Tests that Second Language Teachers Make and Use

Tests that Second Language Teachers Make and Use

Edited by

Greta Gorsuch

Cambridge
Scholars
Publishing

Tests that Second Language Teachers Make and Use

Edited by Greta Gorsuch

This book first published 2019

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-3901-6
ISBN (13): 978-1-5275-3901-3

TABLE OF CONTENTS

Russian

Spanish

CHAPTER ONE

INTRODUCTION: TESTS AND TEACHERS, TESTS AND TEACHING

GRETA GORSUCH

Tests and Teachers

Like it or not, we are "in a relationship" with tests. As learners we have early and continuing experiences with tests. These could have been large-scale standardized tests that all of our school friends took, in all grade levels, in all classrooms, on the same day. These could have been the weekly quizzes we had in high school French, where we had to conjugate verbs. In our lives today, these could be on-site, online tests that we take at the driver's license bureau. Or, our employer requires us to take courses on current laws that are relevant to our roles as employees. We then take a short quiz to certify that we have a basic, as-needed, applied knowledge of those laws.

As second language teachers, our school may require us to prepare learners for large-scale tests so that they may enter university, or graduate. And we have to write and give tests in order to award English-, or French-, or Japanese-language course grades. We may become certified interviewers or graders for a large-scale second language test offered by a testing company.

In some respects, teachers have an uneasy relationship with the simple existence of tests. There are multiple reasons. First, there are those Language Tests, made by commercial testing companies. They command a lot of attention. They are expensive, they may be unfair, teachers may be judged by their learners' results, and preparing for them may consume precious class time. Second language classes are often under-scheduled in relation to what our schools want learners to achieve. So, this last point is particularly painful. Some may feel that such tests capture learner knowledge or skills that the teachers, and even the learners, feel is irrelevant to learners' ability to use a second language to do what they want.

Second, teachers are required to make tests for the purpose of giving grades. Making tests is effortful, time-consuming, and requires technical skills teachers may not feel they have. Making tests and quizzes means that teachers must put into concrete form what they intuitively believe their course is about, and what their learners are about. Such knowledge, called teacher theory by some, is very effective at quickly solving problems, and turning "round" or shapeless content into a linear, moment-by-moment classroom experience. Pinning down our quicksilver teacher theory and turning it into a test, however, is a different matter, and it is challenging to do. And what about the technical skills needed to write and score tests? Are such skills unreachable, and unreasonable to expect working teachers to have? The commercial, large-scale tests may be held up, by ourselves, and by others, as exemplars of "good testing." Yet they are made and scored by professionals, who have specialized training and whose only job is to do testing. This does not often match the reality of teachers' lives.

Third, teachers must balance what they want to test, with what they can test. They may want to test learners' second language conversation ability. They know that their school's required end-of-year test does not capture conversation ability. And yet they also know learners spend a lot of time on conversation practice in classes. But with thirty students, there is no time to give and score such a test. The logistics are impossible. Fourth and finally, giving tests is socially fraught. Teachers must award course grades, presumably based on quizzes, tests, and other sources of information. The resulting scores and grades are bound to create conflicts with learners, and their parents in some cases. Arguing over test scores and course grades may become a major topic of office hour visits. To the extent that our decisions based on tests are called into question, we may have to defend our tests. Even the prospect is deeply unpleasant.

Tests and Teaching

In other respects, second language teachers have a potentially positive relationship with tests. This could be called a "teaching relationship" with tests. There are second language teachers who enjoy making tests, and enjoy what the tests can tell them about their learners. There are many such examples offered by the contributors to this book, found in Chapter Four, "The Tests." Looking at tests through this frame, tests are more supportive of teaching than they may seem at first. Teachers make tests to figure out what learners do and do not know, and what they can and cannot do. Since test scores or teachers' impressions of learners' performances on the test may provide information on this, teachers then have the option of changing

their instruction. They might repeat content, such as that unit on paragraph writing, or skip future content, such as that activity on indefinite articles, since the students already seem to know it. Or, teachers may combine content from two or more lessons into a more elaborate project, such as learners planning an all-day tour of their city for foreign visitors. The teacher had learned from quizzes on multiple lessons on constructing descriptions, reporting past events, linking past events together, and solving communication breakdowns within these topics, that learners were ready to attempt something harder.

Other teachers may enjoy giving frequent, ungraded quizzes as a way to focus learners' attention, and to prepare them for graded tests. Quizzes can be interactive and game-like, raising learners' energy levels. There is support for these ideas. For instance, test effect is the idea that taking tests and quizzes helps with memory retrieval, reducing anxiety, and spacing out study, therefore promoting learning processes. Another idea is pedagogically worthwhile tests. This is where teachers make tests that they would also feel comfortable using as classroom materials. These ideas will be described in more detail in Chapter Eight, "Practical Methods for Using Tests for Teaching and Learning."

So, second language teachers are "in a relationship" with tests, and this is unlikely to change. The relationship may take many forms, both negative and positive, as argued above. This book, *Tests that Second Language Teachers Make and Use*, is intended to illustrate and explore the positive aspects of the relationship between tests and teachers, and tests and teaching.

What is this Book?

Welcome to *Tests that Second Language Teachers Make and Use*. This book is a collection of fifteen actual second language tests that working teachers in six countries have made and used. Many of the tests, or parts of them, are still in use. They are quizzes, series of quizzes, unit tests, mid-term exams, and final exams. In other words they are the common classroom tests used by teachers to assess learners' achievement, to provide feedback to learners, and to award course grades. As such, they are criterion-referenced tests and performance tests, that have, in the eyes of the teachers using them, some relationship to the curriculum upon which the course is based. The tests in this book are not used for program-level decisions such as learner placement or admission to a program. They are not used to compare learners in different classes or programs to each other. These functions would be the work of norm-referenced tests, or general language ability or proficiency tests, such as the TOEFL (Test of English as a Foreign

Language, Educational Testing Service, 2019a), IELTS (International English Language Testing System, IELTs, 2018), or TOCFL (Test of Chinese as a Foreign Language, 2011).

The title of this book uses the term "second language" instead of "second and foreign language." Traditionally, "second language" suggests that a language is being learned in a country where the language is widely used, such as Turkish students learning German in Germany. "Foreign language" suggests that a language is being learned in a country where it does not have widespread or official uses, such as learners of English in Korea. There can be little doubt that teachers and learners in either second or foreign language learning contexts have different experiences and operate under different constraints. However, learning takes place in learners' heads no matter what context they are in. What goes on is grappling with a language other than the mother tongue--a second language. Therefore, the term second language in this book will refer to a second or foreign language, learned in diverse contexts.

The tests in this book are for Chinese (Mandarin), English, French, German, Italian, Japanese, Russian, and Spanish. For non-English language tests, the contributors have provided a basic English translation. The tests reflect the variety of second language courses that teachers teach, and the variety of programs and schools in which they teach. Thus, the tests reflect diverse conceptions of knowledge and skills, content, and learner ability levels.

The tests are coupled with commentary written by the contributors and edited by the book editor. The commentary was built on a principled framework that gets at processes teachers use to write and use tests. See Chapter Two for details on the framework ("A Framework to Probe Tests that Second Language Teachers Make, Score, and Use").

There are five testing content chapters in the book that define key terms and concepts. The chapters are described in more detail below, but they include: "Criterion-referenced Tests and Performance Tests" (Chapter Three); "Communicative Competence and Language Use Description Frameworks and Second Language Tests" (Chapter Five); "Practical Methods for Validating and Improving Tests" (Chapter Six); "Practical Methods for Setting Cut Scores and Making Decisions" (Chapter Seven); and Practical Methods for Using Tests for Teaching and Learning" (Chapter Eight). The chapters present selected, currently and commonly accepted concepts in testing, such as you would find in many available books on language testing. But *Tests that Second Language Teachers Make and Use* goes one step further. The testing concepts in the chapters are illustrated by the contributors' tests and their test commentaries. The commentary probes

their processes of test construction, use, and validation. Thus, readers will see testing concepts and practices in action, in the busy, messy, time-sensitive lives of working teachers. One thing the content chapters do not do, is to support some sort of "test police" analysis. The tests here are not held up against some standard implied by the testing content chapters and the framework used to form the test commentaries. The content chapters are offered as help, and illustrations of what working teachers do. The commentary framework aims at consistency and descriptiveness.

This last idea speaks to one of the reasons this book was done. Tests are an everyday feature in the lives of high school and college-level teachers and learners. Nonetheless, teacher-made tests are under-studied. Perhaps the tests are considered "routine" and idiosyncratic. Or, the tests are seen as poor reflections of what testing professionals can do.

Why this Book?

Classroom tests are an everyday feature of second and foreign language high school and college-level classrooms across the globe. Otherwise known as criterion-referenced tests or performance tests, such tests are the familiar quizzes and exams that teachers use. Yet little is known about how teachers make these tests. What are the processes by which they write tests? What knowledge sources do they draw from? If teachers inherit a test from a supervisor or a previous teacher, what do they make of it? Do they use previous tests or parts of them? What is their process of adaptation, and what knowledge sources do they draw from for that?

Still less known are teachers' preferences about test item formats to use, and areas of learners' communicative competence they capture in their test items. Further, little is known about whether, or how, teachers check the reliability (consistency) of their tests, and whether and how teachers validate their tests. These last two speak to the fairness of tests, and the usefulness of the resulting test scores. Finally, instructional staff in high school and college-level foreign or second language departments are diverse in their language backgrounds and levels of professional development. Some instructors have years of experience, whereas others are being supported as novice instructors and graduate students. What can their test making and test use practices tell us about their different understandings of language learning, and program priorities and traditions?

Rather than view teachers' tests as idiosyncratic and poor shadows of what professional test writers can accomplish, it is more constructive to identify the deeper patterns and order of tradition, innovation, and reasoning

behind teachers' tests and testing practices. Thus, this book is intended as a resource for readers.

Who is this book for?

This book is intended for multiple audiences, including working teachers, teacher educators, and department-level administrators who want to increase their ability to write good, useful, and defensible classroom tests. Teacher educators and department-level administrators also want ways to use tests as a means of professional development (for themselves and for those working under them), and as measures of course and program success. This last speaks to evaluation, which is specialized, applied research of growing importance to the field of second language education. *Tests that Second Language Teachers Make and Use* may help any of these audiences, in multiple ways.

There certainly are scholars interested in second language teachers and testing (called assessment, by some, to describe less formal "language sampling" activities done by teachers, see Rea-Dickins, 2004). They may find this book useful. It should be stated, however, that the tests and commentaries in this book are not intended to be representative of any sort of pattern or consensus of testing attitudes or practices among teachers. Groups of teachers are not compared in terms of their length of experience, geographic location, learner population, language taught, or previous training, even though they provide information on these points in the test commentaries. There was no intention of establishing a predefined sample of responses, or attempting generalizability with that sample. In other words, this is not survey research. Rather, the editor sent out a call for contributions to every conceivable association, listserv, or website she could find, and to every teaching or teacher preparation colleague she could think of. The collection of tests and commentaries here represent what was sent in response to the call, and what was successfully developed in a collaborative fashion between the contributors and the book editor.

What is in the Book

There are five parts to the book. There are five testing content chapters; a chapter outlining the basis for the test commentaries and the questionnaire used to form them; the tests and commentaries themselves; a reference section; and an index.

The testing content chapters. The testing content chapters have a parallel structure. There is a brief introduction which states the reason for

the chapter; definitions of key concepts, which includes examples of applications of the concepts emerging from the tests and commentaries themselves; a chapter summary; and a brief list for further reading.

Chapter Three is on criterion-referenced tests (paper and ink tests) and performance tests (tests where learners write or speak, and their performance judged using scoring criteria). These tests are used to diagnose learners' strengths and weaknesses and give feedback in relation to a course, and to estimate learners' achievement in a course. This chapter comes early, in that criterion-referenced tests and performance tests are the types of tests that appear in this book. This chapter defines concepts needed to understand the function and scope of the teachers'/testers' tests that appear in Chapter Four. It would be less useful to read the tests and commentaries thinking about how the tests would be used for placement in a program, or program admission. Norm-referenced tests are not treated in this book. Because the concepts of criterion-referenced tests and performance tests are foundational to understanding the educational roles and scope of the tests appearing in this book, this is the only testing content chapter that precedes the commentary framework chapter (Chapter Two) and the tests and commentaries (Chapter Four). The remaining test content chapters follow Chapter Four. In this way, readers can digest the tests and commentaries in the order of their own interests, and then consult the remaining testing content chapters as a reference for further understanding.

Chapter Five is on relating test writing to the high- and middle-level theories currently in use in second language education. The high-level theory here is, of course, communicative competence, the idea that language is not only various types of knowledge, but the ability to use a second language. The middle-level theories refer to language-use frameworks such as the ACTFL *Guidelines* (2012a) and *CEFR* (Council of Europe, 2001, 2018). Despite its theoretical impact and professional salience, many working teachers grapple with how communicative competence gets translated into actual test items and tasks. How does textual competence, for example, become test items or a test task? The same may be said of the ACTFL *Guidelines* and *CEFR*. How does a description of what learners are able to do at a particular ability level become test items, a test task, or scoring criteria or rubrics? Perhaps a missing link is understanding better how teachers relate these high-level and middle-level theories and frameworks to their course outcomes. By extension, then, what classroom activities and tasks do teachers then extrapolate from outcomes, and/or the materials available to them? And to what extent then do the activities, tasks, and materials inform the test items or test tasks teachers write? There are probably a host of other missing links that cannot be captured even by

complex and nuanced models such as the Teacher Theory Model (Figure 2-1). Some of the test commentaries from contributors to Chapter Four bring further clarity to these.

Chapter Six is on practical methods for validating and improving tests. Validation refers to how teachers/testers know whether a test measures what they think it measures, and whether using scores from a test is appropriate. Validation sometimes brings on images of professional testers hunched in front of their computers, doing statistical analyses beyond what ordinary teachers can do or have time to do. But many validation strategies can be surprisingly practical, and easy to build into tests during the test writing, administration, scoring, and decision-making stages. Validation strategies will be described and illustrated in tests and test commentaries.

Chapter Seven is on practical methods of setting cut scores. Many teachers score their own tests, but in addition they must make decisions based on test scores. In other words, does a learner pass or fail a test? What grade do they get on a test? Practical methods of cut score methods are given adapted to both pass/fail decisions and grade decisions. Examples from tests and commentaries are highlighted.

Finally, Chapter Eight is on practical methods of using tests for teaching and learning. This chapter is essential to taking a more proactive relationship between tests and teaching. In other words, what are ways tests can be used for teaching? Not only to plan instruction, but to use tests as part of teaching? There are exciting developments in general education, and second language education, that contribute to our practical, working understanding of how to more closely marry tests to our courses. Examples of these kinds of thinking will be illustrated using tests and test commentaries from the book.

Probing teacher-made tests, and test and test commentary chapters. These two chapters appear sequentially. Chapter Two, "A Framework to Probe Tests that Second Language Teachers Make, Score, and Use," introduces two models. The first is the Teacher Theory Model (Figure 2-1), first formulated by Griffee (2012a), which proposes explanations of how teachers develop action-oriented theories in schools to solve the "problem" of classroom tests. The second model is a Life of a Classroom Test model (Figure 2-2) which proposes an explanation, ordered sequentially in terms of the life of a test, of how teachers can plan, score, and use test scores to teach. Both models were needed to design the questionnaire that probes the tests the teachers/testers offer in Chapter Four. The models' components, and relationships between them, are defined. Then, the five-part questionnaire is offered.

The second chapter in this sequence, Chapter Four, "The Tests" chapter, is the focal point of the book. Each test appears with a brief introduction

naming the teacher/tester, a description of the context in which the test is used, the tests themselves and test answer keys and scoring rubrics, and the teacher/tester's responses to the five-part questionnaire, rewritten as prose under the headings suggested by the questionnaire parts.

It is expected that if any part of the tests published here are adapted or used by readers, the readers will then give credit by listing the proper citation of the work on the test or parts of the test they adapt. Here is how to cite a test from this book using APA (American Psychological Association) Guidelines:

Last name, First name (XXXX). Title of test. In G. Gorsuch (Ed.), *Tests that second language teachers make and use* (pp. xx - xx). Newcastle-upon-Tyne, United Kingdom: Cambridge Scholars Publishing.

References. The reference list offers full citations for previous research or commentary cited in-text. The references are both practical and theoretical, and every effort has been made to list references that are readily available either online, or through a standard library document delivery service.

Index. The index is intended for readers to identify page numbers in the book on key topics, concepts, and definitions they are interested in.

CHAPTER TWO

A FRAMEWORK TO PROBE TESTS THAT SECOND LANGUAGE TEACHERS MAKE, SCORE, AND USE

GRETA GORSUCH

Why a Framework to Probe Classroom Tests?

This chapter introduces two models. One proposes an explanation of how teachers develop action-oriented theories in schools to solve the "problem" of classroom tests. The second model proposes an explanation, ordered sequentially in terms of the life of a test, of how teachers can plan, score, and (perhaps) use test scores to teach. Both models were needed to design the questionnaire that probes the tests the teachers/testers offer in Chapter Four. The questionnaire appears at the end of this chapter. Beyond this book, models are also needed to organize, understand, and further discuss complex phenomena or concepts in second language education. And teaching, and teaching and testing, are complex.

Explaining how Teachers Develop Theory

The first model aims at illustrating how teachers develop theory. The model is about teachers, and its basis is how teachers learn, and how they create their own personal theories in order to plan and do coherent actions such as teaching, and making tests. See Figure 2-1. Definitions of the components follows.

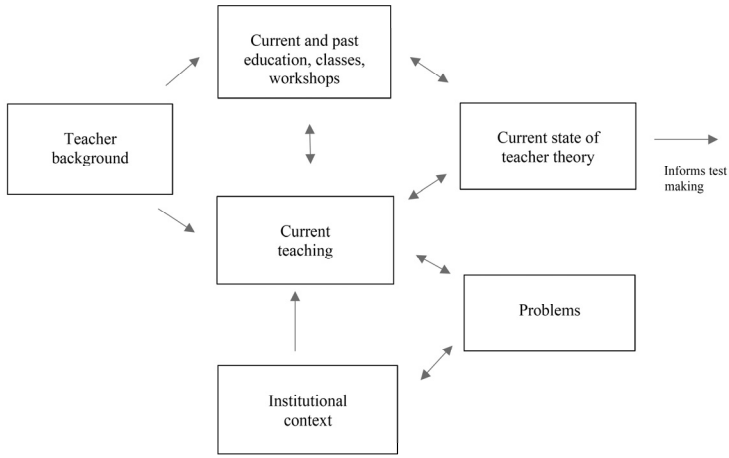

Figure 2-1. **Teacher theory model (based on Griffee, 2012a).**

The teacher theory model originated in research done by Griffee (2012a), who was working with novice second language teachers who were also M.A. in applied linguistics students. His purpose was to learn how novice teachers "construct a teacher-identity that serves as a context for [their] decisions and actions" (p. 202). In related research, Griffee refers to teacher-identity as teacher theory, or low theory, which is "implicit theory held in the minds of individual practitioners" (2012b, p. 52). There will be more detailed discussion of teacher theory, or low-level private theory, and other types of middle- and high-level theories in second language education later in the chapter in the definition section on the model component "Current and past education," and also in Chapter Five on how theories and frameworks of language use descriptions may act as resources for making tests.

Teacher Theory as a Basis for Teacher-made Tests

In Figure 2-1, the component, "Current state of teacher theory" represents teacher theory. Gorsuch and Griffee (2018) refer to low-level theory or teacher theory as "local theory" in that teacher theory is strongly formed by specific contexts with specific learners in courses with a particular curriculum. As can be seen in the model, teacher theory emerges from "Teacher background," "Current and past education," the demands of

"Current teaching" and the "Institutional context," and what Griffee identified as "Problems" (2012a). Thus, teacher theory changes over time, depending on current teaching, among other things. Teacher theory is not a new concept in our field. Widdowson (1993, p. 264) refers to a teacher belief system, or "personal constructs of teachers," representing teacher attitudes, thinking, and decision making.

It is argued here that "Current state of teacher theory" forms the basis of teachers making and scoring tests, and using test scores. How is this the case? Traditionally, teacher theory is thought to be about teaching. It is oriented toward solving problems, such as how to understand what learners need and how to deal with it, and what materials to choose and how to teach with the materials (Gorsuch & Griffee, 2018). Teacher theory seems oriented to translating what teachers think a course is about (the course outcomes and course content), into the beginnings, middles, and ends of successive lessons over time. Teachers create activities and tasks that accomplish the lessons. Griffee (2012a) never specified that his teacher theory model represented teacher theory about tests. And Rea-Dickins (2004), who is no stranger to working language teachers and to teacher theory, calls for a model of teachers and tests. This implies that somehow, teaching must be understood differently than testing (or what Rea-Dickins calls "assessment"). Certainly, Nikolov (2016) found that teachers in Hungary, tasked with writing diagnostic tests for young second language learners, produced tests that "tended to focus on errors, accuracy, and what students cannot do" (p. 75). She felt this was different from the *CEFR* frameworks (Common European Framework of Reference; Council of Europe, 2001) that had been adopted at the teachers' schools, which was thought to guide teaching. Thus, in her mind, teachers' teaching and testing were different.

Yet clearly teachers have theories about tests. Barrette (2004); Cheng, Rogers, and Huiqin (2004); Davison (2004); Kikuchi (2005) and others have reported on teachers' tests and testing practices in various countries. These observed teacher-made tests and testing practices would not exist if teachers did not have theories about tests. There are patterns, and the patterns, the tests, and the testing practices are coming from somewhere. Barrette (2004) commented that second language achievement test drafts she studied for her report were influenced "by the course materials and syllabus, the instructor's preferred teaching methodology, [and] the student population" (p. 58). These are all elements of teacher theory (Figure 2-1), and her comments suggests some desired connection between teaching and testing. However, she also cautions that many tests she examined for her report also had fidelity to "teachers' assumptions about

what a test should be as opposed to matching test items with what learners do in class" (p. 67). So clearly, teacher theory, which traditionally relates to instruction, also generates teachers' apparent assumptions and decisions and practices and attitudes about tests. It may be the case that the products of teacher theory, whether instruction or tests, are different, depending on teachers' education. Note recent calls for teaching practices, informed by language use description frameworks, to more closely link to classroom tests (Nikolov, 2016), and for more support for teachers to learn testing concepts (Brindley, 1997; Edelenbos & Kubanek-German, 2004; Newfields, 2006). See also Moodie and Haany (2008) for a description of a second language conversation performance test that was overtly fitted to what learners experienced in class, through careful planning.

Definitions for the Teacher Theory Model

In this section, the components of the Teacher Theory Model (Figure 2-1) are defined and related to each other. First, some general comments: The general orientation of the model is from left to right. For example, "Teacher background" is on the far left, appearing to be a foundation for "Current teaching" and "Current and past education." "Current state of teacher theory" appears on the far right, suggesting an end state of some sort, generated by "Teacher background" and mediated by "Current and past education," "Current teaching," and other model components. Single headed arrows suggest that one component influences another, while double headed arrows suggest that two or more components have a potential mutual influence. For instance, it would be hard to imagine that "Current teaching" would influence "Teacher background." "Teacher background" existed before "Current teaching." Thus, there is an implied chronological sequence. But, "Current teaching" might influence what summer training workshop a teacher might choose to take ("Current and past education"). And in a turnaround, "Current and past education" will be drawn upon by a teacher to interpret and guide their "Current teaching." Finally, none of these hypothesized relationships are set in stone, and the components themselves have fuzzy boundaries. For instance, "Institutional context" might be seen as being very similar to "Problems." Is it not within an institution that problems occur, and must be solved? This is what makes models so useful. They create and focus discussion.

Teacher background. The definition for "Teacher background" in Figure 2-1 would be aspects of a teacher's history that are relevant to their current role as a second language teacher/tester, including language learning experiences with the target language (Golombek, 1998; Griffee,

2012a), second language ability or status (high competence versus low competence, native speaker versus non-native speaker) (Frain, 2009; Kikuchi, 2005), experiences with tests as a language learner (Barette, 2004), and amount and types of teaching/testing experience (Frain, 2009; Griffee, 2012a; Phillips & Abbott, 2011; Shepard, 2000a). This last aspect, that of teachers' experiences with teaching and testing, is of particular interest because it speaks to "cultures" of teaching and testing that may operate beyond the level of the institution. Cheng, Rogers, & Huiqin (2004, p. 362) noted that second language teachers in Hong Kong and China wrote classroom tests and used "procedures" that "tend to mirror those of the external tests," such as the TOEFL, whereas Phillips and Abbott (2011) noted that American second language teachers had *yet* to be influenced by the ACTFL OPI in terms of their testing practices. The ACTFL OPI (American Council on the Teaching of Foreign Languages Oral Proficiency Interview) is a standardized interview test that has been in widespread use since 1984 (ACTFL, 2018). The two single headed arrows in Figure 2-1 reflect the hypothesis that "Teacher background" influences teachers' "Current and past education" and "Current teaching."

Institutional context. The definition for the component "Institutional context" in Figure 2-1 is related to Golombek's conception of "context" (1998, p. 452) which "includes the institutional and sociopolitical setting along with the time, place, and actors within the setting." This has to do with teachers' knowledge of the institutional setting, such as expectations of teachers, the sorts of learners there are at a school, and what the second language program is supposed to accomplish. What is meant by learners "accomplishing" something in a program may be stated on school websites, etc., but is more likely unstated, and the two might not be the same thing (Griffee & Gorsuch, 2016). Teachers' knowledge of this might be unspoken, but they would know it or have a sense of it after spending time in a program.

Knowledge of an "institutional context" also has to do with constraints on teaching and testing imposed and supported by a specific institution, such as large class sizes, required tests, required materials, etc. Not all institutional contexts impose negative constraints. A school might invest in having smaller classes, and better teacher access to appropriate technology. But often the institutional context is seen as potential clashes "between personal values and institutionalized role requirements and expectations" (Bullough, 1989, p. 79). In other words, the institutional context is the reality in which teachers must operate, and carry out their "Current teaching." If a school wants teachers to use a particular textbook, then a teacher needs to figure out how to use it in line with his or her

teacher theory. Such a textbook might (or might not) supply ideas on test content and perhaps test item formats (fill in the blank, short answer, etc.)(Barrette, 2004; Phillips & Abbott, 2011).

The "Institutional context" partly accounts for the context-dependent nature of teacher theory. When teachers proceed in their careers to teach at another school, perhaps in another country with a different set of learners, they experience a period of disorientation as they further develop their teacher theory to learn the new institutional context (Gorsuch, 2007; Gorsuch & Griffee, 2018). The single headed arrow from "Institutional context" to "Current teaching" reflect the hypothesis that the institutional context influences "Current teaching," as given in the examples above. The two headed arrow between "Institutional context" and "Problems" suggests that whatever "Problems" occur (unmotivated or unprepared students, for example, Griffee, 2012a), may be created or mediated by an institutional context. Conversely, the institutional context may also provide an answer to problems teachers encounter, such as having a supervisor or colleague who can help (Griffee, 2012a).

Current teaching. The definition for the model component "Current teaching" is "whatever courses the teacher is teaching" (Griffee, 2012a, p. 217). This invokes teachers' engagement with the curriculum, syllabus, materials, and outcomes of specific courses, and more importantly, their "interactions with students" in those courses (Griffee, 2012a, p. 217). Current teaching is a powerful impetus in "Current teacher theory." It provides an immediate context and need for goal-directed lesson planning, and planning and giving tests and quizzes. Whatever a teacher's current teaching assignment happens to be, this will change what he or she does with quizzes and tests. Teaching courses with large class sizes may constrain a teacher from doing speaking tests with students individually. They may opt instead to record pairs or groups of students talking to each other (Moodie & Haany, 2008; see Grogan and also Shaver in this book). They may rely more on self- and peer-assessments of those recordings, not only due to logistics, but also because they believe self- and peer-assessment contributes to language learning, and is in line with their conception of the course outcomes (Venema, 2002). An instructor with smaller classes may opt to let learners try a speaking task twice, which may result in greater learner attention form and potentially better performance (Hawkes, 2012). One instructor, De Silva (2014), created scoring scales with five performance levels for both writing and speaking performance tests due to the heterogenous, mixed ability level classes she was assigned to teach. She also believed that teaching the scoring scales to

learners would support their growth in self-assessment, an apparent course outcome.

"Current teaching" is connected to every other component in the Teacher Theory Model (Figure 2-1). The single headed arrow from "Teacher background" to "Current teaching" suggests that a teacher's background influences what classes a teacher can teach and will seek out to teach. The double headed arrow between "Current and past education" and "Current teaching" suggests that a teacher's past or current education will influence not only what classes they are qualified to teach and can teach (the arrow going in one direction), but also how they teach them and use tests in them (the arrow going in the other direction). Teachers will undoubtedly draw upon classes and workshops they have taken to make day-to-day decisions about their teaching and testing. For instance, Edelenbos et al (2004) claim that German-language teachers in the Netherlands working in classrooms in the 1990s had little testing training in that teacher education policy "had opted against" such training (p. 270). Some teachers in the study said that they used diagnostic testing techniques "embedded" into their instruction (asking learners questions to check understanding, for instance) but the recordings of the classes they taught showed little evidence of it. On a more positive note, Purpura (2016, p. 202) believes that the Common European Framework of Reference for languages (Council of Europe, 2001) has improved the "assessment literacy" of teachers across Europe (see Chapter Five on Communicative Competence and Language Use Description Frameworks and Second Language Tests). The single headed arrow from "Institutional context" to "Current teaching" suggests that it is the institution that determines class size, teaching load, and course assignments for teachers.

Problems. The definition for the model component "Problems" is issues or challenges that arise as a function of "Current teaching." For instance, Griffee (2012a, p. 223) learned that novice teachers' response to learner's lack of motivation and uncommunicativeness in classes was to retreat "to a survival mode teaching vocabulary and grammar." Gorsuch and Griffee (2018) found a more constructive response from an experienced second language teacher who worked with a learner who had particularly challenging pronunciation problems. When the learner failed a high-stakes test, the teacher was confronted by an advisor from the learner's department. This is an example of "Problems." Her response to the advisor was to identify problems in specific areas of the learner's performance on the test. She also pointed out features of the test and testing procedure, such as having two independent raters, that ensured consistency of the learner's score. The double-headed arrow going from

"Current teaching" to "Problems" suggests that problems unique to the classes a teacher is working with will arise, and that problems the teacher notices may change their teaching.

Current and past education, and Current state of teacher theory. The two final definitions here are for "Current and past education" and "Current state of teacher theory." They have been touched upon in the previous definitions. Most components of the Teacher Theory Model (Figure 2-1) are hypothetically interconnected (shown as arrows), and thus elements of the definitions are interconnected. But brief, additional definitions are given here. The "Current and past education, classes, workshops" component is any current or past courses, seminars, or workshops a teacher has taken. This could be at the undergraduate or graduate level, leading to a degree or one-off summer workshops. This would also be participation in test development projects that are school-sponsored or sponsored by universities or commercial testing companies, or short training courses intended to familiarize teachers with frameworks of language use description such as the ACTFL *Guidelines* (American Council on the Teaching of Foreign Languages, 2012a) or *CEFR* (Common European Framework of Reference, Council of Europe, 2001, 2018). It will be argued later in this book that language use description frameworks such as the ACTFL *Guidelines* and *CEFR* may constitute a resource for teacher learning in themselves in that they are published and publicized, and construed by some stakeholders in second language education to be a means of establishing accountability (Phillips & Abbott, 2011; Purpura, 2016). Thus, it may not be necessary to have active, externally motivated coursework or workshops in these descriptive frameworks for teachers' theories to be influenced by them.

The "Current state of teacher theory" component represents teachers' action-oriented theories that inform their day-to-day decision-making about testing. "Day-to-day" does not mean that teachers' actions and decisions are random, or lack internal coherence. There are patterns and order, generated by contingent teacher theory, in response to the interaction between "Current teaching," and "Current and past education." The relationships between the three components "Current state of teacher theory," "Current teaching," and "Current and past education" are potentially dynamic, with "Current and past education" about second language testing serving as a possible agent of change with teachers' conceptions of what they do on in the context of their "Current teaching." Yet education is only a potential agent of change. Even with courses and workshops, some testing researchers have found limitations on teachers' "assessment literacy" (teachers' ability to interpret educational literature

on testing, write tests, and assess students with "minimal bias" Newfields, 2006, p. 51). Manley (1995) noted that despite the popularity of "oral testing" at conferences and multiple workshops and discussions on oral testing within a school district, teachers had no way to find "an immediate link to the teachers' classrooms" (p. 94). In other words, teachers grappled with how to do oral testing, and what the test scores would mean in terms of the courses they were teaching. For instance, how much weight should be given to a students' grade based on an oral test? Sixteen years later, Phillips and Abbott (2011) reported much the same. Despite the salience of the ACTFL *Guidelines*, teachers had not readily adopted guidelines of achievement into their theory and practice of tests.

The reason education is only a potential agent of change in teacher theory can be explained in part by the complexity represented in the Teacher Theory Model. Rea-Dickins (2004) notes that teachers act in many roles in the classroom and in the institution, and as a result, face "significant dilemmas" in that they are "sometimes torn between their role as facilitator and monitor of language development and that of assessor and judge of language performance as achievement" (p. 253). Here the idea of "achievement" brings us to learners' grades, which teachers may believe requires formal tests and quizzes such as those presented in this book. Achievement also brings to mind teachers' awareness that their school's standing may be construed by administrators as learners' scores on standardized tests. The Teacher Theory Model takes into account the classroom context in "Current teaching" (the need to award grades), as well as "Institutional context" (how administrators construe school standing). Thus, teacher theory is formed by exigencies other than "Current and past education."

Another explanation is how theory is arranged in the second language education field, which is illustrated by Griffee's (2012b) High Middle Low Model (HML Model). He posited three levels of theory: high-level theory, middle-level theory, and low-level theory, each of which is used by different actors in second language education for different purposes. An example of high-level theory would be communicative competence. An example of middle- theory would be more specific domain theories, such as second language acquisition theories, or language use description frameworks such as *CEFR*. Examples of topics addressed in low-level theory (or local teacher theory) are "What works for me and why," or "How my students learn" (Gorsuch & Griffee, 2018, p. 79). See Table 2-1.

Table 2-1: ***The HML model with examples***

Theory type	Also known as	Characteristics	Examples
High	Grand theories	Articulated in terms most come to agree upon as objective reality, publicly discussed, published Establishes fundamental and overt changes in theory and practice in the field; Answering the question of what reality *should* be	Test validity models and theories Communicative competence Language proficiency models
Middle	Domain theories	Articulated, publicly discussed, published Used to motivate research agendas; May or may not be intended for classroom applications	Theories from communication studies, applied linguistics, education, linguistics, psychology, etc. such as Deliberate Practice Theory, and Exploratory Practice Second language acquisition theories Test constructs Language use description frameworks
Low	Teacher theories	Not articulated, intuitive, not necessarily available for introspection or discussion, private, not published Used to solve the problem of classroom instruction and testing	What works for me and why How my students learn What my students need and how I deal with that What I put on my mid-term exam What I think about students' test scores and do I make future plans according to them

Note. Based on Gorsuch & Griffee (2018, p. 79)

It may not be the case that classroom teachers have a lot of experience with high-level and middle-level theories, even in M.A.-level graduate courses. There may have been survey courses with high- and middle-level theories as their content, but such courses may have been taught in generalities, without sufficient support for novice teachers and in-service teachers to apply the theories to the complexities of "Current teaching" contexts (Griffee, 2012a). And it may not be the case that researchers using middle-level theories to motivate their research projects will make specific suggestions about classroom applications for teaching or testing.

It might be argued that the high-level theory of communicative competence has gained a better foothold in teacher's thinking about classrooms. Certainly, communicative competence has been represented in the Common European Framework of Reference over time (Council of Europe, 2001, 2018; see however Fulcher, 2010a). But reports persist of communicative competence being construed in U.S. college departments as just having learners talk to each other about their opinions, without using reading, writing, or listening for communication and personal enrichment (Griffee, 2012a; Swaffar & Arens, 2005). Such is an anemic realization of communicative competence, existing below even the level of conscious notice and discussion (below middle-level theory).

Explaining how Teachers Plan, Write, Administer, and Score Tests, and then Use Test Scores

This second model proposes an explanation, ordered sequentially in terms of the life of a test, of how teachers can plan, write, administer, score tests, and use test scores to teach. As with the Teacher Theory Model, the model was needed to design the questionnaire that probes the tests the teachers/testers offer in Chapter Four. The model is not intended to be prescriptive, but the sequential steps within the model can be used as guidelines. The model is primarily intended to be descriptive. The model takes into account teacher theory (high-, middle-, and low-level theories, Figure 2-1, Table 2-1) that inform test making and use in many contexts in the second language education field, including classrooms, institutions, countries, regions, and transnational companies. See Figure 2-2.

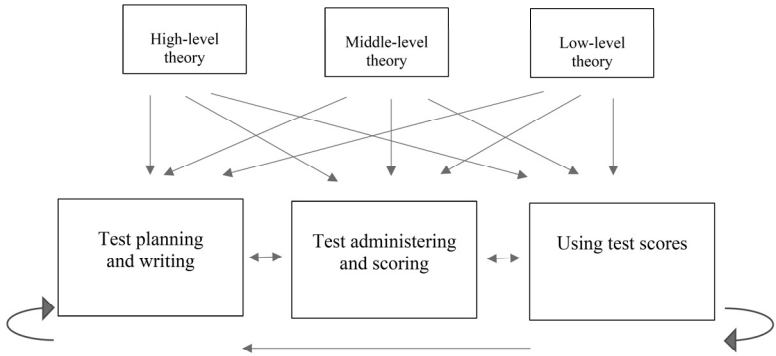

Figure 2-2. **Life of a classroom test model (based on Downing, 2006; Gorsuch & Griffee, 2018; Griffee, 2012a)**

Some general notes: The three larger components on the bottom row represent the focal points of the model, which shows the life sequence of tests. The starting point is the component on the left, "Test planning and writing," which then proceeds to "Test administering and scoring," and then on to "Using test scores." Teachers may then use test scores to return to "Test planning and writing" to improve existing tests, or to use items or test tasks or scoring criteria from the original test to create a new test, hence the arrow circling back to the left and returning to "Test planning and writing." At any point teachers/testers may move leftwards within the sequence. They may begin assembling materials for a test (part of the "Test administering and scoring" component) only to have an insight that would send them back to the drawing board ("Test planning and writing"). Perhaps they found the test directions as written were not clear, or they did not have at hand the right recordings for a listening test. If the teachers/testers had the luxury of giving a pilot test, they may have learned that a test task had seemed clear in the planning stage (listening to a recording and then putting the verbs used into a different form in a different text), but during the pilot learners simply looked at a previous page in the test and got their answers there, as contributor Yesica Amaya found. Within each of the three main components are multiple, more specific, steps that will comprise the definitions for the components detailed below.

The small components in the upper row represent high-level, middle-level and low-level (teacher) theory (Table 2-1). Any of these can inform any stage of testing, hence the multiple arrows leading from each of the

theory types to the three focal components of "Test planning and writing," "Test administering and scoring," and "Using test scores." The components of High-level, Middle-level, and Low-level theory are proxies for the Teacher Theory Model (Figure 2-1). In other words, they represent teacher theory. Teacher theory can draw upon high-level, middle-level, and low-level theories ("Current and past education") as well as knowledge of "Current teaching," or information or resources present in the "Institution," etc.

Definitions for the Life of a Classroom Test Model

Within each of the three focal components are specific steps for classroom tests and performance tests that comprise the definitions for the components. The steps were derived from Downing (2006), and from adaptations of Downing's steps set out in Gorsuch and Griffee (2018) specifically for second language teachers.

Test planning and writing. This component is comprised of the steps in which tests are planned and written. Planning includes deciding test content, purpose, item formats, and making plans for establishing consistency of scoring (test reliability). Writing includes writing actual test items for paper and ink classroom tests, and writing, borrowing, or adapting test tasks and scoring criteria for performance tests. The processes of planning and writing may play off of one another in ways that reflect how teachers must balance what learner knowledge or skills they want to capture or know how to capture, against how much class time or personal time they think they have to devote to administering or scoring a test. In other words, they use their teacher theory (Figure 2-1, "Current and past education," "Current teaching," etc.). See Table 2-2.

Table 2-2: ***Test planning and writing for classroom tests and performance tests***

Step	Definition	Examples
1. Overall test plan	The "big questions" stage. This sets the parameters of what a teacher wishes to do, and where questions are answered: What is the test purpose? Who are the test takers? Is the test high stakes or low stakes? What decisions will the test scores be used for? What do I wish to know about learners in	Questionnaire item examples: *Why did you write the test?* *What were the purposes of the test?* *Were you concerned at how long the test would take to administer?* *Were you concerned at how long the test would take to score?*