

CHAPTER ONE

OPERATIONALIZING LANGUAGE ASSESSMENT LITERACY

Glenn Fulcher

(In Tsagari, D. (Ed.) (2022). *Language Assessment: From Theory to Practice* (pp. 8 – 28). Cambridge: Cambridge Scholars.

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 Munro 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 can say 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.

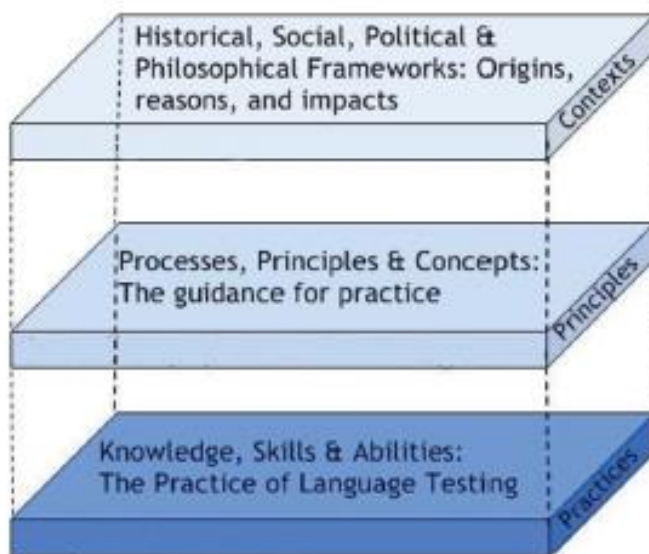


Figure 1: Language assessment literacy: An expanded definition (Fulcher 2012, 126)

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.

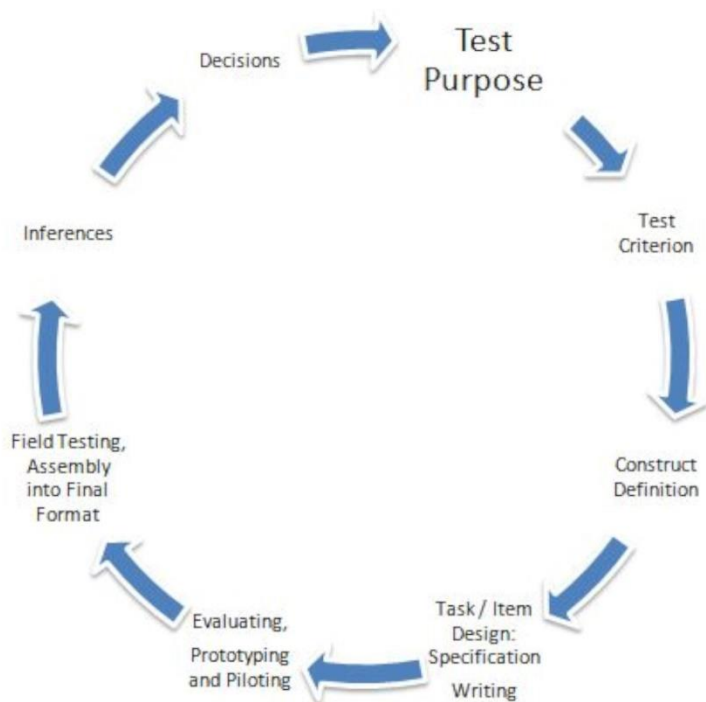


Figure 2: The Test Design Cycle (from Fulcher 2010, 94)

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://language-testing.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://languagetesting.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.

7. Crafted Artefacts

The assessment artefact (step h) is not some isolated essay question, but the outcome of a process of design, building, and evaluation. The report produced in step h is therefore a radically different kind of output from those normally generated in postgraduate programmes, many of which have at best a tangential link to module content. Creativity in thinking about appropriate outputs for assessment is possible when the task characteristics and content are made explicit. In a recent student-driven activity, learners conducted research into the nature, design and intention of test accommodations for test takers with specific learning difficulties in their L1, when taking language tests in an L2. This is an important aspect of test design (sometimes called “universal design”) that is frequently overlooked. The artefact produced at the end of the process was a podcast that problematized issues surrounding construct-irrelevant variance and made the research and potential solutions accessible to a wider audience. Following steps similar to those above, the final artefact integrated design issues with learning theory, accessibility and fairness, validity and reliability. The readers may judge the quality of this particular student artefact for themselves at <http://languagetesting.info/features/test/accommodations.html>, which is made available with the permission of the students. This is a prime example of genuine involvement in doing work for which they will take responsibility.

8. The Apprenticeship Classroom

The apprenticeship classroom for language teachers on an MA programme is primarily one of activity surrounding the content of a model of language assessment literacy delivered through tasks that follow the test design cycle. Some activities focus upon particular aspects of test design craft such as item writing and review, prototyping, or the analysis of pilot data. Other activities use the resources provided on the website languagetesting.info to support debate or group project work, such as the scenarios. Each task interweaves theoretical considerations, practical research, and test building. Students do this by creating small-scale tests and taking them through multiple steps in the design cycle, and then presenting the results to their peers in presentation sessions. Feedback is provided both by peers and the class tutor, which leads to iterative improvement. Assessment is closely linked to these activities. Writing up a scenario-based task design and analysis project as a research report is one option; but other options are possible, such as producing a broadcast quality podcast on some pressing assessment issue in the news, such as how language testing may be used to recruit health professionals to the health service while ensuring that patients are protected and can communicate with those who care for them in hospital. The apprenticeship classroom therefore looks and feels, as Peirce put it, like “doing” the work for real, experiencing the process, and taking responsibility for the outcome, even if it is in the service of completing an apprenticeship.

9. Conclusions

The LAL literature is replete with calls for the low literacy levels of stakeholder groups to be investigated and improved. What has not been addressed is how this is to be done. This chapter provides an initial articulation of one possible approach to LAL teaching and learning for language teachers on an MA programme, which is integrated with a theoretical model of LAL. It may go some way to encouraging teachers of language testing and assessment to consider how they use the resources available to them, including the website considered above.

It is suggested that the apprenticeship model is particularly appropriate for language teachers and others whose careers will require them to design, develop and administer language tests. The quotation from Peirce above makes

clear that the pedagogic model is underpinned by a Pragmatic philosophy that insists upon the interconnection between knowledge, learning, and experience. The “doing” as an apprentice is also an inherently social activity that requires collaboration with others. In education generally Dewey “...urged workers to assess the quality of their work in terms of shared experience, collective trial and error” (Sennett 2009, 288), which is fundamental to an apprenticeship model of LAL pedagogy. The apprenticeship model also resonates with the view that LAL activities are useful for the professional development of practising teachers, many of whom may also be undertaking MA programmes (Koh 2011; Fulcher 2017). It has long been recognized that engaging in research and collaborative design enhances professional practice: the link between doing and growing as a practitioner is an essential feature of the teacher research movement (Hopkins 2008, 38). Through apprenticeship-like tasks teachers acquire theory and knowledge as well as skills, becoming autonomous professionals with the capacity to make informed decisions (Goodenough 2010).

I make no claim that other pedagogic models may prove more useful, particularly for other stakeholder groups. However, I do contend that this is an explicit, deliberate pedagogy, as Brown (2011, 145) has called it, which to date has not been articulated in the literature. That LAL research has reached the stage of operationalization in pedagogies based on sound theory and research is also a sign of maturity within the discipline. Only in established self-confident branches of learning do current practitioners consider how best to serve the needs of the next generation. Interest in the theory and pedagogy of LAL is indicative of a profoundly optimistic field of endeavour.

The primary purpose of addressing the question “How do we do it?” is to achieve the assessment literate society that Stiggins envisioned. In turn, this requires what Fred Davidson (personal communication) calls “testing citizenship”. This term echoes Wenger’s notion of social participation, with specific reference to “the ethics of how we invest our identities as we travel through the landscape [of practice]” (Wenger, 92), bringing a relevant understanding of assessment to the decisions that we are required to make. This chapter makes one small contribution towards realizing such a goal through LAL practice.

6. References

- Berry, Vivien, and Susan Sheehan. 2017. “Exploring teachers’ Language Assessment Literacy: A Social Constructivist Approach to Understanding Effective Practices.” Paper presented at the 6th ALTE Conference, Bologna. Online: <http://eprints.hud.ac.uk/31898/>.
- Black, Paul, and Dylan Wiliam. 2009. “Developing the Theory of Formative Assessment.” *Educational Assessment, Evaluation and Accountability* 1(1): 5-31.
- Bracey, Gerald. 2000. *Thinking about Tests and Testing: A Short Primer in "Assessment Literacy"*. Washington DC: American Youth Policy Forum. ERIC Document No. ED 445 096.
- Brookhart, Susan. “Educational Assessment Knowledge and Skills for Teachers.” *Educational Measurement: Issues and Practice* 30(1): 3-12.
- Brown, A. 2011. Review of Language Testing and Assessment: An Advanced Resource Book. *Language Testing* 28(1): 145-148.
- Candlin, Christopher. 1987. “Towards Task-based Language Learning.” In *Language Learning Tasks*, edited by Christopher Candlin, and D. F. Murphy, 5-22. Lancaster Practical Papers in English Language Education Vol. 7. London: Prentice Hall.
- Conole, Grainne, Martin Dyke, Martin Oliver, and Jane Seale. 2004. "Mapping Pedagogy and Tools for Effective Learning Design." *Computers & Education* 43(1-2): 17-33.
- Coomey, Marion, and John Stephenson. 2001. "Online Learning: it is all about Dialogue, Involvement, Support and Control—According to the Research." *Teaching and Learning Online: Pedagogies for New Technologies*, 37-52.
- Crusan, Deborah, Lia Plakans, and Atta Gebril. 2016. "Writing Assessment Literacy: Surveying Second Language Teachers’ Knowledge, Beliefs, and Practices." *Assessing Writing* 28: 43-56.
- Davies, A. 2008. “Textbook Trends in Teaching Language Testing.” *Language Testing* 25(3): 327-348.
- DeLuca, Christopher, and Don A. Klinger. 2010. "Assessment Literacy Development: Identifying Gaps in Teacher Candidates’ Learning." *Assessment in Education: Principles, Policy & Practice* 17(4): 419-438.
- Dewey, John. 1923. *Democracy and Education: An Introduction to the Philosophy of Education*. Macmillan. Project Gutenberg Edition (<http://www.gutenberg.org/files/852/852-h/852-h.htm>).
- Deygers, Bart, and Margaret E. Malone. 2019. "Language Assessment Literacy in University Admission Policies, or the Dialogue that isn’t." *Language Testing*, <https://doi.org/10.1177/0265532219826390>.
- Fulcher, Glenn, and Fred Davidson. 2007. *Language Testing and Assessment: An Advanced Resource Book*. Routledge.

- Fulcher, Glenn. 2012. "Assessment Literacy for the Language Classroom." *Language Assessment Quarterly* 9(2): 113-132.
- Fulcher, Glenn. 2013. *Practical Language Testing*. Routledge.
- Fulcher, Glenn. 2015. *Re-examining Language Testing: A Philosophical and Social Inquiry*. Routledge.
- Fulcher, Glenn. 2017. "Cultivating Language Assessment Literacy as Collaborative CPD." Paper presented at the BALEAP Conference, Bristol.
- Goodnough, Karen. 2010. "The Role of Action Research in Transforming Teacher Identity: Modes of Belonging and Ecological Perspectives." *Educational Action Research* 18(2): 167-182.
- Harding, Luke, and B. Benjamin Kremmel. 2016. "Teacher Assessment Literacy and Professional Development." In *Handbook of Second Language Assessment*, edited by Dina Tsagari, and Jayanti Banerjee, 413-427. Berlin: De Gruyter.
- Hasselgreen, A., C. Carlsen, and H. Helness. 2004. "European Survey of Language Testing and Assessment Needs. Part One: General Findings." *Online: www.ealta.eu.org/resources [23.07. 2011]*.
- Hill, Kathryn. 2017. "Understanding Classroom-based Assessment Practices: A Precondition for Teacher Assessment Literacy." *Papers in Language Testing and Assessment* 6(1): 1-17.
- Hopkins, David, Bruce Joyce, and Emily Calhoun. 2008. *A Teacher's Guide to Classroom Research*. Open University Press.
- Huai, Nan, Jeffery P. Braden, Jennifer L. White, and Stephen N. Elliott. 2006. "Effect of an Internet-based Professional Development Program on Teachers' Assessment Literacy for All Students." *Teacher Education and Special Education* 29(4): 244-260.
- Huhta, A., T. Hirvalä, and J. Banerjee. 2005. "European Survey of Language Testing and Assessment Needs. Part 2: Regional Findings." online: http://users.jyu.fi/~huhta/ENLTA2/First_page.html.
- Koh, Kim H. 2011. "Improving Teachers' Assessment Literacy through Professional Development." *Teaching Education* 22(3): 255-276.
- Lam, Ricky. 2015. "Language Assessment Training in Hong Kong: Implications for Language Assessment Literacy." *Language Testing* 32(2): 169-197.
- Lester, Richard K., and Michael J. Piore. 2009. *Innovation—The Missing Dimension*. Harvard University Press.
- Manning, Anthony. 2016. *Assessing EAP: Theory and Practice in Assessment Literacy*. Garnet.
- Ofra Inbar-Laurie. 2016. "Language Assessment Literacy." In *Language Testing and Assessment*. Encyclopedia of Language and Education, edited by Elana Shohamy. New York: Springer.
- Peirce, Charles Sanders. 1878/1958. "Research and Teaching in Physics." In *Values in a University of Chance. Selected Writings of Charles S. Peirce*, edited by Philip Wiener, 325-330. New York: Doubleday and Co.
- Pica, Teresa, Ruth Kanagy, and Joseph Falodun. 1993. "Choosing and Using Communication Tasks for Second Language Instruction." In *Tasks and Language Learning: Integrating Theory and Practice*, edited by Graham Crookes, and Susan Gass, 9-34. London: Multilingual Matters.
- Pica, Teresa, Ruth Kanagy, and Joseph Falodun. 1993. "Choosing and Using Communication Tasks for Second Language Instruction." *Multilingual Matters*, 9-34.
- Pill, John, and Luke Harding. 2013. "Defining the Language Assessment Literacy Gap: Evidence from a Parliamentary Inquiry." *Language Testing* 30(3): 381-402.
- Popham, James. 2009. "Assessment Literacy for Teachers: Faddish or Fundamental?" *Theory into Practice* 48(1): 4-11.
- Popham, James. 2011. "Assessment Literacy Overlooked: A Teacher Educator's Confession." *The Teacher Educator* 46(4): 265-273.
- Read, John. 2011. "Review of Practical Language Testing." *Language Testing* 28(2): 302-304.
- Sennett, Richard. 2009. *The craftsman*. Penguin.
- Shaffer, David Williamson. 1993. "Assessment Literacy for Teachers." *Theory into Practice* 32(1): 118-126.
- Sheehan, Susan, and Sonia Munro. 2017. "Assessment: Attitudes, Practices and Needs." British Council ELT Research Papers 17.08.
- Smith, Calvin Douglas, Kate Worsfold, Lynda Davies, Ron Fisher, and Ruth McPhail. 2013. "Assessment Literacy and Student Learning: The Case for Explicitly Developing Students 'Assessment Literacy'." *Assessment & Evaluation in Higher Education* 38(1): 44-60.
- Stiggins, Richard J. 1995. "Assessment Literacy for the 21st Century." *Phi Delta Kappan* 77(3): 238.
- Taylor, Lynda. 2013. "Communicating the Theory, Practice and Principles of Language Testing to Test Stakeholders: Some Reflections." *Language Testing* 30(3): 403-412.
- Tsagari, Dina, and Karin Vogt. 2017. "Assessment Literacy of Foreign Language Teachers around Europe: Research, Challenges and Future Prospects." *Papers in Language Testing and Assessment* 6(1): 41-63.
- Vogt, Karin, and Dina Tsagari. 2014. "Assessment Literacy of Foreign Language Teachers: Findings of a European Study." *Language Assessment Quarterly* 11(4): 374-402.
- Wenger, Etienne. 2010. "Communities of Practice and Social Learning Systems: The Career of a Concept." In *Social Learning Systems and Communities of Practice*, 179-198. Springer, London.

- Xu, Yueting, and Gavin T. L. Brown. 2016. "Teacher Assessment Literacy in Practice: A Reconceptualization." *Teaching and Teacher Education* 58: 149-162.
- Xu, Yueting, and Gavin T. L. Brown. 2017. "University English Teacher Assessment Literacy: A Survey-test Report from China." *Papers in Language Testing and Assessment* 6(1): 133-158.