	O/C	49	
14. Natalia	O/D	1	O/B	1	O/C	1	1	1	O/C	O/B	46	
15. Johanne	O/B	1	O/B	1	O/C	1	1	O/D	O/C	O/A	40	
Correct Response	C	A	D	B	D	C	A	A	B	D		

1. Which item was easiest? (Item 2, because everyone in the class got the answer right.)
2. Which item was the most difficult? (Item 10, because 8 of the 15 students in this class got it wrong.)
3. Do all of the items discriminate equally well between high-performing and low-performing students? (No. Take a close look at Item 7. Who got this item right? Who got it wrong?)
4. Which item best discriminates between high- and low-performing students? (Item 10. All of the highest scoring students got it right; all of the lowest-scoring students got it wrong.)
5. Which items would you revise or throw out?

Before answering question 5, let's calculate the **difficulty** of **each item** by dividing the number of students who got the item right by the total number of students. For example, 11 out of 15 students got Item 1 right: $11 \div 15 = 0.73$; but 15/15 students got Item 2 right: $15 \div 15 = 1$. Only 7 of 15 students got Item 10 right: $7 \div 15 = 0.46$. So Item 10 was very difficult. Does this mean it was a poor item needing revision? Or, should we throw it out entirely? Before answering these questions, let's look at whether Item 10 discriminated between those who performed well overall on the test and those who did not. Basically the discrimination analysis provides us with information about who got the item right and who got it wrong in relation to overall performance in the test.

Think about this. As teachers we have day-to-day experience with our students in the classroom, and we accumulate evidence of their performance over time. We have an informed sense of which students are performing at the highest levels in our class, and which ones are not performing well. So, if the students who knew the most and were the most capable get an item wrong, and the students who knew the least or were evidently the least capable get the item right, what conclusion would you draw? Before answering this question, consider the perspective of an external test developer who is working on a test for the same group of students. How would the external test developer evaluate the quality of the item? Although it is

not foolproof, there is a simple approach to analysing how well an item discriminates and this provides a very useful tool for teachers as well as external test developers.

ACTIVITY 4.6

Beginning with Item 10, analyse its difficulty level by dividing the total number of students who got it right by the total number of students in the class.

Step 1

- How many students got it right? _____
- Divide this number by 15 (the number of students in the class)
- Difficulty level? _____

Step 2

Choose four other items from the test and write their difficulty levels in the spaces provided below. Item 10 is entered for you.

Item: 10	Item:	Item:	Item:	Item:
$7/15 = 0.47$				

What does the item difficulty tell you about each of these items?
What does it tell you about the test?

Now we are ready to consider **item discrimination**, that is, how well each of these items discriminate between high-performing and low-performing students in our test. Basically, we will subtract the item difficulty for the lowest-performing students from the item difficulty for the highest-performing group. A rule of thumb is to select approximately $1/3$ of the students who received the highest scores on the test as a whole, and $1/3$ of the students who received the lowest scores. Look at [Table 4.6](#). Note the

‘Overall Score’ column. The table has been set up so that the responses of all the students are arranged in order from the highest to the lowest score.

Step 3

There are 15 students in the class (so $1/3$ of the students = 5). We would first analyse the item difficulty for the group of 5 students with the highest scores, and then the group of students with the lowest scores.

Highest group: number of students in the group who got Item 10 right = 5.

5 (correct) divided by 5 (students) = 1 [$5/5 = 1$].

Lowest group: number of students with the lowest scores who got Item 10 right = 0.

0 (correct) divided by 5 (students) = 0 [$0/5 = 0$].

Step 4

Subtract the difficulty level of the lower group from the difficulty level of the higher group.

1 (item difficulty highest group) – 0 (item difficulty lowest group) = 1.

4.7 The Test Results

It is very unusual to have an item that discriminates perfectly (i.e., 1) between those who knew the most on the test and those who knew the least. An item like this is what norm-referenced test developers have as their objective. Those who are working on norm-referenced tests have the principal goal of discriminating between (and ranking) individual students in relation to other students or groups. In the classroom, however, typically we are more interested in describing what our students

know and can do. If we find one of our test items discriminates perfectly between the highest and lowest groups in our class, it will help us to identify where we need to focus our attention in supporting the learning of the students who did not perform well on the test (the lower group).

You may not have needed to analyse item 10 if you carefully examined [Table 4.6](#). Just by looking to see who got Item 10 right (the higher-scoring students) and who got it wrong (the lower-scoring students), it is evident that the item marked a dividing line in the class between upper-level scores and lower-level scores.

In 1954, **discrete-point tests** (i.e., in which each test item measures one discrete feature of the construct, such as the meaning of a word in a vocabulary test or adjective use in a test of grammar) dominated language testing. In practice, multiple-choice test formats and norm-referenced testing were the sole focus of many language testing organizations, and issues of validity and test quality were vigorously debated by theorists, researchers and test developers. Ebel (1954) provided guidelines for judging the quality of an item using a discrete-point test.

Drawing on total scores (i.e., the overall information generated by and within the test itself), Ebel suggested that item discrimination values at 0.4 and above should be viewed as high-quality items; and values below 0.2 should be viewed as low-quality items. Low-quality items should be rejected or improved through substantive revision. The higher the item's discrimination (i.e., the closer to 0.4 and above) the better it separated those who had the capability, competence, or knowledge, and those who did not. The lower the item's discrimination, the lower its usefulness.

Ebel's guidelines are useful in classroom assessment even though they originated in norm-referenced testing. Taking the time to examine who got an item right in relation to how much students knew on an overall test is very helpful to us as teachers. The test can tell us where we need to direct our teaching, which skills, concepts, understandings are missing for some of the students in our class. Because our aim in the classroom is to support the learning of all of our students, it is not a bad thing if we find some of the items on our test are answered by all of

the students in our class. This is our ultimate goal – that all of our students should learn what we are teaching them. However, taking the time to analyse some of the items on our test will help us root out problem items (e.g., ones which all of the high achieving students in our class missed on the test, while low achieving students got them right) and improve the quality of items which require some revision. Calculating item facility and item discrimination takes only a few minutes, but the information can inform our teaching in important ways. In Activity 4.7, analyse the items in [Table 4.6](#) for their quality. Continue with distractor analysis (see Activity 4.7).

There are many other approaches to judging item quality in use today. One that you may have heard of is **Rasch analysis**, which draws on a complex theory (Item Response Theory), and requires large numbers of tests to examine item difficulty and discrimination. As was the case with Ebel's guideline,

ACTIVITY 4.7

Working alone or, if possible, with a colleague, select four items from [Table 4.6](#) and analyse their discrimination.

1. Explain why you selected each item?
2. What results did you get for each item?
3. If you look at the overall difficulty of each item and its discrimination, would you recommend revision or rejection of this item on the test?
4. Now examine the table again and consider your review of the items in relation to a distractor analysis.

We can derive more information from [Table 4.6](#) by conducting **distractor analysis**, that is, examining how our distractors are performing. As you know, this was a multiple-choice test format with ten items. When a student got the item wrong, we noted in the table which of the incorrect distractors the student chose.

In other words, for each item stem (question), four choices (distractors) were provided, as in the example below:

(Item 1) Which of the following best describes the author's tone or attitude?

- A. angry (*clearly right*)
- B. enthusiastic (*clearly wrong*)
- C. unhappy (*somewhat right, but the overall tone is angry*)
- D. impatient (*somewhat right, but the overall tone is angry*)

Comments:

What we intend by offering three wrong choices is to spread out, across the distractors, responses to our items. If a distractor is so weak that no one chooses it, it is not functioning effectively (we might as well offer only three choices instead of four).

5. What does distractor analysis reveal about each of the items you selected for analysis? Can you spot a problem?
6. Now, put all of this information together. What would you recommend with regard to each of the items you analysed? If appropriate, share your observations with your colleagues.

Consider this: It is important to remember that we are not trying to *trick* our students in a multiple-choice test, rather we use distractors to test our students' ability to make subtle but important distinctions between correct and incorrect interpretations. We do not want to use distractors that are so clearly wrong that no one chooses them. If we use such a distractor, we might as well have three choices instead of four, and we increase the potential for guessing rather than providing our students with a chance to show how much they know and can do. In developing distractors for a multiple-choice test first consider what the item is testing, then identify the fine distinctions that separate students who have a deep and meaningful understanding from those who do not.

Rasch analysis uses the information generated by the overall test to judge the quality of an item. Rasch analysis is beyond the scope of our discussion of test development and item analysis (and would rarely be used in classroom assessment), but it is an approach that has wide use in large-scale, high-stakes testing. Ebel's guideline was a precursor to Rasch analysis in that it used the overall scores of individual test-takers to provide more information about the quality of an item within a test. For more information about applications of the Rasch model and item analysis, see, for example, Bond and Fox (2007) or Livingston (2006).

4.8 Looking Back at Chapter 4

Whether you are an individual teacher working to develop a test in your own classroom or a member of a team that is designing a test for a level or programme, specification-driven test development will support the meaningfulness of the information you elicit, the usefulness and appropriateness of the inferences you draw from test performances, and the fairness of testing activity. Analysing how your test is working, who is getting an item right, the item's overall difficulty, and (in the case of multiple-choice), how your students are responding to each of the distractors is an exceptionally useful approach to reviewing a test you have developed. It is a means of *testing your own test* – to be sure it is measuring what you intended it to measure in a fair and meaningful way.

Of course, tests are only one assessment alternative of many, as we discussed in Chapter 3. When there is a clear mandate for a test, it is important to understand how to build a test of the highest quality. Tests elicit a snapshot of a student's mastery, ability, proficiency, or achievement. They work alongside other forms of assessment – such as portfolio assessment, which draws on multiple sources of evidence of student's development over time. Many teachers who use portfolio assessment include a section in the portfolio for tests and quizzes.

Like other evidence collected in the portfolio, high-quality tests, which are validated and fair, provide a clear picture of what our students have learned, know and can do.

Suggested Readings

Davidson, F. & Lynch, B. K. (2002). *Testcraft: A teacher's guide to writing and using language test specifications*. New Haven, CT: Yale University Press.

A thorough and comprehensive guide for teachers and teacher trainers/test developers in the development of specification-driven tests. The book provides useful discussion questions, exercises and advice to those engaged in the development of language tests.

Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge: Cambridge University Press.

Douglas provides a detailed account of the assessment of language used for specific purposes (LSP), within target language use (TLU) situations. This branch of language assessment draws its test content and methods directly from the TLU domain. Douglas explains why and how an LSP test designed for medical practitioners would differ from one designed for tour guides or office receptionists. The book offers test developers extensive guidance on the development of LSP assessment.

Who Are We Assessing? Placement, Needs Analysis and Diagnostics

Activate your learning

- How can knowing more about our students improve the quality of our teaching and learning?
- Which assessment tools are most useful in getting to know our students?
- How can placement, needs analysis and diagnostic assessment help our students?

5.1 Which Assessment Tools Can Help Us to Get to Know Our Students? Why is This Important?

The research literature is full of examples of how better understanding of our students' language learning experiences can improve the impact and effectiveness of our teaching and, as a result, enhance learning (see, for example, Gottlieb, 2006; Ivanič, 2010; Zamel and Spack, 2004). In Chapter 5, we discuss assessment approaches and practices that can help us to better understand our students' learning by addressing the following questions about *who* our individual students are:

- What is each student in my class bringing to the learning of the language?
- How does each of my students intend to use the language in the future?
- What are my students' learning goals?

When we recognize the unique and varied cognitive, cultural, educational and emotional differences of our students, we are in a far better position to address gaps in their knowledge, skill and understandings and develop their strengths in learning (Cheng, 2013, Fox and Cheng, 2007).

In Chapter 5 we take a closer look at specific assessment tools that we use in order to better understand our students' needs and capabilities. We examine these tools in relation to a chronology that often applies in our use of assessment in our classrooms or programmes. Although a placement test, needs analysis and diagnostic assessment may be used at any point in a course, we often apply these tools at the beginning of a new course, so that we can shape the content and learning outcomes of our courses in relation to the individual students in our classes or programme.

In Section 5.2, we begin by considering **placement tests**, **needs analysis** approaches and **diagnostic assessment**. We will examine the potential of specification-driven placement testing. We will explore its advantages not only in grouping our students most effectively at different levels within our programmes, but also as a source of quality feedback for both teachers and students. We will next take a look at different approaches to needs analysis and examine how these approaches influence the kind of information we elicit and its use. Just as our own personal philosophy of teaching and learning influences the decisions we make in the classroom (as discussed in Chapter 1), so too our philosophy of needs assessment will influence the choices we make about learning activity in our classrooms. Subsequently, we will examine the increased role of diagnostic assessment in language teaching. We will examine examples of diagnostic assessment in practice, both *before* (e.g., Fox, 2009; Fox and Hartwick, 2011) and *after* students have been admitted to a university programme (Artemeva and Fox, 2010). We will pay particular attention to the development of **learning profiles** (Fox, 2009), and the role of *targeted* instruction in language teaching (Fox, 2009; Fox and Hartwick, 2011; Fox, Haggerty and Artemeva, 2016).

5.2 At the Beginning: Getting to Know Who Our Students Are and How Best to Help Them Learn

In this section, we discuss three main assessment strategies that help us to get to know our students, namely assessment for placement purposes, needs analysis and diagnostic assessment.

5.2.1 Assessment for Placement

Placement tests (as their name implies) have the purpose of *placing* students at levels of a pre-existing programme – in those programmes with a defined structure and progression. In some contexts they are used for *grouping* students based on shared characteristics (e.g., levels of proficiency, backgrounds, interests, goals, needs), for example, in a less structured programme, which defines classes and levels of instruction based on whoever enrolls. We know that when we are working with students who have similar characteristics, it is easier for us, as teachers, to shape our teaching for the common good. It is not unusual, however, to find that although students are placed on the basis of some shared characteristics they may differ considerably in others. These are the differences that we address through ongoing assessment practices in our classrooms. Whether we are working with students in secondary school or college; adolescents or adults; in English as a Second Language (ESL) or English as a Foreign Language (EFL) settings, the placement test results of one student are usually considered in relation to: (1) the other students, who are enrolled on the programme at the same time; and (2) the overall spread of students' scores across the programme.

In a criterion-referenced programme with a placement test that is guided by benchmark levels such as the Common European Framework of Reference (CEFR) or the Canadian Language Benchmarks (CLB), students are grouped in relation to levels which are defined by criteria. Thus, there may be five groups/classes of students at Level B and only one group at Level A. In a norm-referenced programme, comparisons

are made on the basis of a student's relative performance or score in relation to the range of scores on the test. Groups are identified by the scores along the continuum from low to high.

As we discussed in [Chapter 4](#), when tests are specification-driven and reflect negotiated understandings of development that explicitly relate performance on the test to the programme's developmental range, the information provided by the placement test will be especially informative.

Placement tests occur in both higher-stakes and lower-stakes settings. For example, Bachman and Palmer (1996, pp. 253–84) provide a detailed example of a placement test that was developed for the purpose of selecting students for a university-level sheltered academic writing programme. The course was **sheltered** in that only English as an Additional Language (EAL) students were eligible for it, and additional focus was placed on language support and instruction related to the development of academic writing. First language (L1) English counterparts took a *non-sheltered version* of the same course. Students in both types of courses earned university credits toward their degree programmes. The Bachman and Palmer example is particularly helpful for two reasons: it illustrates the usefulness of specification-driven placement testing, and it provides excellent detail regarding task development.

Conversely, Gottlieb (2006, pp. 15–22) provides a 'decision tree for the identification and placement of English Language Learners (ELLs)' in schools, which uses multiple instruments as part of a placement assessment process. The instruments include a *home language survey*, which elicits useful information about entering students' linguistic and educational backgrounds; a **language use survey**, which elicits detailed information on the use of the target languages and additional languages around the home and in the school or community; and external **proficiency tests**.

In some contexts, self-assessment may be used as the primary assessment tool in placement, particularly, but not exclusively, with adult language learners in low-stakes contexts. For example, students may be asked to identify their comfort level with language use in particular situations by responding to items such as:

Self-Assessment for Placement

I feel comfortable using the target language (e.g., English, Spanish, Chinese, Arabic) to:

	No	Yes
Phone for an appointment with my doctor	0 1 2 3 4 5 6	
Listen to a news story on the radio	0 1 2 3 4 5 6	
Read a newspaper article about winter cruises	0 1 2 3 4 5 6	
Fill in an online form about a missing jacket	0 1 2 3 4 5 6	

In recent years, many language programmes have developed *can-do statements*, which translate the criteria that define different levels in their programmes into actions that operationalize mastery and progress from one level of a programme to the next. Can-do statements may be a vehicle for self-assessment, teacher assessment, or both. A student might respond to a series of can-do statements such as the ones below, which operationalize a benchmark criteria associated with a specific level or class grouping in a programme. For example, a benchmark criterion might state: *A student at this level can ask and answer simple questions, and initiate and respond to statements about a familiar topic.*

Can-do statements:

At school, I can

tell my teacher I did my homework.	Yes	No
ask my friend to eat lunch with me.	Yes	No
talk about the weather today.	Yes	No
ask my teacher to explain a word.	Yes	No

If the student is known to a teacher (i.e., has completed a course and is being considered for placement at another level), the teacher may assess the student using the same can-do statements:

Can-do statements:

At school, _____ can

(name of student)

tell me if she/he finished her/his homework.	Yes	No
ask a friend to each lunch with her/him.	Yes	No
talk about the weather today.	Yes	No
ask me to explain the meaning of a word.	Yes	No

This is a particularly informative assessment procedure if both students and teachers respond to the same can-do statements, followed by an *assessment conference*, in which teachers and students compare their answers to the can-do statements and discuss any differences that occur. It is a very useful learning opportunity for our students, because it helps them to understand in real terms their relative capabilities, to improve their skill in judging their own language development, and to set realistic goals for the future.

Regardless of the approach or approaches used in placing a student at the beginning of a course, an important point to consider is the use of multiple assessment tools, whenever possible, in placement decisions. The more information we elicit about our students, the better our position will be in supporting their learning and in placing them in the right course/class. Placing each student in the most appropriate group or class within our programmes helps to increase their potential to derive maximum benefit.

It should be noted that a fully specified and validated placement test can have multiple versions (which are at the same or similar levels of difficulty, sample the same or similar competencies and skills, and so on – as we discussed in [Chapter 4](#)). Versions of the same test (or even the same version) can be

ACTIVITY 5.1

Answer the following questions about your own experience with placement approaches at the beginning of a new course. If possible, compare your own responses with those of a colleague or group:

1. What placement practices have you used?
2. What are some of the issues that arise if students are placed in the wrong levels? Have you ever experienced this misplacement (either as a teacher or as a student)?
3. Which placement approaches seem to work best, based on your own experience as a teacher or a student?
4. In some language programmes, external standardized tests are used to place students. For example, a high-stakes proficiency test, such as the Test of English as a Foreign Language Internet-based Test (TOEFL iBT) or International English Language Testing System (IELTS) might determine placement in a programme. Discuss some situations in which a placement approach like this would be effective? When, in your view, would this be a mistake?
5. Some textbooks provide placement assessment approaches or tests. Have you ever worked with such a textbook? How have textbooks figured in placement decisions within programmes you are familiar with?

used to assess *achievement* at the end of a course as well. If students are continuing in the programme, their achievement can also serve as a key placement indicator for their subsequent course placement. New students can take the placement test and be grouped accordingly in relation to the students already in the programme.

5.2.2 Needs Analysis

Often at the beginning of a new course (but also at any point during a course), we assess our students' *needs*. As we discussed above, assessment for placement has as its purpose the grouping of students for maximum benefit within a programme. The main purpose of needs analysis or needs assessment (we will use these two terms interchangeably) is to elicit information about individuals in our classes or programme in order to inform our course design decisions. Knowing more about our students will help us to identify the activities and experiences that have the greatest potential of supporting their learning. We might, for example, use some of the instruments described above (p. 142) as part of our needs analysis. One point is clear: the more we know, the more effective we will be in shaping our teaching to meet our students' needs and support their language development. When needs analysis is systematic and well designed, it will help us to prioritize the activities in our classes. Needs analysis helps us to answer the difficult questions that are implicit in each of the decisions we make as we orchestrate the learning in our classrooms:

- What is the most useful activity for these students/this student: in this class, at this time, for this purpose?
- How will the activity help to achieve the intended learning outcomes of my course?
- Where should I begin? What should I do first? What is the most necessary?

The first step in a needs assessment or needs analysis is to identify the kind of information about our students that would be the most helpful to know. As we discussed in [Chapter 4](#), once we have decided what we would like to *measure*, we can begin to operationalize the measurement in concrete terms as tests, checklists, questionnaires, interviews, journal responses and surveys – using any of the assessment tools that we have in our assessment repertoire. Needs analysis can take any of these forms.

Because effective needs assessment provides a strong foundation for good practice in our language classrooms, there has been an increasing focus on it in the literature on language teaching and assessment (e.g., Brown, 1995; Graves, 2000). However, needs assessment has been a prevailing interest of teachers and programme administrators across educational levels and contexts for many years, because it is ‘a process that helps one to identify and examine both values and information’ that inform teaching and learning; and, it ‘provides direction for making decisions’ (Stufflebeam et al., 1985, p. xiii). As Stufflebeam and colleagues suggest, the way in which we structure needs analysis speaks again to our philosophy of teaching and learning (as we noted in Chapter 1). The type of information we collect, how much we collect, when we collect it and how we use it, all relate to our underlying philosophy.

There are a number of purposes of needs analysis. For example, we may consider that the purpose of our needs assessment is to identify *gaps or discrepancies* (see Stufflebeam, et al., 1985) between a student’s current level of performance and the intended level of performance (learning outcomes). If this is our purpose, we will typically elicit information on our students’ strengths and weaknesses so that we can address the weaknesses and work on developing their abilities, skills and proficiency to reach the desired/intended level.

We may view our needs analysis as a means of eliciting and supporting student *participation* in the selection of and emphasis on class activities, topics and procedures. We may view our students’ input on the directions taken in the course as an essential feature of engagement: a means of ‘making it theirs’. We may consider their participation in choosing the directions of the course as a way of increasing its meaningfulness. By recognizing our students’ interests, elicited through needs assessment, we encourage the development of their personal sense of responsibility for learning, improved goal-setting and increased awareness of the role of self-assessment.

We may also consider a needs analysis as an *analytical* means that will help us to define next steps. As such, the purpose of the needs analysis will be to elicit specific information about students' current levels in order to establish what we should do next. Our goal is to encourage development, and from this perspective development is best achieved when we have a more precise understanding of what a student brings to the learning so that we can map out the incremental activities to support development.

Alternatively, we may view needs analysis as a *diagnostic* process, which identifies what is missing in a student's learning that might place that student at risk. In other words, in relation to the purpose and context of our course, a diagnostic viewpoint on needs assessment sees it as a means of insuring that critical (threshold) capacities are in place which, if missing, might lead to failure or harm.

Our philosophies of teaching and learning, which we explored in [Chapter 1](#), and the contexts in which we teach, will influence our perspectives on needs assessment. In

ACTIVITY 5.2

Look back at the many assessment tools that have been identified in the earlier chapters and the overview of some commonly used Classroom Assessment Tools and Test Formats (see the Appendix).

Select several assessment tools which you think would be the most useful in addressing the questions raised by each of the needs assessment philosophies described in [Chapter 1](#). Examine [Table 5.1](#). List the assessment tools in the space provided in [Table 5.1](#). Working alone or, if appropriate, with a colleague, discuss the kinds of information these approaches elicit. What differences do you notice? What are the implications?

Table 5.1 Mapping assessment approaches onto philosophies of needs assessment (see Brown, 1995 and Stufflebeam et al., 1985, for details regarding these philosophies)

<i>Needs analysis philosophy</i>	<i>Questions of interest, given the philosophy</i>	<i>Assessment approaches and instruments</i>
<i>Discrepancy</i>	What differences exist between my students' current level of performance and the desired level? What is needed to change the level of performance to the desired level?	(List several that apply)
<i>Participatory</i>	What are my students' interests, goals and recognitions? What do they hope to learn by taking this course? What are the priorities for our learning during the course?	(List several that apply)
<i>Analytical</i>	At the present time, what do my students know? What are they able to do? How do I build on what they know and can do? What are my next steps in increasing their capability?	(List several that apply)
<i>Diagnostic</i>	What do my students need to know and be able to do in order to participate through language in the target domain? What essential skills, capacities, knowledge is missing that might undermine their participation or place them at risk? What are they able to do now that can be extended and strengthened?	(List several that apply)

Activity 5.2, we examined the relationship between the questions a needs analysis is intended to address and the types of assessment approaches or instruments that we might use.

Although we have approached needs analysis as an assessment tool that is frequently used by teachers at the beginning of a new course, as we noted above, it may be used at any point in a course as a means of informing our teaching. Needs are not fixed; they evolve over the days and weeks of our interaction with our students. It may be helpful to repeat a needs analysis at the mid-point in a course to help us to take stock.

In completing [Table 5.1](#), we listed many different assessment approaches and tools that can be used for needs analysis. We can conduct an informal needs analysis at any point, however, using very simple approaches such as the ‘Five-Minute Essay’. Have you ever done this in your own class? At any point in a class, stop the action and ask your students to respond in writing to the following three questions:

Five-Minute Essay

1. What’s clear to me?
2. What’s still fuzzy? (I’m still struggling to understand this.)
3. What would help me most at this point in the course?

Of course, there are many other questions that we can ask our students to address. For example, What do I need more of? What could I use less of? What do I really like so far? What is helping me most? and so on. Limit the number of questions to three and limit the time for responding to around five minutes.

Let your students know that they do not have to write their names on their Five-Minute Essays. Collect their responses and review their feedback. This is an on-the-spot needs assessment that will improve the quality of your teaching and address the specific concerns, issues and so on of your students. It only takes five minutes for this needs assessment and requires little advance planning, but it is a very useful means of getting to know your students’ perspectives on the course – at any time during a course.

To read more about needs analysis and see more useful examples consult Brown's *Example Needs Analysis* (1995, pp. 55–65). In one case, the example is of an informal needs analysis; the other is quite formal. Brown provides an overview of procedures, summarizes and interprets the results, and discusses how the information was used by the two language programmes considered in these examples.

5.2.3 Diagnostic Assessment

A third type of assessment, which most often occurs at the beginning of a course, is diagnostic. We may ask what distinguishes diagnostic assessment from placement testing and needs analysis. Here we rely on the work of Alderson (2005, 2007), Fox (2009), Fox, Haggerty and Artemeva (2016), and others who distinguish diagnostic assessment from other forms of assessment, because it is more *narrowly scientific*. A diagnostic test or assessment procedure is *fully specified* (as we discussed in Chapter 4) to test for specific capabilities that are related to target or intended competencies, skills, or abilities. Further, information provided by the diagnostic test should trigger specific pedagogical interventions, which are designed to address an individual's weaknesses and strengths through classroom activity. In other words, diagnostic assessment is not fully diagnostic unless it leads directly to teaching that responds to an individual's particular language or learning profile.

Whereas assessment for placement has as its purpose the grouping of students based on what they share, diagnostic assessment examines their individual and unique capabilities and weaknesses and identifies specific activities that address those capabilities in order to support an individual's development.

Let's look at examples of diagnostic assessment in practice: in the classroom, across a programme, and across a system.

- *Diagnostic assessment in the classroom*

Example 1: In a conversation class

Have you ever taught or taken a course in conversation? Did you administer or take a diagnostic assessment at the

beginning of the course? A diagnostic assessment can be used to get at the micro-level issues that interfere with an individual student's comprehensibility when speaking a new language. For example, we know that pronunciation differences impede communication. A diagnostic test of pronunciation at the beginning of such a course would probably not resemble the classroom activities you designed for the course, but it would identify the issues that your individual students needed to address with your help – issues that would impede their ability to communicate clearly and would impact their performance in the course if they were not addressed. For example, some students may have difficulty pronouncing certain consonants. A diagnostic assessment would identify which ones and identify potential exercises to address a student's recognition and production of these consonants. Because speakers of different languages have different challenges in pronunciation, we would not have time in our classroom to address each of these differences for our whole class. Using diagnostic assessment, however, we can identify an individual's capabilities and weaknesses (through diagnosis) and then offer a range of exercises specific to the individual's requirements. In this way we can target the specific micro-level challenges that could impede learning and development.

Such diagnostic tests and concomitant exercises and activities are freely available on the Internet and in textbooks. See, for example, <http://www.lex tutor.ca/tests/levels/productive/> (by Laufer and Nation, 1999, adapted for the Web by Tom Cobb), which tests productive vocabulary drawing on the academic word list. Or, to test how well you can distinguish between minimal pairs in listening, try, <http://www.learnenglishfeelsgood.com/listening/esl-listening-practice14.html#>; or http://www.world-english.org/diagnostic_grammar_1.htm for diagnostic tests that address a range of issues.

If we are teaching a homogeneous class (where all our students speak the same first language), it is possible that our diagnosis may lead to a classroom activity for the whole class. In this case, we can raise our students' awareness by

administering the diagnostic assessment, tap into what they have recognized and the strategies that some of the students use to address the micro-level challenges of communicating in the new language, and develop a range of in-class activities to promote production.

In many language classrooms, however, our students are linguistically and culturally diverse. In a conversation class for students with backgrounds and proficiency in many different languages, the issues in speaking the target language will be equally diverse. One of our students may be challenged by the pronunciation of consonants, but another may have challenges with intonation. Still others may speak too quickly so that words blur together in a flood of speech. Or, it may be that a highly motivated student with brilliant ideas and excellent vocabulary is largely incomprehensible, because the student has all three of these micro-level issues when speaking the new language (i.e., consonants, intonation and pace). Each one of these issues can be addressed. The first step is to diagnose what interferes with communication. The next step is to provide individualized activities to address them.

Example 2:. In an English for Academic Purposes (EAP) class

Peggy Hartwick (see Fox and Hartwick, 2011) teaches in a pre-university *English for Academic Purposes* (EAP) programme. She has used what she refers to as *diagnostic assignments* to ‘drill-down’ to her individual students’ strengths and weaknesses in academic English. These diagnostic assignments (see [Figure 5.1](#)) help her to diagnose her individual students ‘skill sets’ (p. 50). She used this assignment recently when she taught a class at the advanced level.

She notes that when her students are placed in her EAP classes, their skills in academic English are not necessarily balanced or even. In fact, many students in the same class may have very different skill sets. For example, some can speak fluently and easily in English; others may be reluctant

Week 2 - September 8 & 10

Sept. 8—Diagnostic Unit



Diagnostic Unit: Pre-reading online 'chat' forum



Diagnostic Unit: Reading Comprehension and Vocabulary Section

Available from 8 September 2015, 8:30 AM



Diagnostic Unit: Listening on Digital Literacy

Available from 8 September 2015, 8:30 AM



Diagnostic Unit: Speaking Sample



Questionnaire



Self-assessment of English proficiency skills

Homework for September 10th

- Read the first reading by Strange and Bayley (2009) in your course pack. Read actively by taking notes and responding to guided questions (you can find this worksheet in "weekly handout folder").
- Do the **Diagnostic Speaking Unit**. Use your cell phone to record an MP3 file. If you are unable to do this, talk to me on Thursday.
- Finish the Questionnaire and Self-Assessment of English proficiency skills from Tuesday's lab.
- We will do **Diagnostic Writing Unit** on Thursday in class.

Figure 5.1 An example of an online diagnostic assignment

to speak and/or may be very difficult to understand. Her students also differ in their development of academic reading, writing and listening skills. She argues that if she really wants her students to develop to their full potential, she needs to understand their development in each of the academic skills and then devise activities that will support their learning.

Peggy typically works with classes of 30 students. She also has the advantage of having a computer-based learning management system to work with in her teaching context. She explains:

I used this [see Diagnostic Unit in [Figure 5.1](#)] last term and all parts were facilitated through the learning management system (CuLearn – a Moodle-based system). Students completed the

following parts in one day. My focus continues to be on drilling down in order to identify individual student strengths and weaknesses. Due to class sizes it has become increasingly difficult to focus on the individual, but the profiles help to provide more of a global picture of student needs for the term. The writing prompt is given in the next class. Students get a mark of 5%.

Figure 5.2 provides the Diagnostic Writing Unit or diagnostic assignment, which helps Peggy to identify strengths and weaknesses in writing.

Diagnostic Unit: Writing Sample on Digital Literacy

We have read about the importance of developing digital literacy skills. With respect to the three identified competencies (i.e., use, understand, create), how are these skills necessary in a business or educational setting? Explain.

I will hand out an additional reading in class (i.e., an article entitled 'Digital literacy and basic competencies') to provide additional support for your writing.

Your answer should begin with your own definition of digital literacy and a brief explanation of the three competencies, according to the reading. Support your answer by providing clear examples (you can draw on your own personal experience).

Your answer must be a minimum of 350 words. Focus on responding to the question. In marking your response, I will consider:

- your focus in response to the question,
- use of information, including the definitions you provide,
- language structure and accuracy,
- organization, and
- the *tone* of your writing (as we discussed in class, remember, this is academic not personal writing).

Figure 5.2 Follow-up diagnostic assignment (writing)

Peggy uses the information she gathers from these diagnostic assignments to develop an individual skills set or *learning profile* for each of the students in her class. She uses diagnostic assignments at intervals in her course to tap into changes in her students' profiles. She also draws on the profiles in grouping students for work within the class to address the areas where they need additional support.

Peggy offered to share her assessment rubric for the diagnostic assignments: 'I am sending you the form I use to assess the diagnostic work. This has evolved over the last few years and identifies the problems that I have come to see "frequently" in my students' work at this level.' Below is a copy of the form of the learning profile that Peggy uses to discuss the results of their diagnostic assignments with individual students at intervals during her course.

Based on the observed outcomes of your diagnostic assessment, I have identified the following learning priorities for you this term. Please review this report carefully and focus your attention on these areas this term.

Pre-reading online chat	/5
active ☐ somewhat active ☐ not active ☐	
• Does not engage/respond to prompt • Does not contribute critically to online chat • Does not use appropriate language	
Reading and Vocabulary	/15
• Does not appear to have read or understood question(s) • Does not locate specific information/details (skimming and scanning) • Does not demonstrate a general understanding of content through answers	

- Reading speed appears below average (quiz not completed in allotted time)
- Does not demonstrate vocabulary knowledge
- Does not demonstrate fluency (word accuracy)

Listening**/5**

- Does not have organized or detailed notes
- Does not show the gist or main idea in short answers or when paraphrasing
- Does not identify specific information in answering questions from listening

Writing**/20**

- Does not respond to or understand the prompt (no claim/off topic)
- Does not convince or persuade reader
- Does not develop content or support claim
- Does not demonstrate logic
- Does not organize writing (use of logical transitions)
- Does not refer to source(s)
- Does not use keywords from the topic or a variety of academic vocabulary
- Does not use complex or accurate structures
- Grammar breaks down
- General: errors with spelling/punctuation/length/copying/referencing/word form

Speaking Sample**/5**

- Does not have clear speech
- Appears to be uncomfortable
- Hesitates frequently (fluency)
- Does not respond to prompt

5 = Very strong
 4 = Proficient
 3.5 = Developing
 3 = Weak

Questionnaire yes/no **Self-Assessment** yes/no

Comments based on learning profile conference: Date: _____ _____ Date: _____ _____ Date: _____ _____
--

Figure 5.3 Diagnostic assessment: student profile of targeted needs (adapted from Hartwick, 2016)

ACTIVITY 5.3

Diagnostic assessment helps us to identify an area (or areas) in a student's proficiency, ability, or skill set that needs particular attention. Providing targeted instruction to an individual student can make a critical difference in the student's development.

Take a look at the following list of diagnostic assessment approaches.

- Which ones have you used? Write a response describing your experience with one or more of these approaches. What were its benefits? What were its limitations?
- If you have not used any of these approaches, which one or ones would you like to try? Briefly explain why?
- What other approaches might be useful in identifying specific areas that an individual student needs help in? In Chapter 3 we looked at portfolio assessment. What use could it serve in diagnosing student strengths, weaknesses and ongoing development?

Table 5.2 Diagnostic approaches

Approach	How does it work?
Learning log	Students keep a personal record of their work during the course. After completing an assignment for the course, they respond to these three questions in writing: 1. What was easy about this assignment? 2. What was the most difficult? 3. What do I need to learn if I am going to do this better next time? Teachers collect the learning logs and review responses alongside the completed assignment. What is its diagnostic potential? This approach allows us to relate the student's awareness of challenges and difficulties in completing an assignment with their actual performance; helps us to identify and address specific areas of weakness; and helps the student to develop more self-awareness and skills in self-assessment.
Test–retest	Administer the same or similar (i.e., parallel or equated) tests at intervals during a course. Students take the same tests at the beginning, mid-points and end of a course to document changes over time. Teachers can draw on tests from the Internet, textbooks and so on which are relevant to the particular course they are teaching. At intervals, the teachers administer the test and keep a running record of how each student is doing. What is its diagnostic potential? Depending upon what the test is assessing, the information in the running record can guide decisions we make about supplementary support or next steps. There are many useful tests available on the Internet which will be of use to teachers. For example, if you are interested in vocabulary and word knowledge tests, see http://www.lex tutor.ca/ for a test of vocabulary in both French and English; or http://my.vocabularysize.com/. It will be important to

Table 5.2 Continued

	remember that at times the test–retest approach may simply suggest that a student is getting better at taking the test, not that their skills or proficiency are improving. It will be important to supplement this approach with other evidence of their strengths and weaknesses.
Diagnostic conferences (see Figure 5.3 for an example)	Students meet one-on-one or in small groups with the teacher. Teachers ask students to do something with language (e.g., to read aloud a short excerpt from a story or newspaper clipping; to share a written response to a question; to discuss a reading assigned for the course). During the conference teachers can provide immediate feedback to support students' learning, and encourage students to think about their work in other and more complex ways. It is important to keep consistent notes on observations during a conference, and/or to collect evidence of students' performance (e.g., written assignments, recorded read-alouds, or discussions). What is its diagnostic potential? The conference offers teachers an opportunity to note specific strengths and weaknesses and record observations, which can then guide and focus next steps in instruction. Conferences need not take much time. If they occur frequently over time they can be an important source of ongoing diagnostic information about a student's learning.

- *Diagnostic assessment across a programme*

Example 3: A university EAP programme

Although they are less common than the use of diagnostic assessment within individual classrooms, in some instances diagnostic assessment may be used for all the students in a programme. An example is provided by Fox (2009) in a context where the traditional placement assessment was replaced by

external proficiency tests which were unrelated to (i.e., misaligned with) the programme's EAP curriculum. Although students' placement in the programme was dictated by the external proficiency test scores, teachers within the EAP programme were provided with post-entry diagnostic information about each of their students through the administration of a diagnostic test of all the students in the programme. The information supported the identification of groups and the development of targeted activities by group for teachers in the programme.

One of the issues addressed in Fox (2009) and Fox and Hartwick (2011) is what to include in a learning profile and how to communicate it to students. Teachers within the EAP programme where these studies took place wrestled with the questions: Is too much information going to undermine the motivation of my students? How much information should be communicated? What kind of information should be communicated to my students? What information would be most helpful in supporting their learning? Fox (2009) demonstrates how she used the consensus-building *Delphi technique* to develop the learning profile for the EAP programme. Specifically, the EAP teachers met with Fox for the purpose of negotiating a format for the learning profile. Because the focus was on the form itself (and not on teacher beliefs, teaching preferences, methods and so on) they worked towards a positive consensus on the most useful information to include. Negotiation of the learning profile provided a context for professional development and supported the coherence of the programme overall.

How the teachers used the information, however, was left entirely to them. One of the EAP teachers used the diagnostic information to set up groups of students with similar language profiles and used individual learning portfolios to document change (see Fox and Hartwick, 2011). She defined intensive activities for the groups, which targeted weaknesses in competencies, knowledge and skills that had been identified through the diagnostic assessment. The learning portfolios allowed for the collection of ongoing evidence of her students' learning and development. Another teacher developed a series of specific

workshops targeting the language weaknesses of different groups of students within her class. These were external to the regular class activity however. Students were rewarded with a small number of bonus points for attending recommended workshops based on the language/learning profiles generated by the diagnostic test, but for the most part their participation in the workshops was voluntary.

As the focus on diagnostic assessment increases, questions regarding learning profiles are receiving more and more attention.

- *Diagnostic assessment across a system*

Example 4: External, large scale diagnostic assessment across colleges and universities

At the time of writing, much work is underway in diagnostic assessment, particularly in English-medium universities and other post-secondary contexts with large numbers of students who come from diverse linguistic and cultural backgrounds. These diagnostic assessments most often occur *after* a student has been admitted to their programme. Universities are increasingly concerned about retaining students they admit and supporting their academic success. Failure is costly for both the institution and, of course, the student.

New acronyms are being popularized to capture trends in using diagnostic assessment early in a student's academic career in order to identify areas of *risk*, which might lead to failure or impede academic success, and to identify specific learning options to address these areas of risk. The name applied to diagnostic assessment of language risk factors is Post-Admission English Language Assessment (PELA). There are an increasing number of such assessments in many English-medium contexts, where a large number of international students are entering post-secondary institutions or where student populations are becoming increasingly culturally and linguistically diverse (e.g., Australia, New Zealand, Canada).

For example, at both Auckland (New Zealand) and Melbourne (Australia) universities, PELA tests are administered to admit undergraduate students. At Auckland University, the Diagnostic English Language Needs Assessment (DELNA) is taken by students who have been admitted to their degree programmes. For further information on DELNA, see www.delna.nz/ and Elder and von Randow (2008) or Read (2008, 2013, 2016). At the University of Melbourne in Australia the Diagnostic English Language Assessment (DELA) is administered to new undergraduate university students (see, for example, Knoch and Elder, 2013).

5.3 Looking Back at Chapter 5

In this chapter, we have taken a closer look at specific assessment tools that we use to better understand our students' needs and capabilities. We examined placement testing, needs analysis and diagnostic assessment. Diagnostic assessment

Table 5.3 A sample diagnostic assessment tool

Criteria	Possible language support
Overall, the flow of information is logical and clear.	
Paragraphs are used effectively.	
The writer uses correct sentence structure (avoids run-on sentences and fragments, and clearly punctuates sentence endings).	
The antecedents of pronoun references are clear.	
Subjects and verbs agree throughout.	

approaches link assessment to specific or targeted instruction. Look at the following list of criteria taken from a diagnostic assessment of writing in [Table 5.3](#). What specific language activities would you use in order to support a student who demonstrated weakness in each of these criteria.

If you received a language or learning profile for each of your students at the beginning of a new course, how would you use the information? Of course, as we have discussed in other chapters, the context will inform your decision.

Suggested Readings

Brown, J. D. (1995). *The elements of language curriculum*. Boston, MA: Heinle & Heinle.

See in particular pages 35–65 on needs analysis, which focus specifically on needs assessment and provide excellent examples of how it is used in different language teaching contexts. It will be particularly helpful for teachers who are interested in developing a needs analysis for their course or programme. Brown introduces and discusses many alternative approaches to needs analysis.

Fox, J. (2009). Moderating top-down policy impact and supporting EAP curricular renewal: Exploring the potential of diagnostic assessment. *Journal of English for Academic Purposes*, 8(1), 26–42.

Fox examines the role that diagnostic assessment plays in supporting learning and teaching within an adult EAP programme. Her research study highlights the development and use of a learning profile generated by diagnostic assessment and negotiated through teacher collaboration. She provides examples of the diagnostic approach that was used, sample learning profiles, and examines teachers' varying use of the diagnostic information in their classes.

Fox, J. & Hartwick, P. (2011). Taking a diagnostic turn: Reinventing the portfolio in EAP classrooms. In D. Tsagari and I. Csépes (eds), *Classroom-based language assessment* (pp. 47–62). Frankfurt: Peter Lang.

Following on from Fox (2009), this chapter provides details about one EAP teacher's use of information provided to her through a diagnostic assessment approach which generated individual learning profiles for each of her students. There is useful information on the use of: learning profiles, targeted instruction and portfolios in language teaching. Fox and Hartwick highlight the role of student motivation in learning.

Who Are We Assessing? Feedback and Motivation

Activate your learning

- How can knowing more about our students improve the quality of teaching and learning?
- Why is ongoing feedback both important and necessary for quality teaching and learning?
- How can we use our assessment practices to support students' motivation to learn?

6.1 How Can Knowing More about Our Students Improve the Quality of Teaching and Learning?

As we discussed in [Chapter 5](#), the research literature is full of examples of how better understanding of our students' language learning experiences can improve the impact and effectiveness of our teaching and, as a result, enhance learning (see, for example, Gottlieb, [2006](#); Ivanič, 2010; Zamel and Spack, 2004). In this chapter, we again discuss assessment approaches and practices that can help us to better understand our students' learning by addressing the following questions about *who* our individual students are:

- What are my students' learning goals?
- What motivates their learning?
- How can our feedback support their learning?

In this chapter we take a closer look at feedback in assessment practice and how we can shape the feedback to support

our students' learning as part of day-to-day classroom activity. Wiliam (2012) elaborated eight possible ways that students may respond to teacher feedback. We will also examine the role of assessment plans in feedback, and consider the benefits of fully specified tests in providing explicit information to us – and to our students – about language development and its relationship to the learning outcomes identified for a course. Here we will re-examine a number of assessment tools from the point of view of feedback by exploring the following questions:

- What is the feedback potential of a specific assessment tool? How does the feedback differ in relation to the assessment tool?
- What is the potential impact on a student's *learning-in-progress* of a particular approach to assessment?
- How can we improve the quality of our feedback for our students?
- How can we know when the assessment information is being used and understood by our students to inform their learning?

Finally, we will examine the effects of assessment on students' *motivation* and *self-regulation* of learning (Cheng, 2013). As teachers, we engage in assessment in our classrooms every day. Our students know us; we know them. Assessment plays an important role in our relationship with our students and their openness and willingness to learn. We would like to discuss this issue and how we can support students' ongoing learning and development through improved assessment practice.

6.2 Ongoing Assessment: Feedback on Learning in Progress

Our ongoing responses to our students' learning in our classrooms are a mainstay of what teaching is all about. Whether we respond in speaking or in writing, the feedback we provide will shape future performance in fundamental ways. The choices we

make when we develop our overall assessment plan for a course influence feedback potential. The assessment tools we choose and the assessment practices we engage in while teaching a course will shape both the kind of feedback on learning that arises as a result and the information it generates in support of our teaching.

There is an extensive history of research on feedback in language teaching. For example, take a look at Dana Ferris's book *Responses to Student Writing: Implications for Second Language Students* (2003). Ferris provides a comprehensive review of feedback on writing and provides teachers with multiple examples drawn from second language classrooms. In our own discussion on feedback, however, we will consider it solely from the perspective of assessment tools and practices. To begin our discussion, look at the list in [Table 6.1](#), which presents a range of assessment practices and teacher responses. Activity 6.1 helps make the point that our assessment practices and responses (feedback) shape student learning potential.

ACTIVITY 6.1

Answer the following questions about [Table 6.1](#).

- What do you think a student will learn from each of the teacher responses below (1–9)?
- What values are implicit in each of the responses? Are there any hidden messages behind these teachers' responses (i.e., what seems to matter most)?

Fill in the missing information in the table, under the heading 'Student learning'. **Note that in [Table 6.1](#) teachers' responses are in italics.**

Table 6.1 Assessment practices, teacher responses are in italics, students' work is underlined

Assessment practices	Teacher responses (feedback)	Student learning
End-of-unit test	1. 79%/B+ (<i>You missed a lot of information in questions 1–4.</i>)	1.
	2. <i>Look at your answer to question 6, what do you think a doctor would think of your answer? Explain this using information from the in-class readings, and I will award you a bonus point on your test (and raise your mark to an A–).</i>	2.
	3. <i>Let's discuss your test at our learning conference on Tuesday. Were there any questions you thought were unfair or confusing? I'll take your comments into account in reviewing the results with you.</i>	3.
Personal essay	4. Because ^(their) <u>English</u> ^(is) poor, <u>these student</u> ^(s) ^(may) <u>never work as mechanic</u> ^(s) .	4.
	5. Because <u>English poor, these student never work as mechanic.</u> [TRY: <i>Because their English is poor, these students may never work as mechanics.</i>] Explain why each underlined change is needed in order for the sentence to be written correctly. You can earn one bonus point for each correct explanation.	5.
	6. Because <u>English poor, these student never work as mechanic.</u> Why do you think English is so important for students who want to be mechanics? Explain, and I'll award bonus marks for your effort.	6.

Oral presentation of a poster

- | | |
|--|----|
| 7. Poster and presentation: (C); please speak more slowly when you give your next presentation. | 7. |
| 8. Poster: Colourful with attractive visuals; easy to read; a clear topic focus. The number of questions the class asked shows how interested they were in your poster. | 8. |
| 9. Oral presentation: Your outline at the beginning of your presentation was very helpful; you may have rushed at times because of the five-minute time limit so next time practise giving your talk in advance. If you find it is too long, cut it back in advance. You could always use a handout if you feel there is important information to cover but no time to cover it. | 9. |

Now that you have considered what a student might learn from the teachers' responses in [Table 6.1](#), we will examine these responses in greater detail below.

6.2.1 Feedback During a Course

The research on feedback (e.g., Ferris, 2003) suggests that there are few hard and fast rules and little agreement as to the kind of feedback that will have the most impact on our individual students. For example, there are some teachers and researchers who argue that *explicit* feedback on errors in writing is essential. Teachers who subscribe to this school of thought might provide feedback on sentence-level errors in a personal essay like the example in example 4 in [Table 6.1](#). Other teachers might *reformulate* the sentence correctly as in example 5. The

problem is that without an extra step, there is no way for a teacher to know if the explicit feedback and reformulation are actually supporting the student's learning.

As teachers, we spend hours and hours responding to our students' work, but how do our students actually use the feedback? If, as is the case in example 5, we attach a follow-up activity to some of the feedback we provide, and in this case motivate the student to respond by offering a few bonus points, there is a much greater chance that the student will use the feedback productively and increase their learning as a result.

For those who prefer implicit or indirect feedback, example 6 illustrates how a question allows the teacher to reformulate the incorrect sentence. There is also an incentive built into the feedback in example 6 (small but effective), which may motivate the student to draw on the reformulation/feedback in answering the question and earn bonus marks as a result.

Whether the feedback is direct or indirect, positive or negative, limited or extended, the key is our students' interpretation and use of the feedback. Compare the feedback provided to the same student by two teachers responding to the oral presentation of a poster (Table 6.1). What would a student learn from the comment in example 7? What would a student learn from the comment in example 8?

Think back to our discussion of criterion-referenced and norm-referenced assessment in earlier chapters of this book. In example 7, there is only one evident criteria: *speaking slowly* (ostensibly so that the speaker is more comprehensible). Perhaps the students will use this feedback in future presentations, but it is difficult to understand how this one change would improve the overall quality of a poster and an oral presentation. The real information or feedback for the student is the *grade* of C (a norm-referenced indicator that this student gave a more or less average presentation and prepared a more or less average poster). How can the student use this information to improve the quality of future posters and oral presentations? The message is that the student is average. Will speaking more slowly improve his or her next presentation? There is little for the student to work on in order to improve.

In order for feedback to be useful we need to support our students' use of it. Look back at the teachers' responses in [Table 6.1](#). Which of these responses are probably the most useful in supporting the students' future learning? Which of these responses are most likely to result in actions on the part of the student?

Example 8 provides several criteria for the student to consider. Implicit in the teacher's comments are all of the following criteria for evaluation:

Poster

Colourful and attractive display
Text is easy to read
Topic focus is clear
Audience interest is evident

Presentation

Well organized
Observed five-minute time limit
Did not rush

There is much more information required for the student to use in improving the next presentation and poster than that presented in example 7. The following two suggestions will increase the ready flow of useful feedback and improve the quality of our students' work:

- Students will learn the most and perform their best if they know in *advance* the evaluation criteria that will be applied in judging a performance.
- Our feedback will also be improved if we share evaluation criteria with our students *before* their performance.

Better yet, if a teaching context allows, *negotiate* the evaluation criteria for a performance or assessment event with your students in advance. In fact, all teachers should try their best to *work* with their students in creating the evaluation criteria.

- Elicit from them, the criteria that make a presentation and poster engaging, useful and informative.
- Let your students tell you what they look for in a personal essay.
- Give your students the opportunity of identifying the key learning they expect to see in the end-of-unit test (i.e., specific vocabulary they should be responsible for; content they should have studied; questions they should be able to answer.)

Engaging our students in the identification of criteria to be applied in the evaluation of a performance supports their self-awareness, goal-setting, self-assessment and ultimately the quality of their work. It also makes our work as teachers/assessors easier because we have spelled out in advance of the performance exactly what we will be looking for. In [Chapter 7](#) we discuss grading practices in greater detail.

6.2.2 Teachers' Feedback: Conflicting Roles

Some have argued (see, for example, Elbow, [2003](#)) that the problem with teachers' feedback is that it simultaneously responds to two conflicting roles – what Elbow refers to as a teacher's role as *coach* and her parallel role as *judge*. In the role of coach the teacher's feedback is intended as formative and supportive information that will increase a student's future development. Conversely, in the role of judge the teacher's feedback is intended to explain to the student why they are receiving the mark they did. It accounts for the evidence the teacher has identified in arriving at a summative mark. It is an accounting of why an oral presentation or writing assignment got the specific mark it received.

Look back at the teachers' responses to the end-of-unit test in [Table 6.1](#). Which responses suggest the role of *judge*? Which responses suggest the role of *coach*? Do you think that all three responses (1–3) to the end-of-unit test might have been written by the same teacher on one student's test? (If you answered yes to this question, you are correct.)

ACTIVITY 6.2

Take a look at the example below of one teacher's responses to an assignment written by a student in her ESL class. For this end-of-unit assignment on letter writing, students were asked to write a letter to a teacher, Mrs. Barton, to ask for more information about a field trip the teacher is planning for the class to pick apples on a farm. Mrs. Barton had invited parents and/or older sisters and brothers to come with the class on the field trip.

For the assignment, students were asked to write the letter as homework and have it checked before handing it in for marking.

In the example below, the teacher's responses are in italics.

Examine the teachers' responses. Which responses are formative (in the *coach* role)? Which are summative (in the *judge* role)? What is your view of the feedback? How does the teacher attempt to make it useful? Do you think there is too much feedback? Would it be better if there were less? If so, which feedback would you remove? Why?

Example: A student's letter

Ali, although the content and format are good here, there are many errors in your letter. Did you have someone check it before handing it in? That was an important part of the assignment. B-

Remember to check that each sentence has a verb. I added 'is' here. Can you see why you don't need 'I am his father'?

Incorrect: this should be 'Dear Mrs. Barton' as we discussed in class.

Dear Teacher,
Thank you for your invitation. My son is Abdu. ~~I am his father~~. He is in your class and the field trip is coming, to pick apples at a farm. I like this. When do you go? Where *should* I come?

Incorrect. Use 'should' here.

Please send me information on the trip. Abdu can bring it home with him.
Sincerely, Ali

The format of the letter is excellent. Well done.

NEXT STEPS: In class today, rewrite the letter using the feedback.

As the example illustrates, there are no easy answers to questions regarding feedback or the conflicting roles of coach and judge. However, it may be helpful to your students to explain this tension. In responding to Ali's letter, the teacher located feedback relating to her role as judge on the left-hand side of the letter. These comments explain why the student did not receive full marks for his work. The formative comments related to her role as coach are located on the right. If we keep these roles in mind when we are providing feedback to our students on their work, we will separate the comments that explain the judgment from those that support their learning and future work.

The next time you are marking an assignment in your class, you might try locating all the 'accounting' information in the left-hand margin and all the formative feedback on the right. Explain what you are doing to your students and see if, as a whole, the feedback becomes more useful for them.

In the process of working with our students over the duration of a course, we will become attuned to the type of feedback that individual students pay the most attention to. Our ongoing understanding of our students' needs, interests, goals and motivations will shape the formative feedback we provide for their performances in our class. Assessment tools help us to understand who our students are, how they are developing and how we can best support their learning. The more we understand about who our students are, the more effective our feedback will be.

6.3 Assessment and Student Motivation

Assessment and motivation are directly related. What teachers assess and how they assess it have the greatest influence on how students learn – how students see themselves as learners and how they see their learning. In day-to-day

classroom practices, teachers use both assessment *for* learning and assessment *of* learning. This combination requires teachers to use both *summative assessment* (involving the evaluation of learning with a mark or a score) and *formative assessment* (providing quality feedback) as demonstrated above. Both practices have tremendous impact on students, and both are necessary in classroom instruction. To support students' learning and motivate them to learn, three main aspects of assessment highlight the relationship between assessment and motivation:

1. Assessment and motivation require high-quality feedback, that is, feedback needs to be
 - clear
 - focused
 - applicable
 - consistent
 - timely
2. Assessment and motivation address individual student needs, allowing for
 - recognition of individual student differences
 - acknowledgment of students' unique prior knowledge and experience
 - increased use of self-assessment
 - encouragement of self-directedness
 - increased student self-reflection
 - increased autonomy (i.e., taking responsibility for learning)
 - setting goals for learning
3. Assessment and motivation engage students by
 - making assessment real (i.e., contextual to students)
 - offering choices in assessment tasks, tools and procedures
 - supporting their connection with/sense of belonging to a learning community
 - including them in assessment processes
 - creating collaborative assessment practices where students see teachers as allies

6.3.1 Assessment, Learning and Self-Determination

The role of assessment in motivating students to learn can be traced to many theories of motivation (Dörnyei, 2001). There are theories focusing on reasons for engagement in tasks; theories that focus on integrating expectancy and value constructs; and theories that integrate motivation and cognition. Particularly fitting for the assessment context is self-determination theory (SDT), introduced by Ryan and Deci (2000). This theory ranges from self-determined forms of intrinsic motivation to controlled forms of extrinsic motivation and, finally, to amotivation, depending on degrees of self-determination. Testing and assessment policies are mostly based on the concept that rewards, punishments and self-esteem-based pressures are effective motivators for learning. SDT thus fits well in the assessment context.

Ryan and Deci (2000) identified four types of motivation (from the most self-determined to the least self-determined): (1) intrinsic motivation, (2) self-determined extrinsic motivation, (3) non-self-determined extrinsic motivation and (4) amotivation:

- *Intrinsic motivation* refers to motivation that makes one feel engaged in an activity that is inherently interesting or enjoyable. If the assessment practices teachers employ make students feel learning is interesting and enjoyable (make assessment real to students), then students will be intrinsically motivated.

In contrast,

- *Extrinsic motivation* refers to motivation that is instrumental in nature. In other words, the activity is a means to an end, but the requirement to engage in the activity is imposed on the individual and may not even be something they feel like doing.

However,

- *Self-determined extrinsic motivation* is present when individuals participate in an activity voluntarily because they perceive the activity is valuable and important. It is extrinsic because the reason

for participation is not within the activity itself but is a means to an end, and at the same time it is self-determined because the individual has experienced a sense of direction and purpose in acting. If the assessment practices teachers employ make students feel that their learning is an important part of the process for self-improvement, students may have self-determined extrinsic motivation to learn.

Further,

- *Non-self-determined extrinsic motivation* occurs when individuals' behaviours are regulated by external factors such as rewards, constraints and punishment. This type of motivation is extrinsic because the reason individuals participate in an activity lies outside the activity itself (e.g., family pressure) – that is, the behaviour is not self-determined. Individuals feel an obligation to engage and are regulated by external rewards, constraints, or punishment. If the assessment practices teachers employ make students feel that their learning is driven by external rewards such as bonus marks in grading or praise from teachers, students are non-self-determined extrinsically motivated.

Finally,

- *Amotivation* is the absence of both intrinsic and extrinsic motivation. It is a state in which an individual lacks the intention to act. In this case, students may feel that they have no control over their actions or that to act is meaningless or without value or importance. When assessment fails to motivate students to learn either intrinsically or extrinsically it is essentially a useless activity.

Ryan and Deci's SDT categorizes motivation along a continuum from self-determined forms of intrinsic motivation to controlled forms of extrinsic motivation, and finally to amotivation. They connect motivation to an individual's degree of engagement. Testing and assessment policies are often based on the concept that rewards, punishments and self-esteem-based pressures are effective motivators for learning. SDT helps to account for the complexity of individual perceptions of assessment, motivation and learning.

To learn more about this theory in relation to assessment, see Cheng and colleagues (2014). This study examined test-takers' motivation, test anxiety and test performance across a range of social and educational contexts in three high-stakes language tests: the Canadian Academic English Language (CAEL) Assessment in Canada; the College English Test (CET) in the People's Republic of China; and the General English Proficiency Test (GEPT) in Taiwan. This study used a questionnaire exploring motivation, test anxiety and perceptions of test importance and purpose to test-takers in each of the three contexts. A total of 1281 valid questionnaire responses were obtained: 255 from CAEL, 493 from CET and 533 from GEPT. Questionnaire responses were linked to each test-taker's respective test performance. The results showed a direct relationship between test performance and motivation illustrating complex interrelationships of test-takers' motivation and test anxiety in their test performance. Differences in motivation and test anxiety also emerged with regard to social variables (i.e., test importance to stakeholders and test purposes). Further, motivation and test anxiety, along with personal variables (i.e., gender and age), were associated with test performance. Given that motivation and test anxiety have typically been examined separately and in relation to a single testing context, this study addresses an important research gap and provides important evidence for the relationship between assessment and motivation.

ACTIVITY 6.3

Reflect on your own assessment practices in motivating students to learn. Which assessment practices may lead to:

- intrinsic motivation?
- self-determined extrinsic motivation?
- non-self-determined extrinsic motivation?
- amotivation?

6.3.2 Assessment Motivation Strategies

An effective way in which teachers can motivate their students is by involving them in the process of assessment through various procedures. For example, teachers can involve students in setting learning outcomes or achievement goals. Although the main responsibility for creating these learning outcomes rests with the teacher, and is usually guided by the curriculum and standards, communicating these goals to students is one effective, practical way of enhancing achievement. Students can collaborate with the teacher to develop additional self-directed outcomes of learning. 'If students play even a small role in setting the (learning achievement) target ... we can gain considerable motivational and therefore achievement benefits' (Stiggins, 2008, p. 244). Stiggins suggests that students keep learning logs as a way to engage them in assessment, increase motivation and help them to reflect on and recognize their own improvement. Receiving frequent feedback from the teacher can also raise students' awareness of progress.

Another way to motivate students is to involve them in designing assessment criteria. McMillan (2014) discusses the importance of creating learning targets, which involves teachers and students specifying: (1) what a student is to know and/or do as a result of instruction; and (2) the criteria for evaluating the performance. We discussed this relationship, between learning outcomes and assessment tasks in [Chapter 2](#). This process of creating the criteria needs to be a collaborative process among teachers, and among teachers and their students, as good-quality assessment is team-based by nature (whether it is about creating a large-scale test, a small-scale classroom test, or classroom assessment criteria). For example, the process of creating assessment criteria can be carried out by using a checklist if the learning goals are the foundation of learning that is, what every student has to know and/or know how to do. This process can also be carried out using an existing rating scale or a rubric, which teachers can use to specify level(s) of achievement. Irrespective, such creation should be

carried out by teachers *with their students*, so students understand the goals and learn how to achieve those goals as specified by the criteria.

Teachers can then choose various assessment tools to use – for example, those specified in [Chapters 3 and 4](#). (For an overview of assessment tools and test formats see the Appendix.) Knowing which tools to use in the classroom requires teachers to ask a number of questions. For example:

1. Which assessment tools provide the most useful feedback for my students?
2. Which assessment tools are most likely to be motivating for my students?
3. Which assessment tools are most likely to connect with or activate my students' prior knowledge – be that cultural, social, or academic?
4. Which tools are easier to design and/or score?

The first two questions are essential in supporting student learning in general. Question 3 acknowledges that in many language teaching contexts our students come from various countries. Question 4 reminds us that it is always important to consider the practicality of assessment. Assessment has to support both teaching and learning in order for quality assessment to take place in instruction. Neither teachers nor students should be 'buried' under assessment (see the points raised regarding the frequency of assessment events in [Chapter 1](#)). At the same time, students are the key stakeholders in assessment. If students do not want to learn (motivation), do not know how to learn (feedback in relation to criteria), or do not have the awareness or metacognitive strategies to learn (assessment as learning), whatever teachers do is not going to support them.

When students and teachers engage in conversations about assessment, students are encouraged to consider their own learning and development, which aids in and supports the learning process. Motivation will continue if students witness and reflect on their growth in relation to learning goals.

Research shows that when students understand and apply self-assessment skills their achievement increases (Black and William, 1998), and self-assessment plays a significant role in increasing students' motivation to learn. Through self-assessment, students directly observe their own improvement and therefore are more motivated to achieve. By involving students in the assessment processes, teachers encourage students to create a sense of internal responsibility for their achievement. Stiggins (2005) remarks that students 'must take responsibility for developing their own sense of control over their success' (p. 296). This, in turn, leads to greater motivation and greater academic success.

This process is sometime referred to as *assessment as learning*. Assessment *as learning* helps students to personally monitor their learning processes, by involving students in self-monitoring or self- and peer-assessment activities, as well as by using feedback to effectively support their own learning. Assessment as learning is the use of a task or an activity to allow students the opportunity to use assessment to further their own learning. Self- and peer-assessments allow students to reflect on their own learning and identify areas of strengths and needs. These tasks offer students the chance to set their own personal goals and advocate for their own learning. This is also important for students from various educational backgrounds because it creates an opportunity for them to connect their prior learning to the assessment tasks in their current classrooms. Research has demonstrated that, without this opportunity to connect, students will not progress effectively.

As mentioned above, if students do not want to learn (have no motivation), do not know how to learn (cannot relate feedback to learning criteria), or do not have the awareness or meta-cognitive strategies to learn (do not understand the potential of assessment as learning), whatever teachers do will not help such learners.

Existing literature has shown that the information that the students internalize from classroom assessment fuels their

ACTIVITY 6.4

Check that you have used the following six principles to enhance the relationship between motivation and assessment in your own assessment practices. Have you

- shared the learning outcomes/goals with your students?
- helped your students understand the standards or outcomes they are working towards?
- involved your students in assessment?
- provided helpful feedback for your students?
- created a positive learning atmosphere in the classroom?
- integrated teaching, learning and assessment?

These principles will help us to learn about who our students are and how best to support their learning through the use of assessment – for example, placement testing, needs analysis, diagnostic assessment. Ongoing assessment and the feedback on learning that it generates play a pivotal role in supporting our students' motivation to learn.

learning system (Stiggins, 2005), and formative assessment in particular is one of the most powerful ways by which to enhance student motivation and achievement. Here are three further assessment examples of how we can support student development:

- Teachers can share with their students the assessment criteria, or even better, create the assessment criteria with their students. Both assessment processes are valuable practices for students to see teachers as allies in their learning. Teachers can then use the criteria as feedback for students on their performance. Teachers should initiate this assessment process in small steps, as we need to consider the learning backgrounds our students come from.

This may be a very unfamiliar practice for some of our students. Ultimately, however, practice helps to bring all students' learning in line, that is, sharing or understanding the same learning goals, or 'destinations', even though the pathways to get there are unique and individual.

- Teachers can use student work as exemplars to illustrate levels of performance. This is best achieved through collaboration, as teachers work together to build a collection of student work over time. In addition, having teachers working together can ensure consistency of assessment, that is, teachers must reach agreement about what is 'good work', and most importantly what such work looks like in terms of language use and the demonstration of knowledge or control of content.
- Teachers can let their students take more responsibility for their own learning, that is, the use of *assessment as learning*. Many of our students are far away from their families, so taking responsibility for their own learning is a big but essential first step towards being successful. That success will have a long-term impact on their lives, and teachers can demonstrate this through the use of self- and peer-assessment tasks.

We can help our students by teaching them to be more analytic about their own learning, by giving them class time and a structure to examine their own work in relation to previously explained criteria, and by clarifying how they can improve their work. For example, we can help our students to identify their mistakes by providing them with item analyses of their tests or rubric scored projects. We can involve our students in thinking about their mistakes. Give them time to consider why they made the mistake, and help them to understand what they will do differently next time.

Thinking About Doing Better (Sindelar, 2015) is an example of helping students to analyse their mistakes on a **forced-choice** or short answer test. Each student has a form and works in a group of two or three. After students analyse their mistakes with a partner, they are asked to set some learning goals. When students examine what they are doing well and what they

need to improve on, they are beginning the process of setting their own goals for learning. Students should be encouraged to set small, realistic goals as the most useful goals are those that reflect steps along the way – not just the final outcome. Taking small steps helps students to self-monitor their way to success.

Thinking about Doing Better

Directions: Identify three items (questions or problems) you missed on the test. Then with a partner decide why you missed the question and how you could fix it. Next, with your partner write down what you will do differently the next time you encounter a similar question or problem. Budget your time to eight minutes per item.

Item number	Why I got it wrong	How I can fix it	What I will do next time

My Goals

Directions: By yourself write down two learning goals and the activities you will engage in to reach them. If you need help identifying activities, ask your partner or your teacher.

Goal One:	Activities for Goal One:
Goal Two:	Activities for Goal Two:

6.4 Looking Back at Chapter 6

The importance of the relationship between assessment and motivation in supporting student learning has been increasingly examined in education. Teachers can exert influence on students' motivation through instruction, assessment and feedback. Here are some additional questions for reflection and discussion. Write a brief response to each of the following questions in bullet points. Then, if appropriate, discuss your responses with others.

1. Most language teachers also have experience as language learners. Thinking back over your experience as a language learner, has anyone ever asked you to describe your *needs*? If yes, can you remember the context? How did you respond? How did it make you feel? If no one asked you about your needs, why do you think this was? What are the implications?
2. When you mark a student's paper, how do you deal with the conflicting roles of coach and judge?
3. What strategies have you used to encourage your students to use the feedback that you provided? Were they successful?

Suggested Readings

Cheng, L., Klinger, D., Fox, J., Doe, C., Jin, Y. & Wu, J. (2014). Motivation and test anxiety in test performance across three testing contexts: The CAEL, CET and GEPT. *TESOL Quarterly*, 48(2), 300–30. doi:10.1002/tesq.105

This is one of the first empirical studies that examined the role that motivation and test anxiety play in students' language testing performance across a range of social and educational contexts. This study provides teachers with actual data from test-takers in three testing contexts to understand the relationship between assessment and motivation.

Stiggins, R. J. (2008). *Student-involved assessment for learning* (5th edn). Upper Saddle River, NJ: Merrill/Prentice Hall.

This leading book for assessment focuses on showing teachers how to develop assessments that accurately reflect student achievement and how to use those assessments to benefit – not merely to grade – student learning. It examines the full spectrum of assessment topics, from articulating targets, through developing quality vehicles, to communicating results effectively – with an exceptionally strong focus on integrating assessment with instruction through student involvement.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14.

Understanding the impact that assessment has on learning requires a broader focus than the feedback intervention itself, particularly the learner's responses to the feedback, and the learning milieu in which the feedback operates. In this article, different definitions of the terms 'formative assessment' and 'assessment for learning' are discussed and subsumed within a broad definition that focuses on the extent to which instructional decisions are supported by evidence.

When We Assess, How Can We Use Assessment to Move Forward?

Activate your learning

- How can we use assessment to move forward – to increase its positive impact on our teaching and enhance our students' learning?
- What is in a grade? How can we use grading to accurately reflect learning?
- What role should test preparation play in our classrooms? Should we prepare our students to take high-stakes tests?
- What can we learn about high-stakes, large-scale testing from our students' experiences as test-takers?

7.1 Using Assessment to Move Forward

If we, as educators, are to first and foremost 'do no harm' (Taylor and Nolen, 2008, p. 10), we need to continue to focus on the relationship between teaching, learning and assessment, and rethink some of the assumptions that we hold about testing and assessment. The academic race to be the smartest, most skilled student in the class does not place the focus of learning on improvement or the act of learning itself, but rather on achievement and outcomes alone. For assessment to be effective and to enhance, not harm, students' learning, students must compete with themselves to continue to improve, and teachers should use assessment events to help students to develop effective learning strategies that will serve them beyond the classroom.

In our discussions so far, we have examined, *why* we assess (Chapter 1), *what* we assess (Chapter 2), *how* we assess in the classroom setting (Chapter 3), *how* we develop high-quality tests (Chapter 4) and *who* we assess from the point of view of needs analysis, placement and diagnostics (Chapter 5), and from the point of view of feedback and motivation (Chapter 6). In this final chapter we ask, *when* we assess and *how* we can use assessment to move forward. Throughout, we have emphasized that in order to ensure high-quality classroom assessment practices, that is, those practices that will support and enhance student learning, we need to recognize the following:

- Assessment takes place during instruction and continuously.
- Knowledge and skills should not be assessed in isolation.
- Students should be informed about and involved in all assessment events (whether our purpose is assessment *for*, *as*, or *of learning*).
- When, as teachers, we actively question, reflect on and learn more about assessment, we increase its quality and positive impact.
- As teachers, we should use a combination of assessment *for* learning, assessment *as* learning and assessment *of* learning.

Teachers' classroom assessment plays a central role in and inevitably influences teaching and learning (for examples, see Cheng, 2014; Colby-Kelly and Turner, 2007; Fox and Hartwick, 2011). Stiggins (2005) notes that, despite its significance, over the last decade classroom assessment has become a 'victim of gross neglect' (p. 10), receiving little attention in terms of its nature, conduct and use. In Chapter 1, we defined four fundamental aspects of classroom assessment activities to include *events*, *tools*, *processes* and *decisions*.

- *Assessment events*, such as an oral presentation or a listening activity, can support students when the events occur with the right frequency, so that the teacher knows whether instruction is successful, which areas need more instruction, and which student or group of students may need additional support.

- *Assessment tools* can support student learning when the tools provide students with clear ideas about what is important to learn and the criteria or expectations for ‘good’ work, and when assessment goals are aligned with learning outcomes and instructional goals.
- *Assessment processes* can support students’ views of their teachers as allies in their education. Feedback can help students to focus and better understand the requirements of a task. Feedback increases students’ self-awareness and their ability to set meaningful and appropriate goals for their learning.
- *Assessment decisions* can support students’ learning when grades accurately reflect what students know and can do. We make a range of decisions based on the results of our assessment. These decisions range from micro-level course decisions, such as what we need to do more or less of in a follow-up lesson, to macro-level decisions, which have important (even life-changing) consequences for our students, such as deciding which class a student should be placed in or whether a student can be admitted into a university.

In this final chapter, we focus on the fourth fundamental aspect of classroom assessment activities: *assessment decisions*. We will discuss the important yet most complex role that grading plays in teaching and learning. We examine the potential dilemmas that may arise from the conflicting roles of *formative assessment*, which supports and informs learning, that is, teachers as *coaches*, and *summative assessment*, which measures, ranks and rates quality, that is, teachers as *judges* (Elbow, 1986). We unpack the salient issues in grading by examining three grading scenarios. We then draw attention to the influence of large-scale testing on teaching and learning in our daily classrooms. We focus on the common phenomenon of test preparation. Test preparation is another potential dilemma of formative and summative assessment in our day-to-day teaching, that is, how we, as teachers, support students to take large-scale standardized testing. Finally, we discuss the essential role that students play in assessment and the importance of listening to them as test-takers (Cheng and DeLuca, 2011; Fox and Cheng, 2015).

ACTIVITY 7.1

Let's revisit three key questions we need to respond to in our considerations of formative and summative assessment:

1. Should formative assessment results be included in grading for summative purposes?
2. Is it possible to provide our students with support for taking a high-stakes, large-scale assessment without narrowing the scope and depth of our teaching? In other words, can we embed large-scale assessment within the instructional programme of the classroom in a meaningful way?
3. Why should we listen to test-takers in response to their experiences with high-stakes, large-scale assessment? How can their voices inform our teaching?

Reflect for a moment on these questions about a class (and students) you are currently teaching (or, if relevant, on a class you are currently taking). Discuss your responses with colleagues if appropriate.

7.2 Grading

Grading, the process of summing up student achievement using a numerical or ordinal scale, is a complex evaluative practice that requires teachers to make judgments about student learning. Grades are used, most notably, to make public statements to students, parents and other stakeholders about student achievement. Thus, grading is one of the most high-stakes classroom assessment practices, with significant consequences for student self-perception, motivation for learning, prioritization of certain curriculum expectations, parental expectations and social relationships (Brookhart, [2013](#)).

Currently, the lack of research on grading practices provides unprecedented challenges for grade interpretation and grade use across educational systems (DeLuca, Chavez and Cao, 2012; Sun and Cheng, 2014).

Despite the significance and impact of grading in classroom teaching and learning, researchers have long recognized the lack of theoretical grounding for teachers' grading practices. Specifically, researchers have called for an examination of grading practices using contemporary validity theories (Brookhart, 2003, 2013; Moss, 2003) instead of traditional psychometric approaches to validity, which are ill-fitted to classroom assessment practices as these traditional approaches rely on standardized assessment protocols and large-scale data sets. In contrast, contemporary validity theories aim to integrate multiple perspectives into a socio-culturally situated argument on the alignment of grading practices, values and consequences.

ACTIVITY 7.2

Before we discuss grading in more details, answer the questions below. Again there is no right or wrong answer. Your assessment practices reflect the context in which you teach. Remember the key is to involve your students in the assessment. As we have stated in this chapter, assessment refers to 'all those activities undertaken by *teachers*, and by *their students* in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged' (Black and Wiliam, 1998, p. 2, emphasis added).

After you have completed the activity, you may want to know what ESL/EFL teachers in other contexts reported about their grading practices; for this, see Cheng and Wang (2007).

Questions:

1. Do you prepare your own marking guide (or scheme, or system)?
2. When do you prepare your marking guide?
 - ☐ At the time you develop your assessments (i.e., before the students have responded)
 - ☐ Just before you start marking the students' work (i.e., after the students have responded)
 - ☐ After reading a sample paper
3. What type of marking guide do you use when you mark your students' performance?
 - ☐ Analytical scoring (i.e., do you give marks for different components of the essay or a spoken presentation?)
 - ☐ Holistic scoring (i.e., do you give one mark for overall impression?)
 - ☐ Rubric scoring (e.g., do you match essays or presentations to one of four performance descriptions that differ according to completeness and correctness?)
4. Regardless of the marking guide you use, do you give your students written comments about their strengths and weaknesses?
5. Do you usually tell students the scoring criteria (expectations of their performance) before you assess them?
6. Do you ever involve your students in
 - ☐ preparing marking guides
 - ☐ marking other students' work
 - ☐ marking their own work
7. How quickly do you get the marks back to the students? Please describe the normal time taken to get a score or mark to your students.

7.2.1 Research on Grading

Research on grading has a long history in education. In the past, educators were primarily concerned with the reliability and standardization of teachers' grading practices. Recent research has further explored factors that influence and shape teachers' grades related to both achievement and non-achievement (e.g., effort and behaviour) (Guskey, 2011; Randall and Engelhard, 2010; Yesbeck, 2011). Teachers try hard to be *fair* to their students as they juggle their dual roles of judge and coach (Bishop, 1992; Fox and Hartwick, 2011; Sun and Cheng, 2014). However, these roles may be in direct conflict in grading practices, and thus may jeopardize the validity of grade interpretation and use.

McMillan (2008) pointed out that even when teachers use the same grading scale and the same grading guidelines, still there is little consistency in teachers' grades (Brookhart, 2003; Liu, 2013). Based on interview data with secondary and elementary classroom teachers in Virginia, McMillan and Nash (2000) proposed a model of teachers' classroom grading decision-making including both internal and external influencing factors. The most salient internal factor was the teachers' philosophy of teaching and learning (as we discussed in [Chapter 1](#)). The major external factors were identified as district grading policies, mandated state-wide learning standards, high-stakes tests and parents' expectations (as we discussed in [Chapter 2](#)). This model is also supported by studies conducted in the context of teaching English internationally. For example, Cheng and colleagues (Cheng, Rogers and Hu, 2004; Cheng, Rogers and Wang, 2008; Cheng and Wang, 2007) investigated teachers' assessment and grading practices in Canada, in Hong Kong and in China. These studies show that teachers' grading preferences are influenced by their values about assessment, their teaching experiences and training, their instructional contexts and the dominance of large-scale testing.

McMillan (2008) argued that one of the most difficult issues in grading is how to deal with non-achievement factors such as effort, work habits and motivation. He refers to these

factors as academic enablers. Teachers consider these enabling factors in grading because they are traits that teachers cultivate and regard as important for student achievement. Zoeckler (2007) used a theoretical framework of truth, worthwhileness, trust and intellectual and moral attentiveness to examine how US English language teachers attempted to assign fair grades while weighting both achievement and non-achievement factors. The results of this study indicate that grading was influenced by local grading systems, teachers' perceptions of student effort, and their concerns for moral development. Similarly, in a survey of 516 teachers in the United States, Randall and Engelhard (2010) found that under most circumstances, teachers abided by the official policies of the participating school district, assigning grades based primarily on achievement. However, in borderline cases, teachers tended to value other characteristics such as ability, behaviour and effort.

While the studies above focused on *what* is graded, other studies on teachers' grading practices have focused on *how* grades are interpreted by key stakeholders: teachers, students and parents. Friedman and Frisbie (1995) conducted a content analysis of report cards obtained from 240 elementary, 96 middle and 140 high schools in Wisconsin and examined how the various characteristics of those report cards contributed to or distracted from the core meaning of grading information. Studied less often are the consequences of grade use, especially for students. Thomas and Oldfather (1997), in presenting a theoretical discussion of their seven-year research agenda, argued that grade meanings are socially constructed, situation-specific and subject to multiple interpretations. They identified the potential relationship between teachers' grades and students' motivation to learning. Figlio and Luca (2004) conducted a unique study by investigating the extent to which the grades distributed by teachers corresponded to their students' performance on the Iowa Test of Basic Skills in the United States, and also how grading affects students' learning. With the global shift toward assessment *for*

learning, studies that further examine the relationship between teachers' grading and student motivation to learn are needed. Waltman and Frisbie (1994) investigated parents' interpretation of grades, and discovered an overwhelming messiness of school-to-home grade communication and inconsistency between teachers and parents in the way grades were interpreted.

As identified by Brookhart (2013), 'grades are acknowledged to have multiple meanings and multiple contexts, including social and legal contexts' (p. 265). From a sociocultural perspective, the grading decisions convey teachers' values, beliefs and assumptions about teaching and learning, which are rooted in and shaped by the sociocultural and historical contexts in which teachers work.

In summary, little is known about the underlying teaching and learning values that contribute to grade decisions. For example, the concept of effort and behaviour as part of a grade is highly valued in Chinese learning cultures (Cheng and Curtis, 2010; Carless, 2011) whereas grading only on achievement is more widely endorsed in measurement-driven communities in Canada and the United States (O'Connor, 2007; Wormeli, 2006; Simon et al., 2010). Furthermore, we do not have sufficient evidence of the consequences of grade use for teachers, students, parents and principals. Given the complexity and diversity of grading practices, we argue that research on grading needs to move beyond its traditional emphasis on reliability; rather, grading research must examine the validity of grading practices. It is time to use contemporary validity theories (Brookhart, 2003; Messick, 1989; Kane, 2006) to examine the values and consequences of grades as generated and situated *within* and *across* learning contexts.

7.2.2 Grading Scenarios

In summary, grading is a complex evaluative practice that requires teachers to make judgments about student learning. Grades are used, most notably, to make public statements to

students, parents and principals about student achievement. Due to this high-stake nature of grading, we as teachers work continuously to make our grading practices, thorough, fair and meaningful.

ACTIVITY 7.3

Below are three scenarios that can help you to explore your own grading decision-making. The three scenarios were created by Sun and Cheng (2014). Working alone or with a peer, follow the steps *sequentially* in responding to these scenarios. Please complete the first two steps before you do Step 3.

1. Read through the scenario and decide what you would do in each case by choosing A, B, or C.
2. Write a few notes to explain your rationale as to why you would choose A, B, or C.
3. Compare your responses to those of a group of Chinese teachers of English reported in Sun and Cheng (2014).

Scenario 1: focusing on working to ability

Scenario 1

Wang Hong, one of the students in your class, has high academic ability, as shown by her previous work, test results, reports of other teachers and your own observations. As you look over her work for the grading period you realize two things: the quality of her work is above average for the class, but the work does not represent the best that she could do. The effort she has shown has been minimal, but because of her high ability her work has been reasonably good. In this situation, you would:

- A. Grade Wang Hong on the quality of her work in comparison to the class, without being concerned about the amount of work that she has done.
- B. Lower Wang Hong's grade because she did not make a serious effort in your class; she could have done better work.
- C. Assign Wang Hong a higher grade to encourage her to work harder.

Now you can compare your thinking with that of the group of Chinese teachers in Sun and Cheng's study (2014). Their responses are listed in the column to the left. Note that N indicates how many of the teachers in the study chose A, B, or C; and % shows the relative percentage.

N	%	
69	23.6	A. Grade Wang Hong on the quality of her work in comparison to the class, without being concerned about the amount of work that she has done.
123	42.1	B. Lower Wang Hong's grade because she did not make a serious effort in your class; she could have done better work.
100	34.2	C. Assign Wang Hong a higher grade to encourage her to work harder.

Scenario 2: focusing on missing work

Scenario 2

You are the English teacher of a class with varying ability levels. During this grading period, the students' grades are based on quizzes, tests and homework assignments. Li Wen has not turned in any homework assignments despite your frequent reminders. His grades on the quizzes have ranged from 60% to 75%, and he received a D on each of the tests. In this situation, you would:

- A. Assign Li Wen a grade of 0 for the homework assignments and include this in the grade, thus giving him an average of F for the grading period.
- B. Ignore the missing homework assignments and assign Li Wen a D.
- C. Ignore the missing homework and assign Li Wen a C.

Now you can compare your thinking with that of the group of Chinese teachers in Sun and Cheng's study (2014). Their responses are listed in the column to the left. Note that N indicates how many of the teachers in the study chose A, B, or C; and % shows the relative percentage.

N	%	
125	43.7	A. Assign Li Wen a grade of 0 for the homework assignments and include this in the grade, thus giving him an average of F for the grading period.
92	32.2	B. Ignore the missing homework assignments and assign Li Wen a D.
69	24.1	C. Ignore the missing homework and assign Li Wen a C.

Scenario 3: focusing on improvement

Scenario 3

You are the English teacher of a class which consists of students with varying ability levels. For this class you set two exams in each term. As you compute Zhang Lin's grade for this term, you see that on the first exam, he obtained a score equivalent to a D and on the second exam, a B. In this situation, you would:

- A. Assign Zhang Lin an overall grade of C, which is the average of his scores on the two exams.
- B. Assign Zhang Lin an overall grade of B, noting that there was improvement in his performance.
- C. Grade Zhang Lin on the quality of his work in comparison to the class, without being concerned about his improvement.

Now you can compare your thinking with that of the group of Chinese teachers in Sun and Cheng's study (2014). Their responses are listed in the column to the left. Note that N indicates how many of the teachers in the study chose A, B, or C; and % shows the relative percentage.

N	%	
27	9.9	A. Assign Zhang Lin an overall grade of C, which is the average of his scores on the two exams.
237	86.8	B. Assign Zhang Lin an overall grade of B, noting that there was improvement in his performance.
9	3.3	C. Grade Zhang Lin on the quality of his work in comparison to the class, without being concerned about his improvement.

7.3 Impact of Large-Scale Testing: Test Preparation

Large-scale testing has been used more and more in educational systems across countries for high-stakes purposes of accountability, gatekeeping and policy-making (e.g., Cumming, 2009; Cheng, 2014). The results of these tests, regardless of subject areas being tested and their test constructs, are used to make inferences on students' proficiency based on their performance on a sample of items drawn from a whole content domain of knowledge. These results from test scores or test

performance are used as indicators of students' academic achievement, and are often directly related to a variety of high-stakes decisions, from students obtaining degrees and academic advancement, to teachers attaining professional certification and promotion, to school boards funding. Due to the *snapshot* nature of testing as indicators of students' academic achievement and also the relationship between high-stakes decision-making and students' test performance, large-scale testing has affected what and how teachers teach and especially what and how students learn. By *snapshot*, we mean learning represented as a test score derived from a single testing event and at one specific time during a student's learning. Consequently, 'teaching has been inordinately *skewed toward test preparation*' (Madaus, 1988, p. 36, emphasis added), and practices (pedagogy) and principles (appropriateness/ethicality) of preparing students to take tests have thus gained increasing attention in many fields, including curriculum, educational measurement and language assessment.

7.3.1 Test Preparation: Defined

Three key terms have been used, in various contexts, to define preparing students to take tests – coaching, teaching to the test and, more commonly, test preparation.

- *Coaching* is applied to commercial programmes operated as extracurricular activities that students participate in outside of school, and refers to short-term instructions targeted to improve students' test performance on a particular examination and does not necessarily have the aim to improve their academic skills. Therefore, the term coaching usually has a negative connotation as students can be coached to maximize their test performance, but may not have a corresponding increase in their academic abilities per se. Coaching (as a term) is often used and studied in the educational measurement field, which examines whether and to what extent coaching might influence students' test scores.
- *Teaching to the test* is often used in school settings and is usually discussed in the curriculum literature. A narrow definition of

teaching to the test implies that teachers' instruction focuses narrowly on actual tests and items with an aim to improve students' test scores, or simply item-teaching. However, a broader definition may mean that teachers build their instruction around the learning outcomes, which are sampled by tests in order to enhance students' knowledge, capability, and performance both in the test and in real terms in the classroom and beyond.

- *Test preparation* can be applied to contexts both inside and outside school activities and is most often used in research. It is more neutral and inclusive, and is defined as a variety of activities, which review and develop capacities or content sampled by tests, and practice test-taking skills, in an attempt to improve students' test scores (e.g., Crocker, 2006).

These terms differ in focus depending on the contexts where they are used and studied. Some researchers also use the delivery mode to understand and study test preparation. Test preparation can be school-based, commercial-based, computer-based and book-based according to the delivery mode (e.g., Montgomery and Lilly, 2012). School-based test preparation is integrated into the curriculum and offered by classroom teachers within school settings. Commercial-based test preparation is fee-charged, short-term instruction operated by commercial business interests or agencies with the claim of effectively increasing students' test scores. Computer-based test preparation is test preparation whose content is administered through a computer, where students have control over the speed, and the amount of test preparation they choose to engage in. Book-based test preparation is structured on commercial publishers' practice books and test publishers' official guides.

7.3.2 Test Preparation: Examined

Since the 1950s, a number of educational researchers have investigated the effects of commercial test preparation courses (coaching) on test scores. Inspired by these early investigations, more researchers have attempted to reach an understanding of test preparation from their particular

research perspectives. Test preparation has been studied in a range of three specific areas: (1) ‘teaching to the test’ and ‘measurement-driven instruction’ in the field of curriculum, because such practice may narrow curricula (e.g., Madaus, 1988; Popham, 2001); (2) ‘test impact’ and ‘consequential validity’ in the field of educational measurement, because of its effects on test scores and test score uses (e.g., Haladyna and Downing, 2004); and (3) ‘washback’ in the field of language testing because of its influence on language teaching and learning (e.g., Alderson and Hamp-Lyons, 1996; Cheng and DeLuca, 2011; Green, 2007). Studies on test preparation across these fields were rooted in a common concern – the influences of test preparation on the accuracy of students’ test scores as an indicator of their academic competency, or theoretically speaking, the influences of test preparation on the *validity* of test scores.

As we have discussed throughout the chapters of this book, validity refers to the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores. Simply put, if a plausible interpretation of a student’s mastery level can be derived from the test score, the validity of the test score is achieved. However, if a test score is not an actual representation of a student’s mastery, the interpretation and inference based on this test score cannot be accurate; therefore, the validity of this test score is threatened.

Messick (1996) has stated that if test preparation emphasizes the instruction of test-wiseness strategies, students might answer some test items correctly using test-wiseness strategies rather than their actual knowledge. **Test-wiseness (TW)** is defined as the ability to respond advantageously to items containing clues and, therefore, to obtain credit without the ability, skill, proficiency, or knowledge of the subject matter being tested. Strategies include, for example, choosing the longest answer amongst multiple choice distractors, when distractors are of unequal length; avoiding any distractors with the words

‘all’ or ‘every’; ruling out as many alternatives as possible and then guessing from the ones that remain. In this case, students’ increased scores cannot represent the equivalent improvement of students’ ability, proficiency, or knowledge, and this type of test preparation threatens the interpretation and use of the increased test scores.

Meanwhile, if students are prepared to become familiar with the test, (e.g., how many sections there are in the test, what they are expected to do in each section, how much time is allocated to each section), and if students are prepared to develop strategies to cope with psychological influences on test performance (e.g., anxiety reduction), they might perform at a level that is more indicative of the level of their mastery. This type of test preparation thus improves the validity of inferences drawn from test scores as it minimizes construct irrelevant variance.

Since the 1950s, a large number of educational measurement studies have examined high school students’ test score gains resulting from coaching programmes (e.g., Montgomery and Lilly, 2012). The results of these studies have shown that coaching programmes *increase* student test scores by 20 to 30 points on vocabulary and maths subtests of the Scholastic Aptitude Test (SAT) (Montgomery and Lilly, 2012). Studies in language testing investigate this issue in a slightly different way; instead of measuring effect sizes in the unit of score gains, these studies have looked at whether students who take test preparation courses perform significantly better in comparison to students who do not. These studies have shown conflicting results: (1) significantly better performance of students taking test preparation courses (e.g., Hayes and Read, 2004); and (2) no significant advantage for students taking test preparation courses (e.g., Doe and Fox, 2011; Green, 2007). Therefore, whether test preparation can significantly influence test scores is still under debate and needs further empirical explorations (Montgomery and Lilly, 2012).

7.3.3 Test Preparation: Pedagogical Implications

Although the degree to which test preparation influences test scores is still under exploration, it is a shared understanding among researchers across educational fields that a criteria or a code of practice should be established to ensure the appropriateness or ethicality of test preparation practices. In other words, guidance is important, practically speaking, to enable teachers to perform preparation activities, which are appropriate for students' improvement in knowledge and the ability within a content domain. Like it or not, large-scale testing is here to stay in our teaching and so it makes sense that we prepare our students to take tests in a principled way. Since the 1980s, educational researchers have proposed principles to examine test preparation appropriateness from two perspectives: theory-based and practice-based (Table 7.1). Teachers can use these principles to guide their test preparation practices.

Theory-based principles can be considered in two ways, by: (1) applying an evaluative continuum to test preparation activities ranging from ethical to unethical behaviour (Mehrens and Kaminski, 1989) and (2) creating sets of specific standards for evaluating the appropriateness of these activities, in terms of professional ethics (Popham, 1991). It can be seen from Table 7.1 that, since the 1980s, theory-based principles have been developed to include more concrete dimensions to evaluate test preparation activities. Crocker (2006) specifies the following criteria: *validity* requires that test preparation improves the validity of test score interpretation; *academic ethics* requires test preparation activities to be consistent with ethical standards of the educational profession; *fairness* means all test-takers should have equal access to preparation opportunities; *educational value* means test preparation should improve both test-takers' scores and their content knowledge; *transferability* requires test preparation to teach test-takers skills that can be used in different examination situations.

Table 7.1 Principles of test preparation practices

Category	Principles
Theory-based	❑ A continuum from ethical to unethical behaviours (Mehrens and Kaminski, 1989) ❑ Professional ethics (Popham, 1991) ❑ Validity ❑ Academic ethics ❑ Fairness ❑ Educational value ❑ Transferability (Crocker, 2006)
Practice-based	❑ Including curriculum objectives; integrating test content into curriculum ❑ Familiarizing students with various assessment approaches ❑ Instructing and reviewing test-taking strategies ❑ Motivating students to do their best in tests ❑ Managing time and frequency of test preparation throughout the year

One concern of these theory-based principles is practicality. Classroom teachers have complained that these standards are too general to follow in judging their own preparation activity (Popham, 1991). Therefore, practice-based principles of evaluating test preparation have been proposed to help teachers focus their test preparation on curriculum instruction or learning outcomes rather than test items. Turner (2009) identified five types of test preparation practices that can support learning and language development (Table 7.1): (1) teaching to the content domain covered by the curriculum, (2) using a variety of assessments, (3) reviewing/teaching test-taking strategies, (4) promoting students' motivation, and (5) managing time and frequency of test preparation. Turner also suggests practical teaching activities for each. For example, he suggested teachers could create opportunities for students to present their understanding of the content knowledge in different forms and

contexts such as independent work, oral presentations and written essays. When teaching students test-taking skills, teachers might review previous years' test papers to analyse task requirements, and help students to become familiar with the knowledge they are required to know and understand, including task formats. When planning a timeline for test preparation activities, teachers are advised to consider test requirements at regular intervals (e.g., bi-weekly or monthly) throughout the school year and to schedule test review activities in the weeks approaching the test.

Teachers can align their test preparation practices with both the theory-based and the practice-based principles in judging the appropriateness of their own test preparation activity. For example, teachers often instruct students to manage their time for test tasks as one common test-taking strategy to ensure test tasks are completed within a designated time limit. This test preparation activity can reduce the possibility of having insufficient time to complete test tasks owing to not managing time well. This particular activity can help to reduce the construct irrelevant variance (in this case, insufficient time) that is probably unrelated to students' academic competence, but can negatively influence students' test scores.

This test preparation activity conforms to Crocker's criteria of validity (2006), because it increases the plausibility of interpreting students' test scores and the validity of students' test scores. Teaching content domains (e.g., specific language skills such as reading, listening and so on) that are sampled by high-stakes tests (e.g., language proficiency tests) are aligned with the criterion of educational value or educational defensibility (Crocker, 2006; Popham, 1991), because it reduces the concern that test preparation practices might be limited to the contents (e.g., test items) that appear on tests. In addition, teachers can use empirical evidence as legitimate resources of performing appropriate test preparation activities.

It has been found that students preferred teachers' diagnosis of their weakness in specific language skills (e.g., pronunciation in speaking, vocabulary in reading and writing and so on) and needed opportunities for participation, questioning and practising language skills in test preparation classes as they do in their regular class (e.g., Alderson and Hamp-Lyons, 1996; Doe and Fox, 2011). Research suggests that some students taking test preparation courses believe teachers' instructions that aim to improve students' English competence could also contribute to better test performance (Doe and Fox, 2011; Ma and Cheng, 2016). These test preparation activities that emphasize general English competence improvement rather than focusing only on coaching test performance were shown to be associated with students' higher test scores in English language proficiency tests. To conclude, further empirical investigations will shed more light on pedagogical implications that enable teachers to perform appropriate test preparation activities.

ACTIVITY 7.4

Re-examine your own teaching and answer the following questions. Compare your answers, if possible, with those of a colleague's.

1. How would you describe the type of test presentation you provide, if any?
2. Have you ever taken or taught a test preparation course? If so, describe the experience? What were its benefits and drawbacks? How might a test preparation course have helped you or your students?
3. Have you ever taken a high-stakes test? If no, would you consider taking the test that your students will be taking?

7.4 Putting Students' Experiences at the Centre of Our Teaching: A Final Word

From the outset we have argued that assessment plays an essential role in language teaching and learning. The day-to-day assessment of student learning is unquestionably one of the teacher's most important, complex and demanding tasks. As teachers, we are the principal agents of assessment, so we need to ensure the quality of classroom assessment practices and need to use these practices in ways that best support our students' learning. If this statement still stands, we cannot ignore the role that students play in assessment. We have argued that assessment *for*, *as* and *of* learning should be integrated in our assessment practices and emphasized its role in supporting learning. Therefore, it is inevitable that consideration of our students' experiences with and responses to testing and assessment practice should be at the centre of our teaching and research.

Research on students as test-takers has investigated their experiences of being tested, their cognitive processes and the conditions in which they have been tested. Research in language assessment in particular has primarily addressed test-taking experiences from the perspectives of testing strategies (Cohen, 2006), test-takers' behaviours and perceptions during test-taking processes (DeLuca et al., 2013; Doe and Fox, 2011; Fox and Cheng, 2015), prior knowledge and preparation (Sasaki, 2000), test-taking anxiety and motivation for taking the test (Cheng et al., 2014).

Cheng and DeLuca (2011) explored test-takers' testing experiences and examined the relationship between aspects of testing experience and test-takers' perceptions of test validity and use. Fifty-nine test-takers at a large English-medium university in Asia participated in the study. Participants were from three parallel English language assessment courses focusing on the theoretical and psychometric properties of English language assessment, as well as on the practical application of principles of language assessment for teachers of English.

Data were collected via written statements of participants' test-taking experiences. Given the similarities in content among all three courses, results from all 59 participants were combined to establish a broader database for analyses and credibility of claims. Specifically, participants were asked to respond to the following prompt:

Write a report of a real-life language testing event in which you participated as a test-taker. Your report should be a reflection upon your positive, neutral, or negative experience with the language test and should address an issue of test validity and test use. Your report should be approximately 300–500 words in length.

The results reflected participants' multiple experiences with a range of large-scale English language tests. Cheng and DeLuca encouraged participants to discuss issues of validity and test use from a variety of testing experiences. They used this approach in order to explore the validity and use of language assessments in general, rather than to examine any one specific language test. As a further caveat, the results point to certain test features that test-takers identified. Eight overarching themes and 26 codes were drawn: (1) test administration and testing conditions, (2) timing, (3) test structure and content, (4) scoring effects, (5) preparation and test-taking strategies, (6) test purpose, (7) psychological factors, and (8) external factors and test consequences (see [Table 7.2](#)).

Table 7.2 Overarching themes and code frequencies

Themes and Codes	Frequency*
Theme 1: Test Administration and Testing Conditions	12
Code 1: Consistency in test administration	8
Code 2: Electronic/digital resources in test administration	7
Code 3: Stationery resources in test administration	4
Code 4: Testing environment	9

Theme 2: Timing	7
Code 5: <i>Time allocation for test components</i>	6
Code 6: <i>Overall time allocation for test</i>	3
Theme 3: Test Structure and Content	20
Code 7: <i>Authenticity of tasks</i>	13
Code 8: <i>Choice in constructed response items</i>	15
Code 9: <i>Psychometric properties and test item format</i>	15
Code 10: <i>Scoring criteria and item instructions</i>	4
Theme 4: Scoring Effects	11
Code 11: <i>Scoring</i>	9
Code 12: <i>Qualitative or holistic scoring approaches</i>	5
Code 13: <i>Examiner effects on scoring</i>	3
Code 14: <i>Marking criteria (awareness of, or lack of)</i>	6
Code 15: <i>Composite scoring practices</i>	2
Theme 5: Preparation and Test-taking Strategies	21
Code 16: <i>Effects of coaching and test-taking strategy preparation</i>	17
Code 17: <i>Consistencies in test preparation and test experience</i>	15
Code 18: <i>Effects of test on future test preparation</i>	2
Theme 6: Test Purpose	10
Code 19: <i>Consistency in explicit purpose and test items/format</i>	9
Code 20: <i>Perceived unfair test purposes</i>	1
Theme 7: Psychological Factors	20
Code 21: <i>Self-efficacy effects</i>	9
Code 22: <i>Negative emotions and anxiety</i>	18
Theme 8: External Factors and Test Consequences	8
Code 23: <i>Impact and perceived 'stake' of test results</i>	3
Code 24: <i>Exemption policies</i>	2
Code 25: <i>Perceived misuse of test results</i>	2
Code 26: <i>Social group privileging</i>	1

* Note: Theme frequencies do not total code frequencies due to double-coding of data

7.5 Looking Back at Chapter 7

Assessment exerts tremendous power on the lives of our students. What we do in assessment has consequences. Take a moment to look through the themes and codes in [Table 7.2](#). These themes and codes are related to both experiential (i.e., testing conditions and consequences) and psychometric (i.e., test construction, format and administration) aspects of testing. As teachers, we can greatly benefit from studies which investigate tests, testing practices and test-takers' responses to them. We can learn how to better support our students' learning by developing our own assessment literacy (Grabowski and Dakin, 2013). After all, supporting our students' success through assessment is our ultimate goal, and also the key message of this book.

Consider the themes and codes reported in [Table 7.2](#) in responding to the following questions. (For a more detailed report of this study, in test-takers' own words, see Cheng and DeLuca, 2011.)

1. Are you dealing with these aspects of large-scale testing in your classroom?
2. How are you supporting your students' learning in taking tests of this nature?
3. If you serve as an invigilator of such testing, what can you do to support students/test-takers to ensure that they demonstrate their ability?

Cheng and DeLuca (2011) advise us to *listen to the voices of test-takers*, and Fox and Cheng (2015) suggest we *walk a mile in test-takers' shoes* if we are to better understand what tests are measuring and their impact. We urge teachers to listen carefully to their own students, to elicit their students' reflections on and understandings of assessment practices in their classrooms. Coupled with an increased understanding of assessment potential and possibilities (which we hope is the outcome of reading and discussing the information in this

book), paying closer attention to our students' responses and reflections through assessment will enhance the quality of our teaching and increase their learning.

Suggested Readings

Cheng, L. & DeLuca, C. (2011). Voices from test-takers: Further evidence for test validation and test use. *Educational Assessment*, 16(2), 104–22.

Test-takers' interpretations of validity as related to test constructs and test use have been widely debated in large-scale language assessment. This study contributes further evidence to this debate by examining 59 test-takers' perspectives in writing large-scale English language tests. These findings offer test-takers' voices on fundamental aspects of language assessment, which bear implications for test developers, test administrators and test users.

Fox, J. & Cheng, L. (2015). Walk a mile in my shoes: Stakeholder accounts of testing experience with a computer-administered test. *TESL Canada Journal*, 32(9), 65–86.

Compares the responses of test takers who wrote both a high stakes computer-administered Internet-based test of English and a high-stakes paper-based test of English. The study investigates whether there are any differences in the proficiency construct being measured as a result of test administration format. This study provides evidence of the importance of test-taker feedback on testing experience in understanding what tests are measuring.

Sun, Y. & Cheng, L. (2014). Teachers' grading practices: Meanings and values assigned. *Assessment in Education*, 21(3), 326–343. doi: 10.1080/0969594X.2013.768207.

This study explores the meaning Chinese secondary school English language teachers associate with the grades they assign to their students, and the value judgments they make in grading. A questionnaire was issued to 350 junior and senior school English language teachers in China. Results of these analyses demonstrate

that the meaning of the construct of *grade* is closely related to two concepts: (1) judgment of students' work in terms of effort, fulfilment of requirement and quality; and (2) judgment of students' learning in terms of academic enablers (i.e., non-achievement factors such as habit, attitude and motivation that are deemed important for students' ultimate achievement), improvement, learning process and achievement.

Appendix: Samples of Some Commonly Used Classroom Assessment Tools and Test Formats

Below is a short list of some classroom assessment tools and test formats that are often used by teachers. The list is not exhaustive and provides examples only. There are many other alternatives. You may want to add others at the end of the list.

C-test	A type of cloze test, most frequently used to test reading, in which the second half of the words are removed at systematic intervals – often every second word in a reading passage. Example: <i>He under-_____ the prob-_____ but could-_____ solve it.</i> Answers: <i>He <u>understood</u> the <u>problem</u> but <u>couldn't</u> solve it.</i>
Checklist	A list of criteria to be considered (ticked or checked) in assessing a task, project, or performance. Checklists are used by teachers (in observing, monitoring and evaluating); they are also used by students when engaging in self-assessment. In recent years, the checklist criteria are often statements of what students know and can do – ‘can-do’ statements.
Cloze	A type of gap-filling test method where words or items are removed from an integrated text and students must supply or identify what's missing. Scoring may require an exact match or allow for any acceptable replacement. Typically there are no deletions in the first sentence or paragraph (of a long text). Deletions are made on the

basis of systematic intervals (as in the example below, where every sixth word is removed), or may test specific content (grammatical items, vocabulary).

Example:

On Tuesday, she had a doctor's appointment because she had had a mild fever for over a week. The doctor examined her and 1 _____ antibiotics. The doctor suggested that 2 _____ wait a few days to 3 _____ if the fever disappeared before 4 _____ the antibiotics. 'It is always 5 _____ to let the body heal 6 _____,' the doctor said.

Answers:

1. *prescribed*
2. *she*
3. *see*
4. *starting*
5. *better*
6. *itself*

Diary Writing about learning over time. Like a journal or learning log, diaries can be kept by both teachers and students to record students' learning. Different strategies can be used for sharing diary entries, but it is important to respect the privacy of the writer. Much can be learned about students' perceptions, understandings and development if diaries are shared.

Teachers/raters do not typically mark the quality of a diary, but rather respond on an ongoing basis with formative feedback on a student's insights and reflections. Marks are awarded for completion of the diary according to the guidelines set out in advance by the teacher.

Dictagloss A type of dictation activity where learners listen to a passage and take notes. Then, working with other learners, they attempt to reconstruct the original passage from their notes.

Dictation Although dictation techniques vary, typically a short passage is read aloud by the teacher and students attempt to faithfully reproduce it. The more accurate their reproduction, the higher their score.

Essay An extended piece of writing, often in response to a prompt or question.

Example (Prompt): *Do you agree with this statement? It is important to increase the amount of physical activity in schools in order to address the obesity epidemic.*

Essays are scored by teachers or other raters using criterion-reference scales or rubrics (either holistic or analytic).

**Gap-filling/
fill-in-the-
blank** Words or phrases are removed and students are required to replace them.

Example:

1. John ate his _____ at noon each day, and his _____ in the evening.
2. He always had bread and fruit in the morning for _____.

Answers:

1. lunch; dinner (or supper)
2. breakfast

Information gap A problem-solving task in which students must collaborate in order to find a solution.

Example:

One student is given a map with detailed information. His partner is given a map of the same location, but without details, and instructions to find the location of a restaurant. Without looking at each other's maps, the pair must exchange information, through question and answer, to locate the restaurant.

Example:

One student is given a picture of four automobiles. The other student is given a picture of five. Without looking at each other's pictures, the pair must exchange information, through question and answer, to identify which car is missing from the picture of four.

The exchange can be recorded (video or audio) and marked according to criteria for communicative interactions (i.e., comprehensibility, vocabulary accuracy, vocabulary range and so on).

Interviews Frequently used for assessing speaking, most interviews are semi-structured. The teacher/tester has a fixed set of questions or prompts that are asked of each student/test-taker but which allow test-takers to respond freely.

Example:

1. *What is your name?*
2. *What do you think is your strongest language skill?*
3. *What do you think is your weakest language skill?*
4. *Tell me something about yourself...*
5. *What do you hope to learn from the course this term?*

The student's/test-taker's responses can be recorded (video or audio) and marked according to criteria for communicative interactions (i.e., comprehensibility, vocabulary accuracy, vocabulary range and so on).

Learning log Ongoing responses to learning which are collected in a 'log' and encourage students to reflect on their learning, take more responsibility for it, and through increased self-awareness set realistic goals for their learning.

Teachers/raters do not typically mark the quality of a learning log, but rather respond on an ongoing basis with formative feedback on a student's reflections. Marks are awarded for completion of the log according to the guidelines set out in advance by the teacher.

Matching A testing technique that asks a student/test-taker to link one set of items with another. Often used in grammar and vocabulary tests.

Example:

Directions: Match the word on the left with its partner (synonym) on the right by drawing a line to connect the pair.

- | | |
|-------------------|------------------|
| 1. <i>Careful</i> | <i>Right</i> |
| 2. <i>Solid</i> | <i>Difficult</i> |

3. *Challenging*

Sturdy

4. *Correct*

Cautious

Answers:

1. Cautious

2. Sturdy

3. Difficult

4. Right

Multiple-choice

A test item which requires a test-taker to choose the correct answer from other choices (distractors). Each item tests a specific part of the construct and is comprised of a **stem** (a question, phrase, or sentence to be completed) and distractors.

Example:

1. *Which of the following would you expect to find at an aquarium?*

a) *lions*

b) *monkeys*

c) *dolphins*

d) *dinosaurs*

Answer:

c) *dolphins*

Observations

While students are engaged in an activity, teachers can record notes which document a student's development or achievement. Checklists (see above) can spell out specific criteria which a teacher wishes to monitor over the duration of a course.

Open-ended/
constructed
response
item

An item or test which requires students/test-takers to generate a response (rather than to identify a correct answer from a list of possibilities). There are many examples of open-ended items on this list, including interview questions, cloze items, gap-filling items or tasks, role plays and so on.

Example:

1. *When driving an automobile, there are many important things a driver must remember, including _____, _____ and _____.*
(3 points)

Answer:

Any reasonable answer is acceptable, for example: the speed limit, to signal when turning, to put on a seat belt, to avoid texting or answering a hand-held phone, and so on.

In an item such as that of the example, note the clues provided to the student regarding the amount of text (see the lines and commas) and the number of responses (there are three blank spaces and the item is awarded three points).

**Paired/
group oral
interaction** An interview or problem-solving activity which involves more than one student/test-taker interacting with the teacher/tester or task.

The student's/test-taker's responses can be recorded (video or audio) and marked according to criteria for communicative interactions (i.e., comprehensibility, vocabulary accuracy, vocabulary range and so on).

Portfolio An assessment approach which involves the collection of multiple samples of a student's work over time as evidence of development, achievement, or both.

Teachers/raters mark portfolios using the guidelines established for their development or, in some contexts, using a criterion-referenced scale or rubric.

Questionnaires While questionnaires can be used to elicit demographic information, they are also very useful in identifying students' interests, levels of motivation, study strategies and so on. The more we know about our students, the better able we are to support their learning.

Role play A task in which roles are assigned to one or more test-takers who enact the role. Often used to assess communicative competence and/or speaking.

Example:

1. Your friend has invited you to have dinner and meet her family. She is living at home with her mother, father and two younger sisters. You bring flowers and candy. Knock on the door, enter when it opens and greet your friend and her family.

The student's/test-taker's responses can be recorded (video or audio) and marked according to criteria for communicative interactions (i.e., cultural appropriacy, comprehensibility, vocabulary accuracy, vocabulary range and so on).

Self-assessment	Student-led assessment of their development. Self-assessment can take many forms and is encouraged through learning logs, diaries, 'can-do' checklists, questionnaires and so on.
Summary/paraphrase	Drawing on an original text (either spoken or written), the test-taker/student attempts to recreate the meaning of the text in their own words. Responses are marked by teachers/raters according to predetermined criteria, such as accuracy, expression, completeness and so on.
Tasks	A complex performance required of a test-taker/student as part of an assessment activity. Tasks require a test-taker/student to speak or write (although they may be prompted to do so in response to what they understand through listening and writing). For example, see the role play task, the dictagloss task, or the summary/paraphrase task in this list.
True/false	An item which has a correct and an incorrect answer. Such items are typically described as dichotomous (because there are only two options). This item type is not as useful as others (e.g., multiple-choice) because there is a 50% chance of getting the item right even if the student/test-taker doesn't have the capability, knowledge, or capacity that the item is testing. In other words, this item type encourages guessing.

Example:

Directions: Identify which statements are correct or not, by circling True or False.

- | | | |
|---|------|-------|
| 1. Some birds are not able to fly. | True | False |
| 2. Of, to and for are all prepositions. | True | False |
| 3. Blue, old and fast are all nouns. | True | False |

Answer:

1. *True*
2. *True*
3. *False*

Verbal protocols	This technique asks students/test-takers to comment aloud about an activity or performance in a task. 'Read aloud' or 'think aloud' protocols require students/test-takers to explain why they are making choices while or shortly after they have engaged with a task. Asking students to comment on their work, while they are working, alters their focus and the complexity of the task. This is a useful technique, however, in identifying why they use language in a certain way, understanding their weaknesses and strengths, and how better to support their learning. This technique has been used frequently for testing research.
Writing conference/ portfolio conference	A meeting between teacher and student(s) – or students and other students – in which work undertaken for a written assignment (i.e., writing conference) or assembled for one or more sections of a portfolio (i.e., portfolio conference) is the focus of discussion. Conferences, scheduled at regular intervals during a course, allow teachers and students to consider work undertaken, provide feedback on work-in-progress, and monitor and support development through collaboration.
Other test formats or assessment techniques	[Please add your own here.]

Glossary

Alignment The degree of agreement among curriculum, instruction, standards and assessments (tests). In order to achieve alignment, we need to select appropriate assessment methods, which reflect or represent clear and appropriate learning outcomes or goals.

Analytic scale A marking scale or rubric, which identifies specific features of language performance (usually with criterion descriptors). For example, in assessing a test of writing, an analytic scale might ask raters to award separate scores for such features as vocabulary use, paragraphing, sentence structure and so on. In assessing a test of speaking, raters might award separate scores for task completion, comprehensibility, pronunciation and so on. Analytic scales are of use in diagnostic assessment because they help to identify specific strengths and weaknesses.

Assessment Assessment is an umbrella term, which includes both large-scale testing, which is externally designed and administered to our students, and our daily classroom assessment practices. In this classroom context, this term refers to all those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged.

Assessment as learning This type of assessment activity occurs when students reflect on and monitor their progress to inform their future learning goals. It is regularly occurring, formal or informal (e.g., peer feedback buddies, formal self-assessment), and helps students to take responsibility for their own past and future learning.

Assessment for learning This type of assessment activity refers to the process of seeking and interpreting evidence for use by students and their teachers to decide *where* students are in their learning process, *where* they need to go and *how* best to get there.

Assessment of learning This type of assessment activity refers to assessments that happen *after* learning has occurred, to determine whether learning has happened. They are used to make statements about a student's learning status *at a particular point in time*.

Assessment plan An assessment plan is an overall guide for how we will assess students' achievement of the learning goals and outcomes relevant to instruction.

Canadian Language Benchmarks (CLB) A set of criterion-referenced descriptors of language proficiency, used by Canadian language teachers, learners and other stakeholders for teaching, learning and assessment in Language Instruction for Newcomers to Canada (LINC) classes. There are 12 benchmark levels.

Common European Framework of Reference (CEFR) A set of criterion-referenced descriptors of language proficiency, developed by the Council of Europe. These descriptors define six levels of proficiency (A1, A2, B1, B2, C1, C2) and are applied across countries that are members of the European Union. They are also widely referenced globally.

Consequences This term is associated with the results of the use or misuse of assessment results. Research into consequences of large-scale testing tends to focus on the after-effects of test interpretations and use on various stakeholders, including value implications and social consequences.

Construct The trait (traits) or underlying ability that we intend to measure through assessment. For example, *motivation* and *language proficiency* are constructs. Constructs are typically informed by theory or research. Tests provide operational definitions of constructs, eliciting evidence of knowledge or behaviour which reflects the presence (or absence) of the trait or ability.

Criterion-referenced assessment A type of measurement, which describes knowledge, skill, or performance through the use of descriptive criteria. Criteria are typically related to levels across a continuum of language development. These levels are often labelled as standards or benchmarks and distinguish one level of mastery from the next. For example, CEFR identifies different levels of language proficiency from A1 to C2.

Curriculum The term refers to the lessons and academic content taught in a school or in a specific course or programme. It is sometime called syllabus, course of study, programme of study, subjects and modules. A curriculum such as the ESLCO cited in this book provides a considerable amount of guidance as to what you can do as a teacher and what your students can do as learners at a particular level of ESL, but these guidelines do not specifically define your assessment activities by stating what your students should *do* to show what they have learned.

Diagnostic Assessment A diagnostic test or assessment procedure measures an individual's unique competencies, skills, or abilities which are necessary for performance in a specific context (e.g., *reading speed* or *knowledge of academic vocabulary* in the context of academic study). The information provided by the diagnosis results in a learning profile and is linked to specific learning activities that address the individual's weaknesses and promote his or her strengths.

Discrete-point items/tests Measures that isolate each item on a test. This is often referred to as item independence. Discrete-point items typically measure one feature of a construct at a time. For example, a test of grammar might have one question or item about the use of articles; the next question (item) might test adjectives and so on. Discrete-point tests typically use formats with right or wrong answers (e.g., multiple-choice, true/false).

Distractor In a multiple-choice test, the distractors are the choices offered to test-takers.

Distractor analysis In a multiple-choice test, we analyse each of the choices offered to test-takers to determine how effective the choices (distractors) are. If, for example, we offer one correct answer and three incorrect answers, we analyse who responded to the incorrect answers and in what numbers. If we find that one distractor attracted no responses from either the high or the low groups of test-takers, we have lowered the difficulty of the item (we might as well remove the distractor); if we find all of the high-performing test-takers choose this distractor (and get it wrong) and all of the low-performing students avoid it, we are probably not measuring the ability or trait we intended to measure. Distractor analysis is a means of helping us to improve the quality of each item. It is sometimes referred to as distractor efficiency analysis.

Ebel's guidelines Suggested guidelines for judging the quality of an item's discrimination (i.e., how well an item separates those students who perform well on the test from those who do not). The guidelines (ranging from 0 to 1) must be interpreted in relation to the type of test. In a norm-referenced context, 0.50 perfectly discriminates between high and low (50% get the item right; 50% do not). In a criterion-referenced context, no teacher would want 50% of her class to fail.

Fairness When students are provided with an equal opportunity to demonstrate achievement, and assessment yields scores that are comparably valid. This requires transparency, in that all students know the learning targets, criteria for success, and on what and how they will be assessed. Fairness also means that the students are given equal opportunity to learn. Fair assessment avoids student stereotyping and bias in assessment tasks and procedures. Appropriate accommodation is provided to students with special needs.

Feedback In language teaching, feedback from teachers to students is one of the most important ongoing sources of learning in the classroom. Feedback is the outcome of our assessment practices: *assessment as learning*, *assessment as learning* and *assessment of learning*. Feedback is the ongoing information provided to students to guide their learning. We call this type of information *formative*: it informs our students and supports their learning, but it also informs our teaching. The feedback we provide to our students also helps to shape our next steps in the classroom – the activities we choose. Feedback in language testing is provided by key stakeholders (i.e., test-takers and others) who respond to their experience of a test as part of test validation or evaluation.

Forced-choice test A forced-choice test is one that requires the test-taker to identify or recognize a previously presented stimulus by choosing between a finite number of alternatives, usually two.

Formative assessment Classroom assessment practices that inform teaching and learning.

High-stakes In language testing, a test which has major (often life-changing) consequences. For example, high-stakes proficiency tests, such as the Test of English as a Foreign Language Internet-based Test (TOEFL iBT) may determine whether or not a test-taker can enter university.

History file A record of test development that stores information on test decisions, changes and evolution over time. A history file is extremely valuable as part of the ongoing process of test development process.

Holistic scale: A marking scale or rubric, which focuses on the overall impression of a written or spoken performance. Levels are typically described with criterion descriptors, which summarize in general terms the quality of the performance.

Integrated task A task combines more than one skill (e.g., reading-to-writing; listening-to-speaking). Integrated testing incorporates two or more skills in a task or item, as opposed to discrete-point testing, which requires item/task independence (see 'Discrete-point items/tests' above).

Item A single unit on a test which elicits a test taker's response. Points are generally awarded by item and add up to the total score on a test.

Item difficulty The degree of demand or difficulty posed by an item on a test. The desired (and intended) level of difficulty will depend on the test's purpose and the type of test. Item difficulty is calculated on the basis of the overall test scores of the group. It is a useful measure of item quality.

Item discrimination A consideration of how well a test separates those who know or can do from those who do not (i.e., high performers from low). See 'Ebel's guidelines', above.

Language use survey An instrument used to collect information about a student's language use. It provides background information of relevance for the placement and the design of learning activities that will support learning.

Learning profile An instrument, which is used to report on individual test-taker's language skill, ability, strengths and weaknesses. It may combine information from multiple sources (e.g., Interest inventories, language use, proficiency test scores) and is used to inform teaching decisions in the classroom. In diagnostic assessment, the learning profile typically highlights strengths and weaknesses. Learning profiles evolve as learners develop. They provide a tool for collecting information about a student's learning over time.

Needs analysis In the classroom, a procedure for collecting information about students' language in order to define meaningful, useful and relevant activities. In language testing, needs

analyses inform test development decisions, particularly in the context of language for specific purposes (LSP) contexts, where the test is sampling language use within a specific domain (i.e., business, engineering, medicine).

Norm-referenced assessment In language testing and classroom assessment, measures, instruments, or procedures which have as their purpose the ranking and comparing of performance or knowledge in comparison to the performance of others in a given group.

Operationalize In language testing, to make what is unobservable or abstract (e.g., motivation, language ability, test anxiety) observable or concrete. For example, a language test is an operational definition of an abstract construct such as language proficiency. A test elicits behaviour, performance, or information from a test-taker which can be observed, scored and evaluated as evidence of the construct (underlying trait or ability).

Peer-assessment Evaluation or feedback provided by one student (or a group of students) for another.

Placement tests These are measures, which have as their purpose the sorting or grouping of students. For example, in language programmes, students may be sorted into levels in relation to their degree of language proficiency.

Proficiency tests Language tests designed to measure how much ability and/or capability a test-taker has in a given language.

Rasch analysis Informed by Item Response Theory (IRT), Rasch analysis assumes that the probability of getting an item correct depends on a combination of both the ability of the test taker and the difficulty of the item. It is widely used in large-scale testing, and is often used in studies of rater consistency.

Rating scale/rubric Guidelines for raters or teachers that define scores (e.g., grades, points) or describe levels, which are awarded for test-taker/student performances, behaviours, or work.

Reliability: The consistency, stability and dependability of the assessment results are related to reliability. This quality criteria guards against the various errors of our assessments. For example, reliability is the indicator of the degree of the potential errors we make in marking students' written work.

Self-assessment An individual's own reflection on and evaluation of their proficiency, capability, knowledge and so on. This type of assessment encourages students to become more aware of their

learning and more responsible for it. It provides students with experience which helps them to set more realistic goals for their learning and to monitor their progress in achieving these on an ongoing basis.

Sheltered course A course which provides instruction not only in a content or subject area, but also in language. For example, a high school or university course in history might be taken for credit towards a diploma or degree, but the teacher would teach not only history but also language (e.g., vocabulary, skills, strategies). Sheltered courses often run alongside and follow the same course outlines as mainstream courses, which do not offer language support.

Stem (in an item) That part of a multiple choice item which sets up the choices (i.e., distractors) for the test-taker. For example, in the following item, the stem occurs first:

1. Which one of the following is the best definition of *summative* assessment?
 - A. Feedback on an initial draft of an essay. [incorrect distractor]
 - B. Evaluation of a final product or outcome. [correct distractor]
 - C. Identification of strengths and weaknesses. [incorrect distractor]
 - D. Placement of a student in a group. [incorrect distractor]

Summative assessment A final evaluation at the end of a chapter, unit, course and so on. A summary of all that comes before within a designated time. An achievement test is a summative assessment instrument.

Target Language Use (TLU) Domain Language is embedded within and responsive to particular contexts. Test-takers who will occupy roles within these contexts (e.g., tour guides, medical practitioners, air traffic controllers) use language in particular ways. The TLU Domain is defined by certain language use tasks, which inform the design of test tasks, and ultimately allow us to generalize from performance on the language test to performance in the TLU domain.

Task On a language test, this is an item type which requires complex performance. Writing (e.g., essays, summaries) or speaking (interviews, role plays) tasks typically involve more than one skill and are scored by raters who judge their quality based on a

criterion-referenced scale. A pedagogical task in the language classroom is a component of an activity that maps onto learning outcomes for a course.

Test–Retest A method used to investigate the reliability of a test, which involves administering a test twice to the same group of test-takers within a short period of time (e.g., not more than two weeks). One efficient test–retest approach involves splitting a test into two more or less equal halves, based on a principled division of items and tasks, and to compute a correlation coefficient between scores on the two halves. This is known as split-half reliability (still a form of test–retest), but involves only one administration – avoiding a possible practice effect.

Test specifications The detailed blueprint or recipe for a test, which documents what a test is testing, how it is testing it and what we can infer from (i.e., the interpretation of) test scores or performance. It allows for the construction of other versions of the test and evolves in relation to evidence collected about the test over time.

Test-wiseness (TW) TW is defined as the ability to respond advantageously to items or test formats that contain clues and, therefore, to obtain credit without the skill, proficiency, ability, or knowledge of the subject matter being tested. Strategies include choosing the longest answer among multiple-choice distractors, when distractors are of unequal length; avoiding any distractors with the words ‘all’ or ‘every’; and ruling out as many alternatives as possible and then guessing from the ones that remain.

Validity The appropriateness of inferences, uses and consequences that result from the assessment. This means that a high-quality assessment process (i.e., the gathering, interpreting and using of the information elicited) is sound, trustworthy, or legitimate based on the assessment results.

Washback This refers to the influence of testing on teaching and learning – and is now commonly employed in applied linguistics. It is related to the term of consequences and impact.

References

- Alderson, C., Clapham, C. & Wall, D. (2001). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Alderson, J. C. (2007). The challenge of (diagnostic) testing: Do we know what we are measuring? In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner & C. Doe (eds), *Language testing reconsidered* (pp. 21–39). Ottawa: University of Ottawa Press.
- Alderson, J. C. & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280–97.
- Allwright, R. (1982). Perceiving and pursuing learners' needs. In M. Geddes & G. Sturtridge (eds), *Individualisation* (pp. 24–31). Oxford: Modern English Publications.
- Armstrong, C. (2006). *Understanding and improving the use of writing portfolio in the second language classroom*. Unpublished M.Ed. thesis. Queen's University, Kingston, Ontario, Canada.
- Artemeva, N. & Fox, J. (2010). Awareness vs. production: Probing students' antecedent genre knowledge. *Journal of Business and Technical Communication*, 24(4), 476–515.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bailey, K.B. & Curtis, A. (2015). *Learning About Language Assessment: Dilemmas, Decisions, and Directions*. 2nd edn. Boston, MA: National Geographic Cengage Learning.
- Biggs, J. & Tang, C. (2011). *Teaching for quality learning at university*, 4th edition. Maidenhead: McGraw Hill.
- Bishop, J. H. (1992). Why U.S. students need incentives to learn. *Educational Leadership*, 49(6), 15–18.

- Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–48.
- Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability*, 21(1), 5–31.
- Bond, T. & Fox, C. (2007). *Applying the Rasch Model: Fundamental measurement in the human sciences* (2nd edn). New York: Routledge.
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5–12.
- Brookhart, S. M. (2013). Grading. In J. H. McMillan (ed.), *Research on classroom assessment* (pp. 257–272). Los Angeles, CA: Sage.
- Brown, J. D. (1995). *The elements of language curriculum*. Boston: Heinle & Heinle.
- Brown, J. D. (1996). *Testing in language program*. Upper Saddle River, NJ: Prentice Hall.
- Canale, M. & Swain, M. (1980). Theoretical bases of communicative approach to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.
- Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. New York: Routledge.
- Carpenter, C. D. & Ray, M. S. (1995). Portfolio assessment: Opportunities and challenges. *Intervention in School and Clinic*, 31(1), 34–41.
- Cheng, L. (1999). Changing assessment: Washback on teacher perspectives and action. *Teaching and Teacher Education*, 15(3), 253–71.
- Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy and N. H. Hornberger (eds), *Encyclopedia of language and education: Language testing and assessment* (Vol. 7, 2nd edn, pp. 1–13). Chester: Springer Science Business Media.
- Cheng, L. (2013). *Language classroom assessment*. Alexandria, VA: Teachers of English to Speakers of Other Languages (TESOL).
- Cheng, L. (2014). Consequences, impact, and washback. In A. J. Kunnan (ed.), *The companion to language assessment* (pp. 1130–46). Chichester: John Wiley & Sons. doi:10.1002/9781118411360.wbcla071

- Cheng, L. & Curtis, A. (eds) (2010). *English language assessment and the Chinese learner*. New York: Routledge.
- Cheng, L. & DeLuca, C. (2011). Voices from test-takers: Further evidence for language assessment validation and use. *Educational Assessment*, 16(2), 104–22.
- Cheng, L. & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly*, 4(1), 85–107.
- Cheng, L., Klinger, D. & Zheng, Y. (2007). The challenges of the Ontario Secondary School Literacy Test for second language students. *Language Testing*, 24(2), 185–208.
- Cheng, L., Klinger, D., Fox, J., Doe, C., Jin, Y. & Wu, J. (2014). Motivation and test anxiety in test performance across three testing contexts: The CAEL, CET and GEPT. *TESOL Quarterly*, 48(2), 300–30. doi:10.1002/tesq.105
- Cheng, L., Rogers, T. & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods and procedures. *Language Testing*, 21(3), 360–89.
- Cheng, L., Rogers, T. & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education*, 33(1), 9–32.
- Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. Andrade & G. Cizek (eds), *Handbook of formative assessment* (pp. 3–17). New York: Taylor and Francis.
- Cohen, A. D. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly*, 3(4), 307–31.
- Cohen, A. D. & Upton, T. A. (2006). *Strategies in responding to new TOEFL reading tasks* (TOEFL Monograph No. MS-33). Princeton, NJ: Educational Testing Service.
- Colby-Kelly, C. & Turner, C.E. (2007). AFL research in the L2 classroom and evidence of usefulness: Taking formative assessment to the next level. *Canadian Modern Language Review*, 64(1), 9–38.
- Connelly, E. & Clandinin, J. (1988). Recovery of curricular meaning. In *Teachers as curriculum Planners* (pp. 81–97). Toronto: OISE Press.
- Cortazzi, M. & Jin, I. (1997). Cultures of learning: Language classrooms in China. In H. Coleman (ed.), *Society and the language classroom* (pp. 169–206). Cambridge: Cambridge University Press.
- Crocker, L. (2006). Preparing examinees for test taking: Guidelines for test developers and test users. In S. M. Downing & T. M.

- Haladyna (eds), *Handbook of Test Development* (pp. 115–28). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cumming, A. (2009). Language assessment in education: Tests, curricula, and teaching. *Annual Review of Applied Linguistics*, 29, 90–100.
- Davidson, F. & Lynch, B. K. (2002). *Testcraft: A teacher's guide to writing and using language test specifications*. New Haven, CT: Yale University Press.
- Davison, C. (2001). Current policies, programs and practice in school ESL. In B. Mohan, C. Leung & C. Davison (eds), *English as a second language in the mainstream: Teaching, learning and identity* (pp. 30–50). London: Longman.
- DeLuca, C., Chavez, T. & Cao, C. (2012). Establishing a foundation for valid teacher judgments: The role of pre-service assessment education. *Assessment in Education: Principles, Policy and Practice, Special Issue: Moderation Practice and Teacher Judgment*, 20(1), 107–26.
- DeLuca, C., Cheng, L., Fox, J., Doe, C. & Li, M. (2013). Putting testing researchers to the test: An exploratory study on the TOEFL iBT. *System*, 41(3), 663–76.
- Doe, C. & Fox, J. (2011). Exploring the testing process: Three test takers' observed and reported strategy use over time and testing contexts. *Canadian Modern Language Review*, 67(1), 29–53.
- Dörnyei, Z. (2001). New themes and approaches in second language motivation research. *Annual Review of Applied Linguistics*, 21, 43–59.
- Douglas, D. (2010). *Understanding Language Testing*. London: Hodder-Arnold.
- Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge, UK: Cambridge University Press.
- Ebel, R. L. (1954). Procedures for the analysis of classroom tests, *Educational and Psychological Measurement*, 14(2), 352–64.
- Elbow, P. (1986). *Embracing contraries*. Oxford: Oxford University Press.
- Elbow, P. (2003). *Embracing contraries: Explorations in learning and teaching*. Oxford: Oxford University Press.
- Elder, C. & von Randow, J. (2008). Exploring the utility of a web-based English language screening tool. *Language Assessment Quarterly*, 5(3), 173–94.
- Ferris, D. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Lawrence Erlbaum.

- Figlio, D. N. & Lucas, M. E. (2004). The gentleman's "A". *Education Next*, 4(2), 60–7.
- Fox, J. (2009). Moderating top-down policy impact and supporting EAP curricular renewal: Exploring the potential of diagnostic assessment. *Journal of English for Academic Purposes*, 8(1), 26–42.
- Fox, J. (2014). Portfolio based language assessment (PBLA) in Canadian immigrant language training: Have we got it wrong? *Contact, Special Research Symposium Issue*, 40(2), 68–83.
- Fox, J. & Cheng, L. (2007). Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first (L1) and second (L2) language test takers. *Assessment in Education*, 14(1), 9–26.
- Fox, J. & Cheng, L. (2015). Walk a mile in my shoes: Stakeholder Accounts of Testing Experience with a Computer-Administered Test. *TESL Canada Journal*, 32(9), 65–86.
- Fox, J., Haggerty, J. & Artemeva, N. (2016). Mitigating risk: The impact of a diagnostic assessment procedure on the first-year experience in engineering. In J. Read (ed.), *Post-admission language assessment of university students*. Cham: Springer International. DOI: 10.1007/978-3-319-39192-2
- Fox, J. & Hartwick, P. (2011). Taking a diagnostic turn: Reinventing the portfolio in EAP classrooms. In D. Tsagari and I. Csépes (eds), *Classroom-based language assessment* (pp. 47–62). Frankfurt: Peter Lang.
- Friedman, S. J. & Frisbie, D. A. (1995). The influence of report cards on the validity of grades reported to parents. *Educational and Psychological Measurement*, 55(1), 5–26.
- Fulcher, G. (2010). *Practical Language Testing*. London: Hodder Education.
- Gorsuch, G. (2000). EFL educational policies and educational cultures: Influences on teachers' approval of communicative activities. *TESOL Quarterly*, 34(4), 675–710.
- Gottlieb, M. (2006). *Assessing English language learners: Bridges from language proficiency to academic achievement*. Thousand Oaks, CA: Corwin Publishing.
- Grabowski, K. C. & Dakin, J. W. (2014). Test development literacy. In A. J. Kunnan (ed.), *The companion to language assessment* (pp. 751–68). Chichester: John Wiley & Sons.
- Graves, K. (2000). Assessing needs. In K. Graves, *Designing language courses*, pp. 97–122. Boston, MA: Heinle & Heinle.

- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 14(1), 75–97.
- Guskey, T. (2011). Five obstacles to grading reform. *Educational Leadership*, 69(3), 17–21.
- Haladyna, T. M. & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practices*, 23(1), 17–27.
- Hargreaves, A., Earl, L. & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal*, 39(1), 69–95.
- Harlen, W. & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education*, 10(2), 169–207.
- Hayes, B. & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS Academic Module. In L. Cheng, Y. Watanabe & A. Curtis (eds), *Washback in language testing: Research contexts and methods* (pp. 97–112). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Herman, J. L., Geahart, M. & Aschbacher, P. R. (1996). Portfolios for classroom assessment: Design and implementation issues. In R. Calfee & P. Perfumo (eds), *Writing portfolios in the classroom: Policy and practice, promise and peril*. (pp. 27–59). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Interference patterns: Applying linguistic theory to lesson production By: Douglas Magrath. *TESOL English Language Bulletin*, 12 August 2016. <http://exclusive.multibriefs.com/content/interference-patterns-applying-linguistic-theory-to-lesson-production/education>
- Ivanič, R. (2010). *Writing and identity: The discursal construction of identity in academic writing*. Amsterdam, John Benjamins.
- Kane, M. T. (2006). Validation. In R. L. Brennan (ed.), *Educational measurement* (4th edn, pp. 17–64). Westport, CT: American Council on Education.
- Knoch, U. & Elder, C. (2013). A framework for validating post-entry language assessments (PELAs). *Papers in Language Testing and Assessment*, 2(2), 48–66.
- Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor, MI: University of Michigan Press.
- Laufer, B. & Nation, P. (1999). A vocabulary size test of controlled productive ability. *Language Testing*, 16(1), 33–51.

- Linn, R. L. (2010). A new era of test-based educational accountability. *Measurement: Interdisciplinary Research and Perspective*, 8, 145–49.
- Linn, R. L. & Gronlund, N. E. (2000). *Measurement and evaluation in teaching* (8th edn). New York: Macmillan Publishing.
- Little, D. (2009). *The European Language Portfolio: where pedagogy and assessment meet*. Council of Europe. <http://www.coe.int/en/web/portfolio>.
- Liu, X. (2013). *Investigating factors influencing grading decisions among teachers of Chinese to speakers of other languages*. Unpublished M.Ed thesis. Queen's University, Kingston, Ontario, Canada.
- Livingston, S. A. (2006). Item analysis. In S. M. Downing & T. M. Haladyna (eds), *Handbook of test development* (pp. 421–44). New York: Routledge.
- Ma, J. & Cheng, L. (2016). Chinese students' perceptions of the value of test preparation courses for the TOEFL iBT: Merit, worth and significance. *TESL Canada Journal*, 33(1), 58–79. <http://www.teslcanadajournal.ca/index.php/tesl/article/view/1227>.
- Madaus, G. F. (1988). The distortion of teaching and testing: High-stakes testing and instruction. *Peabody Journal of Education*, 65(3), 29–46.
- McMillan, J. H. (2008). *Assessment essentials for standards-based education* (2nd edn). Thousand Oaks, CA: Sage.
- McMillan, J. H. (2014). *Classroom assessment: Principles and practice for effective standards-based instruction* (6th edn). Boston: Pearson. (See also editions 1–5.)
- McMillan, J. H. & Nash, S. (2000). *Teachers' classroom assessment and grading decision making*. Paper presented at the Annual Meeting of the National Council of Measurement in Education, New Orleans, LA.
- Mehrens, W. A. & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practices*, 8(1), 14–22.
- Messick, S. (1989). Validity. In R. L. Linn (ed.), *Educational measurement* (3rd edn, pp. 13–103). New York: Macmillan.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–56.
- Montgomery, P. & Lilly, J. (2012). Systematic reviews of the effects of preparatory courses on university entrance examinations in high school-age students. *International Journal of Social Welfare*, 21(1), 3–12.

- Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13–25.
- O'Connor, K. (2007). *A repair kit for grading: 15 fixes for broken grades*. Princeton, NJ: ETS.
- Ontario Ministry of Education. (2007). *The Ontario Curriculum Grades 9 to 12 English as a Second Language and English Literacy Development*. <https://www.edu.gov.on.ca/eng/curriculum/secondary/esl912currb.pdf>.
- Paulson, F. L., Paulson, P. R. & Meyer, C. A. (1991). What makes a portfolio? *Educational Leadership*, 48(5), 60–3.
- Popham, W. J. (1991). Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice*, 10(4), 12–15.
- Popham, W. J. (2001). Teaching to the test? *Educational Leadership*, 58(6), 16–20.
- Prabhu, N. S. (1990). There is no best method—Why? *TESOL Quarterly*, 24(2), 161–76.
- Pulgram, E. (ed.). (1954). *Applied linguistics in language teaching*. Washington, DC: Georgetown University Press.
- Randall, J. & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26(7), 1372–80.
- Read, J. (2008) Identifying academic needs through diagnostic assessment. *Journal of English for Academic Purposes*, 7(3), 180–90.
- Read, J. (2013). Issues in post-entry language assessment in English-medium universities. *Language Teaching*. doi:10.1017/S0261444813000190.
- Read, J. (ed.) (2016). *Post-admission language assessment of university students*. Cham: Springer International. doi: 10.1007/978-3-319-39192-2.
- Rolheiser, C., Bower, B. & Stevahn, L. (2000). *The portfolio organizer: Succeeding with portfolios in your classroom*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Sadler, D. R. (1989) Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–44.
- Sasaki, M. (2000). Effects of cultural schemata on students' test-taking processes for cloze tests: A multiple data source approach. *Language Testing*, 17(1), 85–114.

- Savignon, S. J. (2003). Teaching English as communication: a global perspective. *World Englishes* 22, 55–66.
- Savin-Badin, M. (2008). *Learning spaces: Creating opportunities for knowledge creation in academic life*. New York: Open University Press.
- Selivan, L. (2016). *Seventh International ETAI Conference Program Book*. Ashkelon, Israel, July 4-6, 2016.
- Simon, M., Chitpin S. & Yahya, R. (2010). Pre-service teachers' thinking about student assessment issue. *The International Journal of Education*, 2(2), 1–22.
- Sindelar, N. W. (2015). *Assessment powered teaching*. Newbury Park, CA: Corwin A SAGE.
- Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(2), 5–15. doi:10.1111/j.1745-3992.2001.tb00065.x
- Stiggins, R. J. (2008). *Student-involved assessment for learning* (5th edn). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Stufflebeam, D. F., McCormick, C., Brinkerhoff, R. & Nelson, C. (1985). *Conducting educational needs assessment*. Hingham, MA: Kluwer-Nijhoff Publishing.
- Sun, Y. & Cheng, L. (2014). Teachers' grading practices: Meanings and values assigned. *Assessment in Education*, 21(3), 326–43. doi:10.1080/0969594X.2013.768207
- Taylor, C. S. & Nolen, S. B. (2008). *Classroom assessment: Supporting Teaching and Learning in Real Classrooms* (2nd edn). New Jersey: Pearson Education.
- Thomas, S. & Oldfather, P. (1997). Intrinsic motivation, literacy, and assessment practices: "That is my grade. That's me". *Educational Psychologist*, 32(2), 107–123.
- Turner, S. L. (2009). Ethical and appropriate high-stakes test preparation in middle school: Five methods that matter. *Middle School Journal*, 41(1), 36–45.
- Waltman, K. K. & Frisbie, D. A. (1994). Parents' understanding of their children's report card grades. *Applied Measurement in Education*, 7(3), 223–40.
- Wang, H. & Cheng, L. (2009). Factors affecting teachers' curriculum implementation. *The Linguistics Journal*, 4(2), 135–66.
- Weir, C. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.

- White, R. (1988). *The ELT curriculum: Design, innovation and management*. Oxford: Basil Blackwell.
- Wiggins, G. & McTighe, J. (2005). *Understanding by Design*. Virginia: Association for Supervision and Curriculum Development.
- Wiliam, D. (2012). Feedback: Part of a system. *Educational Leadership*, 70(1), 30–4.
- Wiliam, D. (2015). *Formative assessment and reading instruction*. Presentation made for WSRA, Milwaukee, WI,
- Woods, D. (1996). *Teacher cognition in language teaching: Beliefs, decision-making and classroom practice*. Cambridge: Cambridge Applied Linguistics.
- Wormeli, R. (2006). Accountability: Teaching through assessment and feedback, not grading. *American Secondary Education*, 34(3), 14–27.
- Yesbeck, D. M. (2011). *Grading practices: Teachers' considerations of academic and non-academic factors*. Unpublished doctoral dissertation. Virginia Commonwealth University, Richmond, Virginia.
- Zamel, V. & Spack, R. (2004). *Crossing the curriculum: Multilingual learners in college classrooms*. Mahwah, NJ: Lawrence Erlbaum.
- Zoeckler, L. (2007). Moral aspects of grading: A study of high school English teachers' perceptions. *American Secondary Education*, 35(2), 83–102.

Index

- alignment
 - contexts 17, 41, 48–52, 59
 - definition xiv, 11, 34, 41, 223
 - of learning goals, assessment tasks and classroom activity 31, 34, 36, 41–43, 192
- alternative assessment (*see also* portfolio assessment)
 - x, 82
- assessment
 - activities (events, tools, processes, decisions) 2–3, 10–11, 17, 32, 41, 189–190
 - as learning (formative) x, xviii, 6, 64, 71–72, 181–182, 184, 189, 223
 - feedback xi, xv, 1, 3, 6–7, 10, 64, 166–175, 180
 - peer-assessment xviii, 12, 92, 182, 228
 - self-assessment 6, 30, 40, 61, 92, 143, 147, 173, 182, 228
 - definition and dimensions 1, 4, 7, 223
 - for learning (formative) x, xviii, 4–5, 71, 77, 176, 189, 223
 - definition 4–5
- methods (assessment tools) x, xv, 2–3, 7, 10, 62, 73–83, 108, 139–140, 144, 146, 163, 167–168, 175, 181, 190, 215–222
- observations, conversations, or products 46, 75–76, 161, 219
- open-ended or supply questions (e.g., short-answer, oral presentation) 74, 219
- selection questions (e.g., multiple-choice items, matching items) 74, 218–219
- student-centred assessments (e.g., portfolio, reading response journal) 74–75
- of learning (achievement, summative) x, xviii, 4–5, 8, 10, 62, 71, 145, 176, 189, 224
- definition 4–5, 224, 229
- plan (*see* planning assessment)
 - classroom assessment plans x, xv, 16, 66–73, 97, 167,
 - definition 224
 - examples 68–72
 - to motivate 10, 180–186
 - versus large-scale testing 62–66
- background on the field of language testing and assessment xvi–xviii
- backward design xiv, 41, 45, 52, 59

- benchmarks or standards 19
 - alignment with 49–51
 - Canadian Language Benchmarks (CLB) 19, 104, 141, 224
 - Common European Framework of Reference (CEFR) 19, 50–51, 84, 104, 141, 224
 - English as a Second Language (ESL) curriculum (ESLCO) 32–35, 41, 45–46, 51
- Canadian Association of Language Assessment/ Association canadienne pour l'évaluation des langues (CALA/ ACEL) 99
- European Association of Language Testing and Assessment (EALTA) 99
- International Language Testing Association (ILTA) 97
- consequences (*see also* impact; validity; washback) 12, 65, 99, 190–192, 195–196, 210, 212
 - definition 224
- construct
 - definition 4, 76, 104, 110–11, 224
 - operationalizing (operational definition of a construct) 103, 111, 114, 124, 143, 146, 224, 228
 - irrelevant variance 15
 - representation 14
 - specificity 107–108
- contexts 48–52, 63, 67, 179, 194–196, 201–201, 207
- Target Language Use (TLU)
 - Domain 105–106, 229
- course planning 44–60
 - template for course planning/ syllabus design 54–59
- criterion-referenced assessment 41, 104–105, 141
- can-do statements 143–144
- definition 224
- curriculum (curricular guidelines) ix, 8, 33–35
 - alignment through learning outcomes xiv, 11, 48–51, 223
- commonplaces 20–21
- curricular philosophy (*see* philosophy)
 - definition 225
- sheltered course 142, 229
- syllabus 54
 - template for course planning/ syllabus design 54–59
- diagnostic assessment xi, xv, 151–163
 - approaches 159–160
 - definition 8, 151, 225
 - examples 151–163
 - across a programme 160–162
 - diagnostic assessment tool 163
 - in a conversation
 - class 151–153
 - in English for Academic Purposes (EAP) 153–158
 - of writing 155, 163
 - online 154–155
 - post-admission, e.g., Diagnostic English Language Needs Assessment (DELNA); Diagnostic

- English Language Assessment (DELA) 163
- student (learning) profile of targeted needs 156–158, 163, 227
- system-wide 162–163
- dimensions of assessment 1, 4, 7
- discrete-point items/tests 134
 - definition 225
- distractor 117, 119–120, 129, 203, 219
 - definition 225
 - distractor analysis 135–137, 225
- Ebel's guidelines (*see also* item analysis) 134–135, 137
 - definition 226
- educational philosophy (*see* philosophy)
- English as a Second Language (ESL) curriculum (ESLCO) 32–33, 45–46, 48, 50–51
- ethical testing practices 97–99
- fairness xiv, 109
 - definition 11, 226
 - in rating and grading 121, 194
 - test-taking experience 12–14, 205
- feedback 1, 3, 166–176
 - definition 226
 - motivation xv, 166, 180–183, 186
 - shaping student learning 64, 92, 140, 150, 169–173
 - teacher's role as coach and judge 173–175, 190
 - test-takers' 129
- forced-choice test 184
 - definition 226
- formative assessment 226
- grading xvi, 6, 8, 191–192
 - of portfolios 91
 - research 194–196
 - scenarios 196–200
- high-quality assessment
 - xiv, 11, 34, 92, 102, 107–109
- high-stakes testing xvi, 16, 49, 51, 86, 200, 207 (*see also* large-scale testing)
 - definition 226
 - motivation and test anxiety 65, 179
 - Canadian Academic English Language (CAEL) Assessment 179
 - College English Test (CET) 179
 - General English Proficiency Test (GEPT) 179
- history file 111–112
 - definition 227
- impact (*see also* consequences and washback)
 - of benchmarks and standards 49
 - of large-scale tests 50, 65–66, 200, 203, 212
 - of portfolio assessment 91
 - of test methods 116–117
- integrated task 117
 - definition 227
- item
 - definition 116, 227
 - stem 229
 - versus task formats 116–117

- item analysis 129–137
 - discrete-point 116, 134, 225
 - distractor analysis 135, 225
 - Ebel's guideline 134–135, 137, 226
 - item difficulty 129–134, 227
 - item discrimination 129–134, 227
 - Rasch analysis 135, 137, 228
- large-scale testing (*see also* high-stakes testing) x, 2–4, 12, 62–65, 73, 137
- impact on test preparation 190, 200–201, 205
- International English Language Testing System (IELTS) 2
- Ontario Secondary School Literacy Test (OSSLT) 50
- testing experience 201, 209–213
- Test of English as a Foreign Language (TOEFL iBT) 2
- versus small-scale 62–66, 73, 180
- learning outcomes
 - xiv–xv, 36–44
 - alignment 11, 31, 34–35, 48–52
 - assessment tasks and learning outcomes 38, 41
 - sample task analysis 38–40
 - curricular 32, 51
 - defining learning outcomes 36–38
 - evaluating the quality of a learning outcome 42–44, 60
 - in course planning 44–47
 - in portfolio assessment 85
 - in test preparation 206
 - learning profile 158, 163, 227
 - mandate (for test development) 104, 109–110
 - method effect 117–118
 - motivation xv, 90, 175–184
 - assessment as learning 6, 180–184
 - examples of assessment that supports learning 161, 183–186
 - grading 191, 194–196
 - theories of motivation 177–179
 - needs analysis or assessment 18, 139, 146–148, 150
 - alignment through learning outcomes 52
 - definition 227–228
 - Five-Minute Essay 150
 - philosophies 149
 - purposes for 147–149
 - student and teacher perceptions 18–19
 - norm-referenced assessment 104–105, 133–134, 141–142, 171
 - definition 228
 - peer-assessment 12, 79–80, 92, 182, 184
 - definition 228
 - philosophy (of teaching, learning and assessment) 15–27, 34–35
 - educational (curricular) philosophies 17–22

- classical humanism 18
- post-modernism or
 - eclecticism 19–20
- progressivism 18–19
- questionnaire 23–25
- reconstructionism 19
- teachers' personal assessment
 - profile 26–27
- teaching philosophy and
 - grading 194
- placement testing 141–145
 - as achievement 145
 - assessment conference 144
 - decision tree 142
 - definition 228
 - language use survey 142, 227
 - self-assessment 143
- planning assessment
 - backward design 41
 - classroom assessment
 - plans 66–73
 - definition 224
 - examples 68–73
 - horizontal and vertical
 - perspectives 44–47
 - in course or syllabus
 - design 46, 54–59
- policy (*see also* benchmarks or standards) 49
 - alignment 49
 - No Child Left Behind* 49
- portfolio assessment x,
 - xv, 82–96, 137–138
- benefits and challenges 90
- conferences 89
- definition 83
- e-portfolios 83, 90
- guidelines 85–89
- planning 96
- purpose 83
- requirements 89
- showcase versus working
 - 83–84, 89
- practicality xiv, 12, 181, 206
- proficiency tests (*see also* high-stakes testing; large-scale testing) 15–16, 51,
 - 66–67, 106, 142, 161, 179, 207–208
- definition 228
- purposes of assessment 7–10, 83
- rating scales/rubrics 123–128,
 - 228
 - analytic scales 125–127, 223
 - holistic scales 124–125, 227
 - partially correct answers 122
- reliability
 - definition 11, 228
 - in grading 196
 - in scoring 122
 - inter-rater reliability 125
 - test-retest 159–160, 230
- self-assessment
 - can-do statements 143–144
 - definition 228–229
 - in placement 143
 - of writing 91
 - Thinking About Doing Better* 184–185
- summative (*see also* assessment, of learning)
 - definition 224, 229
- Target Language Use (TLU)
 - Domain 105–106, 229
- task 116–117
 - alignment with learning
 - outcomes xiv, 38, 44–45
 - analysis 38–40, 129

- task (*cont.*)
 - assessment tasks 11, 37–41, 44–49, 52, 55, 58–59, 62, 72, 176, 182
 - definition 221, 229–230
 - formats 116–118, 207, 221, 227
 - in test development 105–116, 114–116
 - task and item
 - writing 116–117
- teaching philosophy (*see* philosophy)
- test development 102–122
 - construct definition 104–111
 - history file 112
 - item analysis 128–132
 - item versus task 116–117
 - overview of a test development process 108–120
 - test specifications 107–108, 111–118, 230
 - text identification or development 115, 118–122
 - Table of Specifications 113–116, 122
- test preparation practices 190, 201–202, 206–208
 - alignment with theory and practice 205–207
 - definition (types) 201–202
 - pedagogical
 - implications 205–208
 - research 202–204, 208–211
 - test-wiseness 203, 230
 - types 206
- test taking experience 12–15, 209–211
 - feedback on a new test 129
 - large-scale tests 210
- test specifications (*see also* test development)
 - definition 230
- textbooks 30, 48
 - alignment through learning outcomes 52
 - in assessment plans 69–72, 96
 - in test development 112, 115, 120
- validity (*see also* consequences, impact, washback) xiv–xv, 11, 203–205
 - consequential evidence 64–65, 109
 - Crocker's criteria of validity 205, 207
 - definition 11, 230
 - in classroom assessment 64, 192, 194–196
 - in test development 109–112, 125, 134
 - in test preparation 125, 203–206
 - validation xvii, 14–15, 109, 210, 213
- washback (*see* consequences; impact; validity)
 - definition 12, 65, 230
 - positive and negative 65–66, 203
 - potential 66, 72