

Language Assessment Literacy

Language Assessment Literacy:

From Theory to Practice

Edited by

Dina Tsagari

Cambridge
Scholars
Publishing



Language Assessment Literacy: From Theory to Practice

Edited by Dina Tsagari

This book first published 2020

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

Copyright © 2020 by Dina Tsagari and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-4835-X

ISBN (13): 978-1-5275-4835-0

TABLE OF CONTENTS

Introduction	1
<i>Dina Tsagari</i>	

PART I: Defining Language Assessment Literacy

Chapter One.....	8
Operationalizing Language Assessment Literacy	
<i>Glenn Fulcher</i>	

Chapter Two	29
Finding out What Learners Know—and?	
Reflections on Teachers' Language Assessment Literacy	
<i>Gudrun Erickson</i>	

PART II: Understanding Language Assessment Literacy

Chapter Three	50
An Empirical Study on Romance Language Teachers' Subjective Theories Regarding Assessment Purposes and Good Tests	
<i>Anastasia Drackert, Carmen Konzett-Firth, Wolfgang Stadler and Judith Visser</i>	

Chapter Four.....	71
Greek Language Teachers' Conceptions on Purposes and Approaches of Assessment	
<i>Georgia Solomonidou</i>	

Chapter Five	97
How Much Do We Really Know About Alternative Assessment? Investigating YL Teachers' Perception of Assessing Through Games	
<i>Christina Nicole Giannikas</i>	

Chapter Six.....	116
Will A Boom Lead to A Bloom? Or How to Secure a Launch of Language Assessment Literacy in Ukraine <i>Olga Kvasova</i>	
Chapter Seven.....	137
National Literacy Assessment and Language Assessment Literacy <i>Yiasemina Karagiorgi and Alexandra Petridou</i>	
PART III: Expanding Language Assessment Literacy	
Chapter Eight.....	160
Inside the Black Box of L2 Reading Ability Measurement: Interfaces Between Assessment Literacy and Validity <i>Asma Maaoui and Dina Tsagari</i>	
Chapter Nine.....	192
Students' Voices on the Unspoken Oracy Construct: "Find out How to Do it on Your Own!" <i>Anne-Grete Kaldahl</i>	
Chapter Ten.....	214
Sociopragmatic Assessment Literacy in the Russian Language Classroom <i>Anna Dreher and Wolfgang Stadler</i>	
Chapter Eleven.....	237
Assessment Aligned with Active, Lifelong Learning: Evidence from Cyprus <i>Sviatlana Karpava</i>	
Chapter Twelve.....	265
Designing an Assessment Environment for Gifted Students: Focus on Teacher Education <i>Marina Perevertkina</i>	
Chapter Thirteen.....	286
Washback Effect and Teachers' Assessment Literacy: The Stepping Stone for More Learner-centered Exam-classes <i>Irini-Renika Papakammenou</i>	

Chapter Fourteen	304
The Challenge of Aligning Vocabulary and Grammar Difficulty Levels with CEFR Proficiency Ratings: The Case of the E-CAT	
<i>Jack Burston, Androulla Athanasiou and Maro Yiokari Neophytou</i>	

INTRODUCTION

The field of language testing and assessment (LTA) has recently recognized the importance and underlying theoretical and practical underpinnings of Language Assessment Literacy–LAL, an area that is gradually coming into its own with an evolving research agenda (Taylor 2013; Pill and Harding 2013; Inbar-Lourie 2008; Fulcher 2012, Scarino 2013; Vogt and Tsagari 2014).

This book is an edited volume based on the collection of theoretical and research papers within LAL titled “Language Assessment Literacy: Theory and Practice”. The book was motivated by the international conference organized as part of the dissemination of an EU-funded project that aimed at promoting teachers’ LAL: “Teachers’ Assessment Literacy Enhancement – TALE” Erasmus+, Cooperation for innovation and the exchange of good practices, Strategic Partnerships for school education, KA2, European Lifelong Learning Programmes, 2015-1-CY01-KA201-011863 (<http://taleproject.eu/>). The chapters in the volume comprise papers that were presented during the conference and others that were recruited through invitations sent out via online LTA networks and professional lists.

The volume addresses the notion of LAL through a multidisciplinary and multifaceted approach for the wide language assessment research, teaching and academic community and beyond. The chapters of the volume examine various LAL issues that range from definitions of assessment literacy, conceptualizations and research in LAL to relationships between LAL and teachers’ assessment practices, test washback, assessment of young learners and gifted students, computer-assisted language assessment, alternative assessment paradigms, and specific language skills such as reading and oracy skills.

The book comprises 14 chapters organized in three thematic areas:

Part I: Defining Language Assessment Literacy (2 chapters)

Part II: Understanding Language Assessment Literacy (5 chapters)

Part III: Expanding Language Assessment Literacy (7 chapters)

The chapters intertwine and entangle these issues from different perspectives. Contributors to the volume address the issues identified from

theoretical and empirical points of view with important implications for teachers as the critical stakeholder group. Studies of both a cross-sectional and longitudinal nature are included as well as studies conducted with young and adult test takers in either high- or low-stakes environments. The languages examined are many and varied, such as English, German, French, Italian, Spanish, Russian, and Greek, and so are the contexts of inquiry: Austria, Cyprus, Greece, Norway, Tunisia, and Russia.

The first two chapters outline and reflect on the concept of language assessment literacy. Specifically, in the first chapter, Glenn Fulcher builds upon his previous research (Fulcher 2012) to extend the concept of LAL and presents a model of LAL acquisition for language teachers. The model has a philosophical basis and is illustrated with practical examples of materials and activities within the context of an MA programme for language teachers. In the second chapter, Gudrun Erickson touches on a number of very important issues that can be used for and by teachers (pre- and in-service), students and other stakeholders and provides concrete examples of international sources and resources for the enhancement of teachers' language assessment literacy.

Underpinned by considerations of a broad view of validity, the next three chapters focus on teachers' conceptions of assessment and assessment literacy in second language assessment. In the first of these chapters Anastasia Drackert, Carmen Konzett-Firth, Wolfgang Stadler and Judith Visser present selected results of an empirical study investigating language assessment practices and training of in-service secondary school teachers of French, Italian, and Spanish in Austria and Germany. The survey results present teachers' reported practices and their subjective theories on the purposes of assessment. The authors emphasize the need to involve teachers' subjective theories in assessment practices. In the following chapter Georgia Solomonidou, also stresses the same point that perceptions of assessment are important on building healthy systems of LAL in student assessment of the Modern Greek language as a first language. Through her sequential mixed-methods research she shows that the conceptions of secondary school teachers are in alignment with the current shift of assessment in the local context to be used for enhancing teaching and learning.

Christine-Nicole Giannikas in her chapter tells us that it is also important to consider the assessment practices of teachers of YLs. Data were gathered via a global online survey regarding the use of games as an assessment tool, teachers' perspectives, and assessment training. The outcomes show the need for the enhancement of LAL, and the need to include alternative assessment in particular in teacher training programmes.

The next chapter written by Olga Kvasova focuses on three LTA courses developed by the author and delivered in a classical university in Ukraine at undergraduate and graduate levels. The author discusses the course components and the methodology of course delivery, e.g. modes and formats of instruction and techniques of training and assessing the LTA knowledge and skills of participants. The findings of the study are of use to educators and institutions who aspire to enhance teachers' LAL at national and international levels.

This section of the book closes with a chapter by Yasemina Karagiorgi and Alexandra Petridou which presents a nationwide study that focuses on the Programme for Functional Literacy (PFL), implemented in Cyprus that aims to identify students "at risk" at Years 3 and 6 across all public primary schools of the country. The chapter argues in favour of maintaining in place the longitudinal national assessment programme as it enhances aspects of teachers' language assessment literacy.

The third part of the book comprises equally interesting chapters that research various aspects of LAL. For example, the first chapter of this section written by Asma Maaoui and Dina Tsagari examines teacher- and context-related factors in reading assessment in Tunisian higher education within the interface of LAL and test validity. The results show how intrinsic and extrinsic factors shape the teachers' LAL at the individual level and identify the challenges caused by inadequate teacher assessment literacy underlying the test design process.

In the north of Europe, Norway, Anne-Grete Kaldahl examines the lower secondary students' perceptions of the assessment of oracy across subjects, drawing on a rhetorical topos analysis of qualitative data generated through focus group interviews in two public schools. The results have important implications for teachers and their LAL enhancement.

Wolfgang Stadler and Anna Dreher look at sociopragmatic assessment literacy in the Russian language classroom. The authors examine coursebooks and test collections that reveal implications for the various stakeholders and offer useful information for improving teaching and assessment competence thus supporting the assessment literacy of key stakeholders.

Svetlana Karpava looks at how alternative assessment can be aligned with active, lifelong learning. In her study set in Cyprus the author investigates assessment in language courses for 20 university EFL tutors and 64 secondary school English teachers. The analysis of the data shows that the tutors like the idea of communities of shared practice and

developing into reflective practitioners but factors such as restrictions of time and curricula aims are not conducive to raising levels of LAL.

Marina Perevertkina researches an equally important aspect of LAL, that of designing a conducive assessment environment for gifted students. The study, set in Herzen State Pedagogical University of Russia, reviews the content of a course that was specifically designed with this aim in mind and outlines the strategies in choosing valid techniques of assessment, in terms of developing, scoring and interpreting results and assessment methods appropriate for gifted students.

Irini-Renika Papakammenou sheds light on another very important aspect of language assessment, that of test washback and its relation with teachers' assessment literacy. The chapter suggests how to decrease or even eliminate negative washback and provides ways to employ communicative methodologies and alternative assessment techniques based on a research study in a multi-exam context in Greece.

The final chapter written by Jack Burston, Androulla Athanasiou and Maro Neophytou addresses the challenge of aligning vocabulary and grammar difficulty with CEFR proficiency ratings. The authors propose a computer-adaptive test of English, namely the E-CAT, which can considerably enhance LAL levels for its users.

In conclusion and given its scope and nature, this edited volume is of use to teachers and institutions committed to making assessment literacy sustainable in diverse national and international educational contexts. The book also hopes to chart new avenues of research in the field of LAL in the future and become an important reference source and reading material in undergraduate and postgraduate courses in teacher education in colleges and universities around the globe.

References

- Fulcher, Glen. 2012. "Assessment Literacy for the Language Classroom." *Language Assessment Quarterly* 9(2):113-132.
- Inbar-Lourie, Ofra. 2008. "Constructing a Language Assessment Knowledge Base: A Focus on Language Assessment Courses." *Language Testing* 25(3):385-402.
- Pill, John, and Luke Harding. 2013. "Defining the Language Assessment Literacy Gap: Evidence from a Parliamentary Inquiry." *Language Testing* 30(3):381-402.
- Scarino, Angela. 2013. "Language Assessment Literacy as Self-awareness: Understanding the Role of Interpretation in Assessment and Teacher Learning." *Language Testing* 30(3):309-327.

- Taylor, Lynda. 2013. "Communicating the Theory, Practice and Principles of Language Testing to Test Stakeholders: Some Reflections." *Language Testing* 30(3):403-412.
- Vogt, Karin, and Dina Tsagari. 2014. "Assessment Literacy of Foreign Language Teachers: Findings of a European Study." *Language Assessment Quarterly* 11(4):374-402.

PART I:

**DEFINING LANGUAGE
ASSESSMENT LITERACY**

CHAPTER ONE

OPERATIONALIZING LANGUAGE ASSESSMENT LITERACY

GLENN FULCHER

Defining Language Assessment Literacy (LAL) is important; but only as a precursor to achieving it (Stiggins 1991). It is arguably the case that existing texts designed for use in language testing and assessment programmes attempt to do just this (Davies 2008), and one recent introductory text in assessment is subtitled “theory and practice in assessment literacy” (Manning 2016). Yet, there is little explicit articulation between survey research into LAL needs, definitions of the construct, and teaching/learning strategies, even in research that is explicitly aimed at developing learning programmes and materials (Huai et al. 2006; Berry and Sheehan 2017; Sheehan and Monro 2017). Furthermore, when it is claimed that attempts to develop LAL are based upon a named philosophy such as “constructivism”, there is no attempt to demonstrate how the philosophy informs practice (see Conole et al. 2004, 17). This chapter builds upon the survey research of Fulcher (2012) to extend the definition of LAL presented there into an overt theory of pedagogy that can inform teaching and learning for LAL in one specific context, namely the professional development of practising language teachers undertaking the language testing modules on an MA programme. The approach described is embedded within a pragmatic theory of learning through doing that draws upon the metaphor of the apprentice. The resulting pedagogy is illustrated in practice through examples drawn from print and online learning materials. The chapter adds to the LAL debate by extending the discussion from definition and needs analysis to learning and teaching.

1. LAL Research

Traditional approaches to defining Assessment Literacy (AL) depend on classification rather than empirical research and tend to draw heavily upon psychometric theory and technical questions of validity (Bracey 2000; Popham 2009). This is encouraged by rather narrow definitions, such as: “Assessment literacy consists of an individual’s *understandings* of the *fundamental assessment concepts* and *procedures* deemed *likely to*

influence educational decisions” (italics in the original) (Popham 2011, 267). Similarly, a-priori “standards” have been produced upon which tests of AL have been created (Brookhart 2011). When these definitions or standards have been used as a basis for learning and teaching, they have led to very restricted course content that has changed little since Shaffer’s (1993) survey of curriculum content for AL.

Research into LAL has largely avoided such a-priori definitions. It does, however, begin with the work of Stiggins (1991). Although he attempted to define what was meant by a person who is “assessment literate”, he also raised the question of how assessment literate various stakeholders might be. These two questions have dominated LAL research to date, explaining the concentration of effort into questionnaire and survey design with the dual purpose of discovering the needs of specific groups of learners, and extracting from these needs a working definition of LAL (see Inbar-Laurie 2016).

Hasselgreen, Carlsen and Helness (2004) and Huhta, Hirvalä and Banerjee (2005) conducted a survey designed to uncover the assessment training needs of teachers in Europe, and reported that the most important areas for developing LAL are in the areas of portfolio assessment, classroom tests, peer- and self-assessment, feedback (all arguably assessment for learning), interpreting scores, validity, reliability, statistics, item writing, and rating performance tests (all more related to summative assessment). DeLuca and Klinger (2010) surveyed 700 trainee teachers in Ontario using a questionnaire based upon an a-priori understanding of LAL, arriving at the following definition based on factor analysis and labelling: “...the understanding and appropriate use of assessment practices along with the knowledge of the theoretical and philosophical underpinnings in the measurement of students’ learning” (DeLuca and Klinger 2010, 429-420). Crusan et al. (2016) conducted an international survey of writing teachers, but this focused primarily on attitudes rather than needs and definitions. Sheehan and Munro (2017) conducted qualitative research into teacher attitudes to assessment, and their assessment needs, using a small convenience sample. However, no clear definition of LAL emerges. While there is a claim that the research was conducted as a preliminary to producing training materials there is no link between LAL theory and materials, and no discussion of pedagogy or principles to guide the construction of learning materials.

Fulcher (2012) conducted an online survey to discover the needs of language teachers, also conducting factor analysis on the closed items and expanding the factor definitions with reference to the qualitative accounts

provided by individuals. He arrived at an “expanded definition” of LAL as:

The knowledge, skills and abilities required to design, develop, maintain or evaluate, large-scale standardized and/or classroom-based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order to understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals. (Fulcher 2012, 125)

While the generic definition of LAL provided by Fulcher (2012) seems to have been widely accepted as a useful starting point for future research (Tsagari and Vogt 2017, 41), the question of stakeholder literacy still preoccupies researchers. Taylor (2013) quite reasonably suggests that different profiles of LAL might be needed for different stakeholder communities, such as professional testers, teachers, and admissions tutors. As a result of Taylor’s theoretical suggestions, researchers have concentrated on the development of ever more complex survey instruments in an attempt to provide an empirical basis to claims about the assessment literacy needs of the different stakeholder groups (Harding and Kremmel 2016). This avenue of research will continue to bear fruit, particularly when it seeks to elucidate the needs of identifiable groups of stakeholders, such as policy makers (Pill and Harding 2013), trainee teachers (DeLuca and Klinger 2010; Xu and Brown 2016; Hill 2017), students (Smith et al. 2011), and university admissions tutors (Deygers and Malone 2019), especially if new instruments can solve the sampling and psychometric issues outlined by Fulcher (2012) and confirmed as problematic in other studies (e.g., Xu and Brown 2017, 149).

However, Stiggins raises a third question that has so far not been addressed: the role of teaching and learning to achieve assessment literacy. He did not specify what should be learned, or how it was to be learned; but he considered teaching and learning to be the ultimate goal to be served by creating theoretical definitions and conducting needs analyses for stakeholder groups.

What will it take to create a society that is assessment literate? We must start with the right attitudes and then undertake a great deal of basic assessment training. (Stiggins 1991, 538)

Little has been done to address this question, although where the outcomes of LAL training have been investigated the evidence suggests that there are significant problems with pedagogy (Lam 2015). In what follows I present a model of LAL acquisition for language teachers as a critical stakeholder group. The language teachers are undertaking two language testing modules as part of their MA degree in Applied Linguistics and TESOL. The model has an explicit philosophical basis, illustrated with practical examples of materials and activities. Both the philosophical basis and the practical materials may not be the most appropriate (or even relevant) to other stakeholder groups, but they may nevertheless act as a starting point for considering other targeted pedagogies.

2. An Apprenticeship Model

I argue that literacy in language testing and assessment for language teachers is about designing and building tests for their own use, and the institutions for which they work. Many will also work with local or national educational authorities to build tests for wider use. It is therefore appropriate to use the metaphor of the apprentice for this audience.

While the modern conception of apprenticeship is purely practical, its original meaning included a grounding in theory and an understanding of society, as well as an ability to make things. This is echoed in Fulcher and Davidson (2007):

The practice of language testing draws upon, and also contributes to, all disciplines within applied linguistics. However, there is something fundamentally different about language testing. Language testing is all about building better tests, researching how to build better tests and, in so doing, understanding better the things that we test.

Sociolinguists do not create ‘sociolinguistic things’. Discourse analysts do not create discourses. Phonologists do not create spoken utterances. Language testing, in contrast, is about *doing*. It is about *creating tests*.

(Fulcher and Davidson 2007, xix. Italics in the original)

In ancient Greece craftsmen were referred to as “*demioergoi*”, derived from the two words for *public* (*demios*) and *work* (*ergon*). Hephaestus, the god of craftsmen, was responsible not only for the creation of objects, but also of civilization itself. There was no separation between theory and practice, knowledge and doing. Each informed the other. It was only in classical times that the link was severed. Aristotle changed how we see the world by referring to a craftsman as a “*cheirotechnon*”, literally translated

as a “hand worker”. From then on artisans were seen as makers of things divorced from the theory that informs the making. The result was a binary class system with the artisans in the inferior role. By Roman times the rift was institutionalized in the hierarchical structure of society. In his *Ten Books of Architecture* (circa 20 BC), Vitruvius said that “The several arts are composed of two things—craftsmanship and theory. Craftsmanship belongs only to those who are trained...in the work; theory is shared with all educated persons” (cited in Lester and Piore 2004, 98). Making and building things were no longer considered thinking activities.

This rupture was healed to some degree during the Enlightenment. Diderot argued that it was necessary to understand theory and make mistakes in order to become a master craftsman. The process of error resulted in greater knowledge and understanding of the theory that informs quality work. “Become an apprentice and produce bad results so as to be able to teach people how to produce good ones” (cited in Sennett 2009, 96). The concept of quality and the desire to achieve it inevitably require understanding, setting standards (in the sense of what counts as a “quality product”), and searching for better ways of achieving desired goals.

Knowledge must come through action; you can have no test which is not fanciful, save by trial. (Sophocles 496 BC to 406 BC)

This principle is illustrated most vividly in movements like the Arts and Crafts movement, and the workshops of great craftsmen like Stradivarius.

The metaphor of apprenticeship has four critical implications for education. Firstly, the learner is at the centre of the process. Secondly, the learner learns by doing. Thirdly, there is a master craftsman who is on hand to guide and mentor the learner. Fourthly, in their apprenticeship Peirce argues that (1878/1958, 328): “They should be made to feel that they are doing real and important work which was to appear in the digests of science and for the accuracy of which they are responsible....”

The notion of “responsibility” is essential for any apprenticeship model of learning. By doing and creating, the tests (or test fragments) produced by learners should be evaluated in terms of their usefulness for the purpose for which they were designed. Both the design effort and the evaluation are therefore to be informed by theory. But the apprenticeship model also implies a task-based pedagogy, which requires a pool of tasks that leads the apprentices to literacy through action. The structure and oversight are provided by the master craftsman, the person who is responsible for guiding learning, designing learning tasks, and providing critical feedback on both the product and process of the test item/task design.

This pedagogic model may encompass “reflective practice” and “community activities” (Xu and Brown 2016, 158), but it goes beyond them by proposing a structured pedagogic model that consists of apprenticeship tasks and the content to be learned. We deal with each of these in the next two sections.

3. Characteristics of Apprenticeship Tasks

Coomey and Stephenson (2001) conducted a meta-study to extract from research the key elements of successful online learning, which can be used as a starting point for understanding how both text- and web-based multimedia content can be used in LAL learning. They discovered from the range of research they reviewed that the structure and design of learning materials are considered successful when they engage four positive behaviours.

The first is the generation of *dialogue* around the input material and tasks. Interaction is varied by constructing either convergent or divergent goals for the participants (Pica et al. 1993, 13). Convergent tasks encourage collaboration and the achievement of specified outcomes. In language test development these are advantageous to show learners that test design cannot be done in isolation. Divergent tasks that resemble “game playing” with an element of competition have the capacity to generate challenges and critiques of ideas and practices, and they are particularly useful for considering opposing views of ethical test use or the role of tests in society.

The second is *involvement*, defined as the extent to which learners become engaged by the material, such that it generates intrinsic motivation. This is achieved if tasks are sufficiently challenging. While involvement is difficult to assess, I would argue that it grows out of *responsibility*: the feeling that valuable work is being done that will benefit the students, institutions and communities that language testers serve. This requires learners to engage with the social functions of language testing.

Thirdly it is necessary to provide *support* to learners while engaging with tasks. Feedback from other apprentices and the master is critical to learning. It is not an afterthought but integrated into the structure and progression of learning tasks (Black and Wiliam 2009).

Fourth is the structuring of *control* over an activity such that beginners are provided with more guidance, while more advanced learners are given more freedom to engage with tasks as they wish. This recognizes that with

greater knowledge comes the freedom to innovate; with innovation comes error; with error comes learning.

These four characteristics may be referred to as the “**DISC** qualities” (**D**ialogue, **I**nvolvement, **S**upport and **C**ontrol); but as useful as this research might be to help define how a “good task” may engage learners, it is not sufficient for task design purposes. We also require a classification of task elements that provides a blueprint for design; similar, indeed, to a task specification template in language testing. There are many to choose from. However, one of the most useful, having stood the test of time, is that provided by Candlin (1987), which I have adapted for the purpose of LAL operationalization in Table 1 below.

Input	Stimulus to generate features of DISC
Roles	The assignment of participant duties within the task
Settings	The context for which test use if required
Actions	What participants must do to achieve goals
Outcomes	The goals of the task
Objectives	What you expect participants to learn (learning outcomes)
Feedback	Evaluation of performance and outcomes to inform iterative learning and improvement

Table 1: Classification of Elements of LAL Tasks, adapted from Candlin (1987)

This structure will be exemplified in the discussion of sample tasks below.

4. The Content of Apprenticeship Tasks

Developing a content model for apprenticeship tasks was the purpose of Fulcher (2012), whose definition of LAL is provided above. That definition is drawn from three aspects of LAL that are presented in Figure 1.

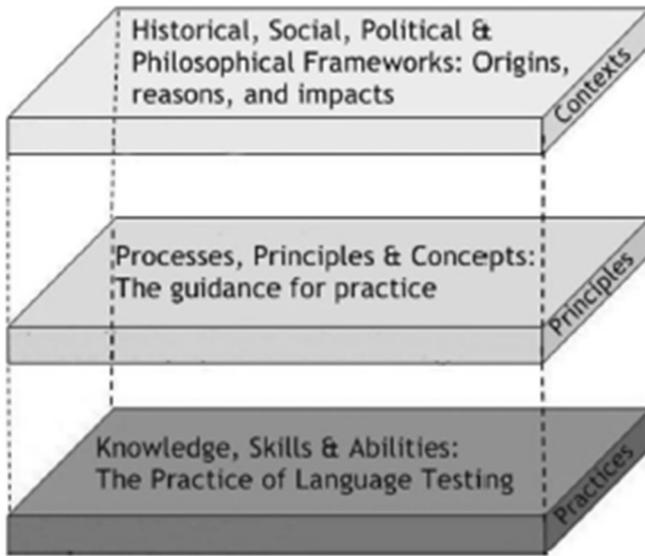


Figure 1: Language assessment literacy: An expanded definition (Fulcher 2012, 126)

Harding and Kremmel recognize the intention behind the apprenticeship content model:

Importantly, Fulcher (2012) proposes that theoretical concepts in testing textbooks and courses should be presented within the context of practical test construction, using the test development cycle as the scaffold and introducing core principles and core terminological knowledge along the way rather than merely introducing them as decontextualized components of LAL. An implementation of this approach can be found in Fulcher's (2010) textbook *Practical Language Testing*. (Harding and Kremmel 2016, 419)

Additionally, Fulcher (2015) made a deliberate attempt to address the top level of the model presented in Figure 1. Essentially, the model became a plan for the content of a LAL programme realized through inter-related texts with supporting web-based materials.

The content model in Figure 1 is not intended to represent a hierarchy, but a movement from the foundation in practice to the abstraction of theory. The exploration of theory is realized in the practice of assessment design, using theory to inform practical decision-making. In practice, this is achieved by structuring tasks around the test design cycle as illustrated

in Figure 2. Each element in the cycle is characterized by specific tasks, each of which involves decisions that are informed by theory and research.

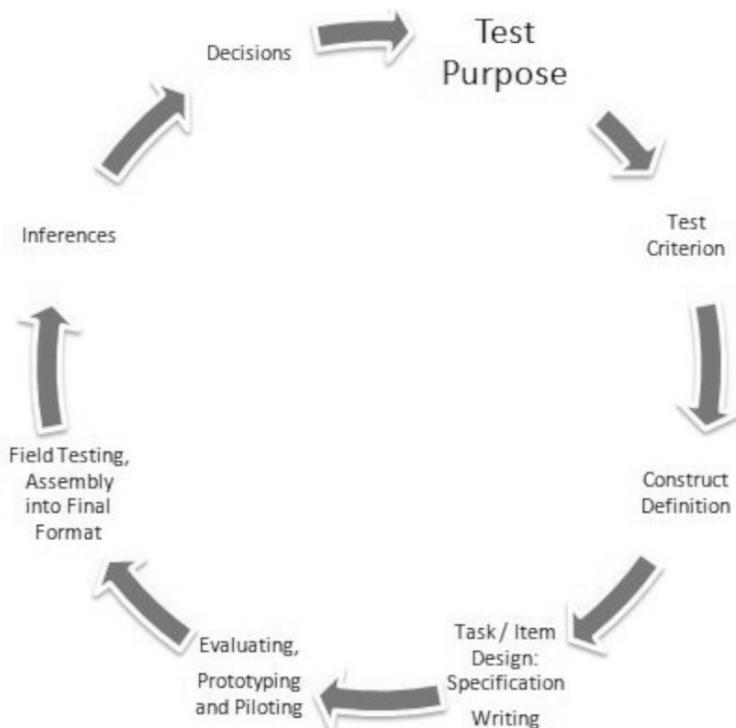


Figure 2: The Test Design Cycle (from Fulcher 2010, 94)

At each step in the cycle the aim is to create sets of learning tasks that implement DISC characteristics through tasks with explicit structures that aim to foster LAL in one or more elements of the content from the LAL model.

5. Illustrative Text-based Task Types

Below I present two tasks to illustrate the principles outlined above and relate each to both the characteristics and content of the apprenticeship model. The first would be used as part of the exploration of test purpose and use. The second is a design and review task that encourages the application of theory and data to item evaluation and revision. In each case, I make explicit the task elements from Table 1.

Task Type 1: Debate (Fulcher 2010, 23)

Content: Understanding social and political frameworks, impacts on individuals and society, through the exploration of a radical and controversial policy for a new use of educational assessments. Students have read Foucault, Mill, and a commentary on their differing views of tests before coming to the debate.

Sample Task

Read the following article. Do you believe that the creation of the database with individual dossiers for life is legitimate? Or does this just go to show that Foucault was right about the true intentions of governments? List the pros and cons on both sides of the argument. If you are working with a group of colleagues you may wish to organize a formal debate, with the motion “This house believes that a ‘testing record for life’ is an infringement of personal liberties and damaging to the future of the individual.”

Adapted from *The Times*

Every child in school numbered for life

All 14-year-old children in England will have their personal details and exam results placed on an electronic database for life under a plan to be announced tomorrow.

Colleges and prospective employers will be able to access students’ records online to check on their qualifications. Under the terms of the scheme all children will keep their individual number throughout their adult lives, *The Times* has learnt. The database will include details of exclusions and expulsions.

Officials said last night that the introduction of the unique learner number (ULN) was not a step towards a national identity card. But it will be seen as the latest step in the Government’s broader efforts to computerise personal records.

Last night teachers’ leaders, parents’ organizations, opposition MPs and human rights campaigners questioned whether this Big Brother approach was necessary and said that it could compromise the personal security of millions of teenagers.

Stimulus: Media items and news reports of a controversial nature.

Roles: The proposer and seconder for the motion, the principal opponent. Other participants join in the debate on one side or the other.

Settings: Social and political issues of interest to a wider audience than just language testers. Issues of privacy, selection, privilege and individual rights are particularly suitable.

Actions: Group members consider the arguments for or against together; preparation of arguments and initial speeches outside the classroom; preparation of a presentation (transferable skills).

Outcomes: A debate lasting up to one hour, involving two groups of protagonists. Primarily involving divergent interactions.

Objectives: Deeper understanding of the role of test scores, qualifications, and records of other personal achievements in life chances and employment. Explore the ethics of personal data sharing and access.

Feedback: Post-debate debriefing, perhaps inviting all participants to “vote” on whether or not they would favour the introduction of a lifetime database.

Task Type 2: **Item Review** (Fulcher 2010, 192-194)

Content: Prototyping and evaluating new test items. This task type is situated within a series of tasks designed to sensitize students to potential problems with closed-response items before attempting to craft and prototype their own. The level of control is relatively high as the flaws in the items have been identified for their usefulness in representing common errors made by novice item writers.

Sample Task

The following four items were designed to test the pragmatic competence of intermediate level adolescent learners of English. Review each item and identify any flaws that you find. Each item may contain multiple problems.

Item 1

Tony: Finishing the packing for our holiday is going to take forever at this pace.

Linda: Yes, and we’ve had quite a few late nights recently.

Linda implies that they

- (a) will miss their flight.
- (b) will be up packing late into the night*.
- (c) are both very tired.
- (d) need to work faster.

Item 2

Presenter: So, what do people feel about binge drinking on the streets of our towns and cities, and particularly the rising incidence of drunkenness among young girls? We went out and about with our microphone to find out. Here's what Tom, an office worker from Middlington had to say.

Tom: Well, I mean, it's up to them isn't it? Okay, you know, perhaps they don't have the money, so they have to get it from somewhere, and it could damage their health. But it's what they want to do. So I don't see the problem.

Tom's view is that

- (a) drinking causes social problems.
- (b) young people need more money.
- (c) heavy drinkers get liver disease.
- (d) the young can do as they please*.

Item 3

Listen to the exchange and answer the question.

Rebecca: There are only three tickets for the concert on Friday, and I've invited Sonya and Terry.

Angela: I guess I'll get over it in time.

Angela

- (a) is very upset.
- (b) feels left out*.
- (c) doesn't have time.
- (d) has her own ticket.

Item 4

Economist: It is highly likely that the credit crunch will become excessively tighter as the year progresses, forcing more small- to medium-

sized businesses into liquidation, and even resulting in many larger companies and high street brands being forced into the hands of the administrators. As governments become more involved with the banking sector many analysts foresee increasing levels of regulation that will bring an end to many opaque practices such as the trade in derivatives.

The economist argues that in the coming year

- (a) we will all have a harder life.
- (b) governments will buy banks.
- (c) some businesses will close*.
- (d) shops will get better managers.

Stimulus: Test items produced by colleagues or taken from existing tests. These need not be multiple choice.

Roles: All participants are critical evaluators. Ideally two or more groups should evaluate the items independently.

Settings: Prototyping new item types/evaluating existing items and tests.

Actions: Group members work convergently towards agreement on the strengths and weaknesses of each item.

Outcomes: An analysis of each item, and a decision about whether the item should be retained, revised, or rejected.

Objectives: Understanding what makes a “good” item. Acquiring the skills to engage in an item review, including key checks (for m/c items), a bias review, and an editorial review. Investigating the congruence of an item to the specification.

Feedback: Groups compare analyses, agree and/or critique the analyses of other groups.

This approach to operationalizing LAL has to date received favourable reviews (Read 2011, 304) because the application of the apprenticeship model creates a learner-centred learning environment in which theory informs decision-making by the learners. By doing so, they exercise judgement in order to arrive at responsible design decisions, taking into account the likely impact of their designs on individuals, institutions and society.

6. Illustrative Web-based Apprenticeship Tasks

Text-based tasks can target specific skills or knowledge in the test design process, but web-based tasks with their multi-media content can be used to structure integrated tasks that bring together many elements of the test-

design cycle into a single activity. The most integrated are scenario tasks. Six of these are available online (<http://languagetesting.info/whatis/scenarios/list.php>) as part of a package of multimedia resources to support apprenticeship-type learning programmes.

Scenarios were developed to simulate real-life assessment problems in which learners develop the craftsmanship of item and test development. In each case the stimulus is a set of multimedia materials that provide the setting for describing the need for a language test within a particular language use domain. Text and video are static, but news items about the use of tests within a domain, and links to relevant tests, are all updated on the fly every week. Learners take on the roles of professional test developers who have been hired to design an operational test for use in the domain. The outcome may be either a small number of tasks that would form part of a test, or a complete test.

The scenarios currently cover six areas: aviation, peacekeeping, call centres, medical communication, legal interpreting, and international student university admissions. I follow an eight-stage apprenticeship learning process, which I illustrate here with reference to the aviation language assessment scenario.

- (a) Test Purpose and Issue Awareness: students first study the language standards of the International Civil Aviation Organization (ICAO), and consider the consequences of making decision errors within this context. Students show how the ICAO communication standards might guide test development. Key constructs relevant to aviation communication are identified and listed.
- (b) Language Study: students study the restricted code of aviation language and then go to the website of the Aviation Safety Network to map communication failures against recorded accidents or near accidents. They are also presented with recordings of aviation communication where language contributed to accidents. In each case they are asked to identify the source and reason for miscommunication. Consideration is given to the use of non-standard language and its use when the restricted code is either violated or is not sufficient in unusual circumstances.
- (c) A literature review to discover how others have approached this problem, as well as the social and ethical issues that come into play.
- (d) Task design: students listen to and record live communications from an airport while tracking a plan on flightradar.com. The recording is used to develop a listening task for an aviation assessment battery. This may involve using the authentic recording,

or a scripted adaptation that the students write themselves and re-record for the test. Decisions at each stage of the audio design are recorded along with reasons for those decisions.

- (e) Item types to assess identified constructs using the listening texts are designed in groups. Students record design decisions and reasons for the decisions taken. Item specifications are created.
- (f) Qualitative Evaluation: students compare their tasks with those in existing tests used for the aviation industry. Where possible a prototype task is used with a language learner to practise protocol analysis leading to iterative task revision. Item specifications are evolved, each evolution is given a new number, and the rationale for each evolution is recorded.
- (g) Quantitative Evaluation: It is normally extremely difficult for learners to prototype and pilot the new tasks in this domain, but in others (such as higher education) access to a small number of participants is often possible. Where it is not possible artificial datasets may be created for students to analyze using a range of tools such as the reliability or DIF tools in SPSS, or Excel spreadsheets for item analysis like those provided on the website (<http://language-testing.info/statistics/excel.html>). Students are introduced to the software for quantitative analysis through practice under guidance in computer laboratories.
- (h) Report Writing: writing a group or individual report on the process of developing, prototyping and piloting a sample task.

In integrated projects of this kind all elements of DISC are present. Control is exercised only through the provision of stimulus material from the website. Support is available from the expert tutor, but the role of the tutor is primarily to guide the students to make their own design choices as apprentice craftsmen, to justify those choices, and to evaluate them using the data that they collect. The involvement comes through the requirement to create and evaluate an artefact with a critical purpose in the real world that serves a social good; and to achieve this discussion and debate are essential. Just as important, the students learn that the process of test design and development is also a research process from beginning to end. This research is not possible without the integration of theory and practice, the careful weighing of social, ethical and societal demands, and weaving these into an effective and practical assessment tool.