Mark my Words: The linguistic, social and political significance of the assessment of signed language interpreters

Chapter · January 2011

CITATIONS
5

READS
84

1 author:

Lorraine Leeson
Trinity College Dublin
100 PUBLICATIONS **311** CITATIONS

SEE PROFILE

"Mark my words"

The linguistic, social, and political significance of the assessment of signed language interpreters

Lorraine Leeson
Trinity College Dublin

This paper examines a critical area of interpreting research – assessment – as it pertains to the training of signed language interpreters (SLIs). The idea of testing as a wholly objective assessment of competence is problematized and issues that impact the testing cycle are teased out. These include the design and use of appropriate testing mechanisms. The attitude and expectations of external raters is analyzed, as is internal self-assessment as a means of creating active engagement in learning, and developing critical evaluation skills for later use in professional practice. We briefly consider a small number of empirical case studies that have explored aspects of assessing student interpreters. Finally, we draw attention to established pan-European frameworks, specifically the Bologna Process, for modelling fitness to practice at undergraduate and graduate level and usefulness in assessing the competencies of graduating interpreters. Given the limited empirical research that exists on the assessment of SLIs, this chapter suggests a number of themes where further studies would be beneficial.

Introduction

A crucial area of research that has been underdeveloped in the signed language interpreting (SLI) literature is the assessment of students and practitioners. Assessment of signed language interpreters is fraught with concern about fitness to practice, the competencies required to interpret effectively in a broad range of settings, idealized notions of desired competence versus minimal skill levels required to undertake the task at hand, as well as issues relating to language teaching, language status, and societal attitudes toward deaf communities and signed languages. Yet, many trainers of interpreters and many of those engaged in the assessment of interpreter quality have never had any formal training in applied linguistics generally, or specifically, in the area of language testing.

Fitness to practice as an interpreter entails more than being able to mediate between language A and B. In addition to bilingual fluency, interpreters must demonstrate competencies in "soft" skills (Humphrey and Alcorn 1996; Stewart, Schein and Cartwright 1998; Mindess, Holcomb and Langholtz 1999; Roy 2000; Janzen 2005; Napier, McKee and Goswell 2006; Napier 2009). These skills require assessment, both within training programmes as students develop skills (*formative assessment*), and as they complete training (*summative assessment*). In some countries, including the US, Australia, Canada, Sweden, and the UK, there are also established, sometimes statutory, accrediting licensing bodies that have responsibility for assessment of fitness to practice and specific domain related competencies.

Assessment of interpreters is a significant issue for academic progress and also has an important social function – policing access to and progress within the profession. Despite this, little has been published about the practice of assessment of SLIs; the available literature focuses mainly on descriptions of assessment protocols for registration or accreditation with national bodies. Traditionally, the act of interpreting (i.e., "doing" interpreting) has been more highly valued than giving thought to or theorizing about interpreting (i.e., "knowing about" interpreting). This is revealed by the relative dearth of interpreter research until the 1990s, reflecting prescriptive rather than empirically based descriptions of the act of interpreting. This is particularly true of the lack of empirical data on the assessment of interpreters of spoken and signed languages. Indeed, we are only beginning to consider what it is that students and professionals need to facilitate internalization of professional ideals and practices (e.g., Mindess, Holcomb and Langholtz 1999; Roy 2000; Napier 2005).

This chapter considers issues of assessment as they pertain to the training of interpreters. We problematize the notion of testing as wholly objective assessments of competence and tease out central issues that impact testing including (1) test type: achievement versus proficiency; (2) the testing cycle; (3) the Bologna Process; and (4) a review of some case studies that look at the assessment of student interpreters. Finally, in terms of the social and political rationale for testing, we will briefly discuss the issue of registration of interpreters versus the accreditation of training courses.

Test type: Achievement vs. proficiency

Testing is a facet of everyday life. We test learner drivers, we test language proficiency, and we test interpreters. In some cases, we test before, as well as during and at the end of training. In order to practice, some countries require

mandatory testing after completion of training to be registered with a state license or its equivalent. Our focus here is the testing of student interpreters, while acknowledging that there is what is known as "*washback*" from external factors such as national registration testing requirements (e.g., Sweden, UK, Australia, US, Canada).

Test purpose is a critical factor in determining the kind of test (i.e., pen and paper test, performance test, etc.) to be applied in a given setting. Test purpose is typically associated with the distinction between achievement tests and proficiency tests. *Achievement tests* are associated with the process of instruction. They assess what a student has achieved with respect to the curriculum taught on a programme. McNamara (2000) notes that achievement measures tend to be associated with portfolio-based assessments, course tests, as well as the assessment of course work. With achievement tests, assessment is focused on student development vis-à-vis course learning outcomes. There are a number of features of achievement testing: (1) it should support the teaching it relates to; (2) it may be self-enclosed, that is, focus on aspects of language grammar or use that has been covered in the curriculum rather than on language use in the wider world; (3) it can be highly innovative, and (4) is often associated with "alternative assessment", which stresses that assessment is integrated with the goals of the curriculum, and pushes for a constructive relationship between teaching and learning.

While achievement tests assess skill development in individual students with respect to what they have already learned, *proficiency tests* are concerned with future language use, without reference to the teaching process. Future language use is considered as the criterion against which proficiency is measured. Performance features are frequently incorporated into the design of proficiency testing. For example, aspects of the linguistic environment that medical personnel will encounter are included as test criterion (i.e., can a doctor communicate effectively with a patient who speaks language X?).

Here we see relevance for interpreter trainers and the potential difference regarding what educational programmes are asked to assess (most frequently, achievement) versus what post-graduation testing (typically registration bodies and assessment boards) is concerned with, namely proficiency. Educators, particularly towards completion of courses, are concerned with proficiency, though what is measured is required to be related to what has been taught, so there is always an achievement aspect involved in course-related testing.

An excellent example of sound creation of proficiency tests for interpreters is presented by Angelelli (2007), who outlines the processes involved in creating authentically driven performance features associated with medical interpreting in the test design phase. Working with stakeholders (native speakers of English, Spanish and Hmong, medical practitioners and pharmacists and practicing interpreters),

Angelelli's team created authentic, criterion driven test data to assess language proficiency and interpreting readiness of candidates before and on completion of specialist training. She notes that criterion-referenced tests allow testers to "make inferences about (a) how much language ability a test taker has (LP) and (b) how much interpreting ability a test taker possesses (IR), rather than merely how well an individual performs relative to other individuals." (2007: 71). McNamara (2000) makes the related point that testing is "about making inferences; this essential point is obscured by the fact that some testing procedures, particularly in performance assessment, appear to involve direct observation" (p. 7).

These factors are vitally important in evaluating interpreters. We are, on the one hand, assessing a student based on a given performance, but we also have to differentiate between language proficiency and interpreting competence. In practice, however, we seem to treat all tests as an indicator of all possible future performances in authentic settings. This is why it is critical to separate out the distinction between the *criterion* (i.e., the relevant communicative behaviours in the target setting) from the test.

The criterion can be described as an unobserved series of performances subsequent to the actual test. Those future, unobservable performances are in fact the target in proficiency tests. It is the characterisation of the essential features of the criterion that influences the design of the proficiency test (i.e., the real life settings that we aim to simulate in the test will influence the way the test process is structured). In contrast to the criterion, the test itself is a performance or series of performances, which simulates, represents, or is sampled from the criterion. It is only the test itself that is observed. Thus, with proficiency tests, we are making inferences about what a student *might or should* be able to do on the basis of our observation of the test situation.

One of the things that must be borne in mind is the idea that all language testing, which includes the testing of interpreters, links to real-world ability. While materials and tasks included in any kind of language test can be relatively realistic (or "authentic"), they can never be "real". Interpreter students are not "really" interpreting for a job interview, or at a meeting with the bank manager, and all parties are conscious of the fact that they are participants in a test environment. When assessing SLI candidates, for example, we ask questions like the following: "Will this student be able to cope with situation x or y in the "real world"? Or in a given "specific domain?" This in turn leads to the question of whether final assessments of students really are achievement focused or proficiency focused. Indeed, are faculty, students, employers, and registration bodies clear about what kinds of tests we have in place and why?

A number of other restrictions also apply with respect to criterion based testing or proficiency tests. First, limits always apply to the authenticity of tests

because of the differences in the conditions under which a test is administered. For example, a student may be prepared for interpreting a lecture on a specific issue for a test, but may, in future "real life" situations, not conduct detailed preparation, and the attention given to preparation for the test scenario may inflate their result vis-à-vis future performances. Thus, changing conditions can jeopardise validity, and with it, the generalizability of test results.

As McNamara notes, "The point is that observation of behaviour as part of the activity of assessment is naturally self-limiting, on logistical grounds if for no other reason … most test situations allow only a very brief period of sampling of candidate behaviour … oral tests may last only a few minutes" (2000: 9). Thus, most testing situations allow only a brief sampling of candidate performance and behaviour, the test is restricted to what it tells us about candidate performance in the test context, and from this context, we infer behaviours in other more generalised (or specific) settings. Given this, test validation issues also arise. We are obliged to consistently investigate the defensibility of the inferences made on the basis of test performance. We must also bear in mind that the act of observation can impact on behaviour, described as the "Observer's Paradox" (Labov 1969). This applies to test candidates as much as to the subjects of sociolinguistic studies. That is, the very act of observation can change the candidate's normal behaviour, and it is their normal behaviour that we want to see.

The issue of how we judge student interpreters' performances is critical. McNamara (2000) suggests that:

> In judging test performances … we are not interested in the observed instances of actual use for their own sake; if we were, and that is all we were interested in, the sample performance would not be a test. Rather we want to know what the particular performance reveals about the potential for subsequent performances in the criterion situation. We look, so to speak, underneath or through the test performance to those qualities that are indicative of what is held to underlie it. (p. 10)

Following from this we can ask what other factors influence the outcome of assessment and what can we do about these factors beyond being aware of them?

The testing cycle

The testing cycle is key. Very often those outside of education assume that testing is a straightforward process: it is assumed that you teach, you test, the students pass or fail, and their percentile score is a direct indicator of ability (e.g., a 60% grade is seen as equally getting the interpretation 60% right). However, a cycle of evaluation and review is required in creating appropriate testing frameworks. The actual

operational use of the test generates evidence about its own qualities – that is, as educators, we need to learn from the experiences we have in running tests, and take elements of that learning into account when we design and run subsequent tests.

Test design

Test design involves three key phases: (1) background issues, (2) test content and test method design, and (3) review validation and revision. The kinds of questions that arise influence test method and design, and include consideration of the constraints that impact on test design and implementation as well as the resources (financial, physical, human) that are available for test development and operation. The issue of test security is another essential component. Is test content unseen? Is it partially shared with candidates? Is test content known in advance? Finally, external factors must be considered including examination protocols within the institution that must be followed, assessment bodies who have agreed on specific formats for testing and reporting back or for accreditation of programs.

Test content and test method design

The first issue with regard to test content and test method design entails making decisions about test content – namely, what goes into the test. This links to how we see language and the use of language in a test situation (i.e., our view on test construct), and how we link test performance to usage of languages in a real-world interpreting context.

In major test projects, McNamara (2000) notes that teams may start by defining the test construct. The theoretical framework of the test may be the first step taken in test design, and this is frequently the case in vocational training where the training approach will determine the approach to assessment (or indeed, where external criteria for assessment determines the approach to training). This is a constraint that clearly operates on assessment within SLI training.

The second issue is that of identifying *test domain*. Doing so involves careful sampling from the domain of the test. We must identify the set of tasks or the kinds of behaviours that arise in the criterion setting. This may include introductions, managing turn-taking and ratification behaviours. In addition, it might also incorporate sociolinguistic norms in a given context, such as use of names and titles, or maintenance of register across the task as in Angelelli's (2007) study. In further considering the medical environment, other behaviours possibly incorporated are the capacity to interpret across registers and to bridge perceived gaps in education and world knowledge between medical practitioners and patients.

Further issues for consideration are *test method* and *authenticity*. Test method includes aspects of test design and scoring and issues of authenticity of the test, which we have mentioned already. Earlier, we outlined how criterion for proficiency tests should be based on job sampling, linked to what interpreters do in practice, but some constraints also operate. McNamara (2000: 27) notes that "… test design involves a sort of principled compromise". He says this because:

> On the one hand, it is desirable to replicate, as far as possible in the test setting, the conditions under which engagement with communicative content is done in the criterion setting, so that inferences from the test performance to likely future performance in the criterion can be as direct as possible. On the other hand, it is necessary to have a procedure that is fair to all candidates, and elicits a scorable performance, even if this means involving the candidates in somewhat artificial behaviour. (ibid.)

He also notes that, "As assessment becomes more authentic, it also becomes more expensive, complex and potentially unwieldy" (2000: 29). From this stems another important issue – that of validity. If tests cannot be controlled in terms of contextualisation to a greater or lesser degree, then there are issues arising in terms of how valid the test is. The issue of resource limitations is real and one that we must also acknowledge as impacting on what we do when we test.

A fourth issue is *test specification*, which refers to the set of "rules" for the test, comprising written instructions for implementation. Instructions function to make explicit the design decisions regarding the test and explicate the test's structure, duration, authenticity, source of testing material, the extent to which authentic materials are altered, response format, test rubric, and scoring system. Test materials are then written to these specifications.

The next stage in the test cycle is trialling the test, which should include taking feedback from test takers, followed by information gathering regarding modifications necessary before its implementation. Trialling tests can be difficult to do because of the constraints on time, resources and sample populations that exist in many countries, but the process is worthwhile in terms of solidifying the validity of the test process. It is highly probable that the small community of interpreters and interpreting students is also at the heart of the limited amount of published data on interpreter assessment for both spoken and signed language interpreting. Leeson (2007) and Bartlomiejcyk (2007) identify the limited number of empirical studies focused on the assessment of interpreters, especially interpreters in training. A determining factor may be that the anonymity of participants can be compromised because of the small pool of interpreting students in many countries. Making "mock" examinations available and ensuring that students have access to the test specifications prior to the test are hallmarks of university education. This

principle is applied in many "high stakes" testing domains too (e.g., in Canada, AVLIC make sample tests available to candidates), though in other regions, this is unfortunately not the case.

Rater attitude and expectations

Another key issue is who rates performance and their impact on testing outcomes. While informal judgment forming by peers and members of the deaf community is a standard component of being an SLI, judgment also impacts in a subjective manner on formal testing. However, if we set testing contexts up as objective, reliable indicators of ability, then we also have to account for the subjective judgment calls that raters make. Much testing focuses on SLI proficiency in communicative situations, with data marked live or, where recorded, marked *post-hoc*. Ratings awarded to a candidate are not solely a reflection of the candidate's performance, but are also a reflection of the qualities of the person who has judged that performance.

Following McNamara (2000), we can say that most rating schemes entail the assumption that if rating category labels are clear and explicit, and if raters are trained to interpret and apply these labels as per the intention of the test designers, then an objective rating process is possible. The reality is that rating is an intractably subjective process, containing a significant degree of chance associated with both the process and the rater. Given this, there are two choices – avoid direct testing or acknowledge the need for frameworks to be established which facilitate judgment by the raters. The latter can entail the establishment of "cut-off points" in hurdle tests (i.e., establishing the minimum cut-off point for passing on the basis of "good enough"/"not good enough") or employ a gradient continuum of marking (i.e., provide feedback to students in terms of their progress, mapping their performance to institutional marking scales). Crucially, raters must be trained to work with rating scales and understand what it is that they are being asked to mark, and for what purpose. They must have clearly outlined sets of rating criteria that they can return to when determining borderline cases, and they must demonstrate an understanding of clearly defined outlines of attainment aligned to institutional marking schemes are central to the process of training.

From our discussion thus far, we can say that both achievement and proficiency testing are used in SLI training, though sometimes we may combine (and perhaps also confuse) these. The majority of interpreter tertiary level training programs seem to include multiple points of testing over a significant period of time (i.e., two to four years) in order to evaluate student development and performance with respect to stated course objectives, learning outcomes, and external testing requirements. Despite the widespread use of assessment, there is a severe lack of empirical data assessing the processes applied or the outcomes of the approaches

we adopt. However, in Europe, there have been moves to "force" greater attention to the relationship between teaching and learning and assessment via the Bologna Process.

The Bologna Process

Institutional protocols don't exist in isolation; in the European Union (EU) there are EU-wide systems that have been implemented over the past number of years, under what is known as the Bologna Process[1]. This effort is a pan-European move towards greater transparency in education at third and fourth level, which has focused attention to issues such as course description, student workload, course learning outcomes and assessment criteria. This focus demonstrates a commonality among all facets of higher teaching and learning across the European Union (e.g., the introduction of so-called BAMA courses (i.e., a five-year pathway to Bachelor and, ultimately, Master level qualifications). The driving principle behind this initiative is that European higher education is vital to realizing a knowledge-based, creative, and innovative region.

The Bologna Process has led to the publication of sets of descriptors for use in assessment, known as the Dublin Descriptors. The document states that students with a Bachelors degree (Level 5 in their terms) should demonstrate the following competencies upon graduation:

1. have demonstrated comprehensive, specialized, factual and theoretical knowledge within a field of work or study and an awareness of the boundaries of that knowledge that builds upon and supersedes their general secondary education, and is typically at a level that, whilst supported by advanced textbooks, includes some aspects that will be informed by knowledge of the forefront of their field of study;
2. have the ability to apply expertise in a comprehensive range of cognitive and practical skills in developing creative solutions to abstract problems;
3. can apply their skills and competence in management and supervision in contexts of work or study activities where there is unpredictable change and review and develop performance of self and others;
4. can apply their practical skills in a manner that indicates a professional approach to their work or vocation, and have competences typically demonstrated through devising and sustaining projects and arguments and solving problems within their field of study;

1. See www.ehea.info/.

5. have the ability to gather and interpret relevant data (usually within their field of study) to inform judgments that include reflection on basic social, scientific and ethical issues;
6. can communicate information, ideas, problems and solutions to both specialist and non specialist audiences; and
7. can develop learning skills that are necessary for them to continue to undertake further study with a basic degree of autonomy and practical work within defined responsibilities (see http://www.mqc.gov.mt/pdfs/Grid%20of%20 statements.pdf).

The Bologna Process may also be useful in thinking about the assessment of interpreters trained in Europe and further afield. We can consider the Dublin Descriptors with respect to what we expect interpreters to do, and consider how relevant these are for fitness to practice upon graduation. We can further use these descriptors as a starting point for posing research questions in an empirical analysis of graduate interpreter competencies:

1. Requires graduates to demonstrate comprehensive, specialized, factual, and theoretical knowledge within a field of work or study and an awareness of the boundaries of that knowledge that builds upon and supersedes their general secondary education. This is typically at a level that, whilst supported by advanced textbooks, includes some aspects that will be informed by cutting-edge knowledge from their field of study. Educators expect that SLIs will have an understanding of the social, cultural and historical context that deaf people operate within. It is also expected that they will have a comprehensive understanding of the main issues that impact on interpreting. SLIs need to be able to harness theory; including empirically driven analyses of interpreting situations (e.g., Van Herreweghe 2002) on turn taking in meetings in Flanders, Brennan, Brown and MacKay (1997) regarding interpreting contexts Johnson (1991) regarding the importance of preparation, Stone (2005) regarding how deaf interpreters prepare in contrast to how hearing interpreters prepare). Further, interpreters should be able to utilize this knowledge in analyzing their own practice and, potentially, the practice of their colleagues, facilitating the development of critically reflective professionals (Leeson 2007).

2. States that graduates should have the ability to apply expertise in a comprehensive range of cognitive and practical skills in developing creative solutions to abstract problems. This follows from the last point. There is little benefit in teaching interpreters to "just do it" without giving time to reflection on why they do it "that way" in one situation yet another way in a different situation. Building bridges between theory and practice is essential, and educators and researchers alike need to consider further how we can assess student decision-making

processes. One approach is the Think Aloud Protocol (TAP), which has been used widely in translation studies, and more recently in aspects of signed language interpreting research (Sadlier 2007, 2009; Stone 2005).

3. States the expectation that graduates can apply their skills and competence in management and supervision in contexts of work or study activities where there is unpredictable change, and review and develop their own and others' performance. Interpreters need to bring their training to bear on their professional work, and working as an SLI entails unpredictable situations. A major issue is how we assess this competency. Clearly there is overlap with the last point (i.e., self-assessment of performance) but this also falls within the scope of internships or practical placement-based assessment. Interpreter placement is one of the most problematic situations in assessment, since work-based raters are usually not trained in assessment, and may provide "naïve" responses to student performance. While facing issues of inter rater reliability, educators still need to know how other interpreters, deaf clients, and interpreting agencies assess our students. At the same time, we want them to recognize that in-process assessment is formative (and achievement based) while successful completion of the program, and final assessments may be proficiency based. This is something we will return to again later in this chapter. An area ripe for empirical research, investigations may identify what criteria "naïve" assessors are judging and how they map onto the criteria established within formal training programs; examine how learning outcomes for established programs map onto work-oriented competencies, and explore how newly qualified interpreters skill-sets compare with interpreters with a certain level of experience.

Lets consider (4) and (5) together. We want interpreter students, on graduation, to solve problems relevant to their field of work, e.g., ethical dilemmas. We can assess this through Problem-Based Learning (PBL) (Sloane 2005), an approach that engages students in actively seeking solutions to problems by using critical thinking skills. PBL is a departure from traditional learning in that students become responsible for finding creative responses to problems as members of teams and are partners in the assessment of the group's work. Critically, PBL also entails student engagement in their own formative and summative assessment, with both their own and their peers' evaluations contributing to their actual scores for individual modules or courses. Research that examined the benefits of the PBL approach in supporting critical thinking skills and absorbing the key learning objectives for a given interpreting course would be fruitful.

6. Focuses on communication, a key theme for SLIs is that graduates can communicate information, ideas, problems and solutions to both specialist

and non-specialist audiences. This is an essential skill, with researchers arguing that the main prerequisite of translation activity is linguistic and cultural competence for both source and target texts (Nord 1992). In training, we assess this over time by testing skill in multiple domains, looking at register flexibility, and through the assessment of written materials. This descriptor is heavily emphasized, becoming *the* skill-set on which SLI assessment traditionally concentrates. Despite this, very little empirically-driven data has been generated that examines development of skill in SLI students. In turn, there seem to be very few research-driven programs in place. This lace of evidence-based programs is a crucial area requiring attention. It is one which the cognate field of language teaching has engaged with, for example, via the Common European Framework of Reference for Languages (CEFR) and the associated European Language Portfolio (ELP) (Council of Europe 2001)[2].

7. Requires graduates to be able to develop learning skills that prime them for further study and for practical work within defined responsibilities. Making progress in the profession entails the need to provide postgraduate options for interpreters (e.g., additional languages or in specialist domains as per the EU-MASLI programme offered collectively by Herriot Watt University (Scotland), the Humak Institute (Finland), and the University of Magdeburg (Germany)). These types of graduate options will not only ensure more reflective, highly trained practitioners, but will also encourage the development of what Daniel Gile calls "practicesearchers," professional interpreters engaged in research on interpreting. This, in turn, will lead to the potential for greater, and more critically evaluated understanding of our work, which can then feed back into training.

The question remains as to how we assess these skills at completion of undergraduate training bearing in mind that in many countries, training is not available at university level at all. In part, these are the traditional skills that link to academic autonomy, which we also seek to assess in internship or placement programmes. Aside from the important issue of promoting the academic as well as practical aspects of interpreter education, we should ask if the Dublin Descriptors are adequate to the task we set for ourselves as teachers and evaluators of SLIs. Questions include:

---

2. See Leeson and Byrne-Dunne (2009) and Leeson and Grehan (2009) for discussion of how the CEFR has been adapted for the teaching and assessment of Irish Sign Language and a range of other signed languages included in D-Signs Project. D-Signs is a Leonardo da Vinci project funded by the European Commission. The project is led by the University of Bristol's Centre for Deaf Studies. Partners include the British Deaf Association (Wales), the Centre for Deaf Studies, Trinity College Dublin (Ireland), Systema (Greece), the University of Nicosia (Cyprus) and Charles University (Czech Republic). See http://www.dsigns-online.eu/ for further details.

As interpreter educators, are we clear about the criterion that forms our proficiency test bases? Do we clarify our criteria to students for transparency purposes? Do we clarify our criteria for examiners to facilitate norm-referencing across raters and test score validity testing? We should bear in mind that interpreting research suggests that a "clear explication of the assessment criteria used in exams enhances the learner's autonomy and may exert a considerable influence on the quality of students' work" (Bartkomiejczk 2007: 251).

We can also ask if we are efficiently linking core competencies for interpreter performance to syllabus design and assessment criteria. That is, are we actively checking that test content assesses what it is we say we want students to be able to do on completion of their program? This entails that our programmatic learning outcomes are aligned to the competencies that stakeholders and professionals associate with fitness to practice, and that these competencies are embedded in the curriculum and explicated in module learning outcomes and assessed in a range of ways.

While very few studies examine the process of student testing within interpreter programs, a small number of case studies are available, which are outlined in the next section.

Student self-assessment

One of the few studies that investigates student interpreters' self-evaluations is Bartlomiejcyk (2007). She takes as a starting point the quality assessment of professional and student interpreters in a range of contexts, focusing on two specific studies of student interpreters of Polish-English. She reports on a study of 18 students who were asked to evaluate their performances, focusing on strategic processing they applied while on task. Her informants were completing their second or third year of simultaneous interpreting practice and her study focused on their analysis of a ten-minute speech from English to Polish (B-A) by a Dutch Prime Minister.

Bartlomiejcyk reports a significant trend towards negative self-assessment combined with attention given to how faithful the target language (TL) is to the source and to issues of completeness of the TL message. In contrast, Bartlomiejcyk notes that students rarely focused on issues of presentation such as monotonous presentation, hesitant voice and long pauses.

Bartlomiejcyk suggests that the focus on negative self-assessment may owe much to the fact that in class, teachers (who have limited time at their disposal) typically present feedback regarding the gravest errors in student performances. She also notes that neither problems with faithfulness to the original nor lack of

completeness of the TL message can be rectified simply by student self-awareness. She suggests that certain errors (such as errors of sense and errors of omission resulting from poor proficiency in the TL) can only be overcome by increased TL proficiency. In contrast, Bartlomiejcyk notes that if errors arise as a result of inadequate strategic processing or what she calls "imperfect allocation of processing capacity" (ibid.: 263), then focused training can assist in improving performance output[3]. She further suggests that product evaluation (i.e., a focus on the form, completeness and quality of the TL output) is a task that comes easier to student interpreters than retrospection on strategic processing.

While Bartlomiejcyk proposes that students at this level are focused on product rather than process, Leeson (2007) found that a sample of student Irish Sign Language (ISL)/English interpreters, at a much earlier stage of training, exhibited the capacity to reflect on both product and process, albeit at a fairly basic level. Considering meta-cognition as a crucial skill for interpreters, Kruger and Dunning (1999) note that "…students self-perceptions were not good predictors of their skill level: those who were unskilled tended to be unable to assess accurately their ability". They go on to say, "Not only do these people make erroneous conclusions and make unfortunate choices, but their incompetence robs them of the meta-cognitive ability to realize it" (p. 121).

In contrast, lack of meta-cognitive awareness leads to a lack of weighting of specific data, poor cohesion and increased difficulty in getting to the meaning (see Jones 1997; Marschark et al. 2005; Russell 2007). Given this, we suggest it is critical that SLIs have the metacognitive skills necessary to appraise their performance and skill levels in a way that reflects actual – rather than imagined – skill level. The objective is to integrate self-analysis skills with knowledge of how to develop practical methods for improving areas of weakness (via guidance). Such critical introspection allows for reflection on a range of issues including:

1. Appreciation of the fact that meaning cannot be known simply by understanding all of the words or signs of a language. Instead, meaning is co-created in context and is, by nature, intersubjective in nature (Wilcox and Shaffer 2005). Pragmatics leads language use regardless of whether the linguistic event is monolingual or if it entails interpreted interaction (Janzen 2010).

---

3. Of course, we would not wish to suggest that a conduit-based understanding of language is one that should be embedded in teaching or assessing interpreters. For further discussion on this point, and for discussion of a cognitive model of interpreting, See Wilcox and Shaffer (2005).

2. Indicators of what might linguistically constitute "quality" signed language interpreting; for example, Cokely's (2004) miscue analysis and discussion of lag-time effect in simultaneous interpreting, Baker's (1992) translator linguistic management strategies, as well as work by interpreting studies researchers in Europe, the US, Canada, Australia and New Zealand, e.g. the work of Harrington and Turner (2001), Baker (1992), Stone (2007), Roy (2000), Metzger (1999), Mindess (1999), Wadensjo (1998), Gile (1995), Janzen (2005), and Wilcox and Shaffer (2005), among others.
3. Consideration of the quality of the cultural turn in interpreted events such as the modality specifics of signed-spoken language interpreted interactions as well as linguistically referenced cultural specific indicators (e.g., Brennan and Brown 1997). This is perhaps the most difficult for student interpreters to appreciate in a deep way. Native signer SLIs often struggle with contextualizing cultural data from ISL into English, while non-native signer SLIs often don't identify signals of cultural significance beyond the most general levels.
4. Attentional issues arising in student interpretation (e.g. Giles' Effort Model).
5. Ethical issues that arise, including co-interpreting issues or conflicts.

Some simple caveats must be noted. Despite Bartlomiejcyk's (2007) observation that student interpreters automatically assess their own work, students cannot make informed reflections without guidance. They cannot reflect on how interpreting theory and practice relates to them if they are not introduced to the literature. The "language" of self-reflection has to be taught and developed through, for example, Think Aloud Protocols, group work, and self-analysis. Finally, feedback from teachers and mentors in structured programs is essential.

This type of guidance might facilitate and impact on the quality of student performance in a number of ways. First, reflection on practice and implementing the learning from reflection is one of *the* defining pathways in the development of expertise. Practice in isolation is futile as mistakes, uncorrected, can become habitual, to a sense of "I'm alright" or conversely, "I'm no good – why bother", which links to Kruger and Dunning's (1999) comments on meta-cognition, as discussed earlier. Leeson (2007) suggests that such *guided* self-reflection provides student interpreters with the tools to continue to make progress on the path to expertise beyond the scope of their training.

Leeson also considered student reflections on how signers and speakers package information, on the form and function of the source language (SL), and on options that interpreters have for managing that in the target language (TL). She found that that even at a relatively early stage of formation, students reflect on both product and process. The kinds of language that students use to analyze their own performances include:

(1) In staying away from the form of the source language, when working into the TL, the interpreter was able to paraphrase and to extract the meaning of the speaker to gain equivalence in the TL. Seleskovitch (1978) suggests that there should not be word for word translation but rather, the interpreter needs to attain a search for equivalents in two different languages.

(2) Considering Gile's Effort Model, I must have been putting too much emphasis on listening and analysis and neglecting the production element in the process.

Example (2) demonstrates capacity to consider processing strategy. This student is reflecting on a segment where the TL is not cohesive and simply provides literal sign for word renderings of what was present in the SL. The student identifies where her particular problem arises, and, with coaching from the lecturer, seeks strategies to redress the balance in performance. As in (2), the student in (3) notes that problems arise because they do not adequately parse information in the TL appropriately.

(3) Jones (1997) talks about the importance of actively listening. I made a lot of errors in this piece but they were mainly down to the fact that I was not actively listening. Instead I was trying to get as much of the information across on a surface level. This meant that my choice of words and grammar were heavily influenced by my first language, which is English, rather than the language I was working into (ISL).

Strategic on-task thinking is also reported on. For example, one student discussed how she strategically omitted SL components in a bid to produce a cohesive TL. Here, we must take into account that in ISL, there are significant linguistic outcomes associated with gendered generational signing (see LeMaster 1990, 1999–2000; Leeson and Grehan 2004; Leeson 2005; Leonard 2005; Leeson et al. 2006; Leeson and Saeed 2011).

In (4) below, the student comments on her interpretation of two pieces: one by a woman in her sixties (Patricia) and another by a younger male signer in his late twenties (Sean). The woman made great use of sandwiching fingerspelled items and signs, when introducing nominals for the first time. This student wrote:

(4) As Patricia is a different age from Sean, she often signed a word and then fingerspelled the same word. I found it hard to know what Patricia was doing, however, I realised the fingerspelling was not entirely necessary as most of the words [spelled had a sign which had already been used…This shows the interpreter is required to have a wide knowledge of variation in ISL…

In stating the case for student self-evaluation, it is important to emphasize that there has to be a balance between space to reflect and space to "do" interpreting. There is clearly a need to balance out assessment to ensure that students are not rewarded for knowing where they are going wrong in situations where they need to *act* appropriately. At the same time, meaningful self-evaluation (i.e., where the student's assessment of their own work has a value associated with their result) has value when it is built into training and is seen as forming part of the pathway from unconscious incompetence to conscious competence and beyond.

Sadlier (2007, 2009) presents on student assessment, considering the views of both test-takers and evaluators. She discusses a case study that explores both test-taker and developer/rater perceptions as well as linguistic and sociolinguistic challenges that arose during the test with respect to theoretical interpreting frameworks. She also considered how testers can work within test design constraints to find an appropriate balance between authenticity and reliability.

Looking at a liaison interpreting test, Sadlier (2007, 2009) notes that students are not expected to have anything approaching the same level of competency as a working interpreter at this stage of development and thus, this test is viewed as an achievement (rather than a proficiency) test. This test is designed to correspond to the learning outcomes in its construct, purpose and content. Test construction is based on the module's learning outcomes, which outline the aspects of knowledge to be measured (Bachman 1990). The test purpose is an achievement test (McNamara 2000). The test assesses the basic skills expected of trainee interpreters at this stage and is not a measure of:

> … end product' proficiency … and … content relates to a "real world" scenario which correlates both to the real world of interpreting and to the practice role-plays that occurred during the term, thus aiding validity and reliability (Sawyer 2004).
> (Sadlier 2009: 186)

Sadlier's case study entails three distinct elements: a focus group, a Think Aloud Protocol (TAP) exercise, and an analysis of exam performance and her findings focused on three main themes: (1) general issues, (2) linguistic issues, and (3) sociolinguistic issues. She found that students had concerns around learning Irish Sign Language (ISL) vis-à-vis previous language learning experiences, and concerns about the test setting itself. Given the importance of language proficiency for student interpreters, and the central role assessment of language proficiency plays within interpreter training programmes, it is interesting that all participants reported a prior negative (spoken) language learning experience which contrasted with their signed language learning experiences. Sadlier points out that such negative views can impact on test-takers' examination performances and the impact of

previous tests can factor as determiners of success and failures in the minds of the test-takers (Shohamy 1982; Ushioda 1996).

In discussing test setting, all participants commented on interpreting process factors, with a particular focus on memory. Students tended to seek to place culpability for poor interpreting process on extrinsic factors. For example, they attributed their difficulties with memory within the test scenario to external factors such as inaccurate or lengthy source language messages even though they had noted capacity management as a key difficulty when acquiring interpreting skills. Sadlier suggests that test-takers are only too aware of the impact that these tests will have on their self-esteem and indeed their future as students, and potential professionals, noting the importance associated with face validity (i.e., ensuring that test takers see a test as valid and authentic) and that they believe in the testing process.

Other issues that challenged students were linguistic in nature. Specifically, the bimodal nature of the communication was challenging, and when using ISL (none of the test takers were native ISL users), they had difficulty in maximizing use of Non Manual Features (NMFs) in their ISL TL product for morphological, syntactic, and affective purposes. They also unintentionally added meaning non-manually – students displayed a frown-like expression, which is problematic as furrowed brows can bear grammatical meaning in ISL, signifying a question rather than a statement or bearing a negative attitude towards the subject under discussion and thus altering the intent of the message.

Another key challenge Sadlier pinpointed for this cohort was the management of numeric and phonetic information that required literal transfer to the TL via the use of number signs or fingerspelling. During the TAP, Sadlier's informants noted that while fingerspelling was not difficult *per se*, while on-task in an interpreting situation, it causes extra pressure due to the requirement to manage all other information that co-occurs with it. Students also experienced difficulty with the use of loci for establishing and tracking reference in the ISL data, as well as sociolinguistic factors such as the management of turn-taking, and ratification of both deaf and hearing participants in the interpreted exchange (following Metzger 1999).

Following from this process, Sadlier built on test-taker feedback when preparing the following year's cohort's test procedures. For example, the test marking scheme was made available to students who were encouraged to apply the system when analysing their own in-class performances. This process assists in developing a clear understanding of the assessment criteria applied and provides clear-cut criteria for self-evaluation independent of testing. This, along with ensuring that students know about the layout of the assessment space leads to better conscious awareness of the test protocols, and such processes are focused on improving face

validity, alleviating the "mystery" of the test (Ushioda 1996; Shohamy 2001), and providing test-takers with greater transparency in assessment. Sadlier's participants reported that this helped reduce assessment process related fears.

Looking forward: Toward better ways of testing SLIs?

Angelelli (2007) and Sadlier (2009) provide pragmatic, focused approaches for dealing with test taking. While these provide an excellent starting point for considering the detail of how we might modify the practice of testing for specific purposes (Angelelli 2007) or for skill development in specific interpreting modes (Sadlier 2009), there are also larger-scale issues with which we must contend.

Most notably, generic tests are frequently used as the basis for assessment of specialist skill, for example, in registration tests with certain thresholds used to mark capacity for domains where no specialist test has been created (e.g., Signature – formerly CACDP, in the UK). Tests that are effectively demonstrating achievement of aspects of the curriculum are forced to function (or, perhaps more accurately, are perceived as functioning) as proficiency tests with respect to what employers want graduates to be able to do in the real world. For example, the consecutive interpreting test that Sadlier (2009) discusses is an achievement test and is not designed to test interpreter proficiency, although some of the simultaneous interpreting tests offered later in the students' program function as both proficiency and achievement tests. Her results suggest a need to increase clarity regarding what purpose a tests serves, whom the test is for and why we test this item in this manner at this stage of student development.

Striking a balance

One way of moving forward is by accepting that assessment is sometimes about ticking the box on what students have achieved relative to the curriculum that we cover. We tend to introduce students to elements of performance that are gradient in terms of difficulty. For example, SLI programs tend to introduce students to shadowing exercises, then paraphrasing exercises, then consecutive interpreting tasks before moving towards simultaneous interpretation in unilateral settings and then move toward simultaneous interpreting in bilateral and interactive environments. We assess student progress at salient points of development, typically marked by examination periods in institutional calendars. Assessment then, is not typically individuated, despite moves towards "learner centred teaching".

However, in tandem with the achievement testing component of our work, we are also asked to stand over students' fitness to practice on successful completion of our programs. The difficulty lies in identifying the minimum levels of competence required to be a SLI. We may need to consider adopting a set of agreed professional competencies[4], starting with language proficiency.

An increasing number of European universities are working with the Common European Framework of Reference for Languages (CEFR), the Council of Europe's framework for learning, teaching and assessment of languages (Council of Europe 2001). This allows for clearly defined descriptors for working languages, aligned to the CEFR, and mapped to curricula (Leeson and Byrne-Dunne 2009; Leeson and Grehan 2009). From this, there is scope for assessments linked to the CEFR, allowing for cross-linguistic mapping of graduate language competencies. A significant advantage of this system is that students are full participants in tracking their learning, and they engage in regular self-assessment exercises and, in a sense, validate their self-assessments through feedback from lecturers.

Regarding interpreting, we know that assessment of competence is not *just* about language proficiency. As Angelelli (2007) notes, language proficiency *and* a minimal (defined) threshold of interpreting competence must be attained. But this is not enough either. SLIs are required to be "people people", and because of this, we are also looking to measure competency in things that we don't actually teach students to do, such as empathy or the ability to remove one's own ego for the purpose of ensuring that communication amongst participants go smoothly.

Much of this is the "touchy-feely" element of subjective assessment, as pointed out in the earlier discussion on rater influence on test outcomes, maps onto the "soft-skills" that interpreters are expected to have. Importantly, these link to what deaf people have called "attitude" – namely, that preferred interpreters are those with a good attitude toward the deaf experience, which has traditionally been given precedence over language skill. In contrast, there is a falling away from the deaf community in some countries, in part as a result of the lack of linkage between training and practice. Cokely (2005) notes that while the deaf community was historically responsible for selecting who would serve their community as an interpreter, this is no longer the case, as deaf communities in the USA are not widely involved in the selection of candidates for interpreter training and assessment. This is an issue in parts of Europe too, often made more complicated because of the low numbers of deaf professionals working at tertiary education

4. Such a discussion would book-end discussion on the aptitudes that potential interpreters should have at entry to interpreting programmes, for example, as per the 2009, "Aptitude for Interpreting. Towards Reliable Admission Testing" conference at Lessius University College, Antwerp, Belgium.

level where interpreters are trained. No data is available with respect to the European context, but informal discussion with deaf and hearing colleagues from various parts of Europe suggest that there are barriers to deaf people's participation in interpreter training (e.g., because higher institutes in some countries demand masters or PhD qualifications that many skilled teachers do not yet have for many reasons). Further, we know that there are problems with interpreter quality in practice (e.g., Brennan, Brown and MacKay 1997; Brennan 1999). The impact of this divide on the selection, training, and perceptions of signed language interpreters in deaf communities is yet another area ripe for research.

Conclusion

Assessment is a process, which entails a shuttling between review of test design, criteria, resources, washback issues, rater training, validation and review, potentially leading to revision of what we are doing. SLI training does not exist in a vacuum. High stakes testing, external to our programmes (i.e., registration testing) also impact on our teaching and assessment protocols in some countries.

Proficiency testing must be explored, debated, and analysed more fully by programme deliverers and by pan-continental organisations of interpreter trainers. We must consider a range of questions – both in-house, and via empirical research processes – including: (1) Are we adequately separating out language proficiency from interpreting proficiency?; (2) Are we are clear about the criterion we are using in preparing proficiency tests? Are they appropriate?; (3) Are we overgeneralising inferences drawn from certain test domains and applying them beyond the scope of their applicability? (e.g. general tests used to mark proficiency, but not specified competence, to work in medical/legal domains); (4) Are we adequately training assessors and particularly, those working as "naïve" raters in the field?; (5) Do we share a common view on what interpreters should be able to do in order to be "fit for practice" in a specific geographical territory (e.g. the European Union)? If not, why not?; (6) What are the core competencies that we wish to see graduates achieve?; (7) Are we mapping these onto our training programme in an adequate manner? Are we effective in doing what we set out to do? Are these adequately benchmarked vis-à-vis the Dublin descriptors that underpin the Bologna Process for European training institutions or equivalent pan-continental educational markers in other instances?; (8) Are we building student self-assessment into the process or excluding it? Why?; and (9) Are we empirically cataloguing what we do when we test interpreting students and professional interpreters? Are we learning from our experiences of testing or embarking on assessment predicated by flawed assumptions of what testing entails?

As we have seen, assessment is a complex issue and we must be sufficiently sophisticated in our response to the challenge of ensuring fair, appropriate and authentic tests that we can stand over, which external parties see as valid and reliable, and which serve to appropriately reflect fitness to practice requirements. We should be mindful of how transparency with respect to assessment protocols can aid student progress, and we should empirically look at the value of student self-evaluation as a tool in developing critical thinking and autonomous learning for student interpreters.

Acknowledgements

Thanks are due to the staff and external examiners at the School of Linguistic, Speech and Communication Sciences, and particularly those at the Centre for Deaf Studies, for their willingness to discuss issues of assessment. A special thank you to Professor David Little for his mentoring around the development of testing and for feedback on earlier drafts.

References

Angelelli, Claudia. 2007. "Assessing medical interpreters: The language and interpreting testing project." *The Translator* 13 (1): 63–82.

Bachman, Lyle F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bartlomiejcyk, Magdalena. 2007. "Interpreting quality as perceived by trainee interpreters self-evaluation." *The Interpreter and Translator Trainer* 1 (2): 247–268.

Brennan, Mary. 1999. "Signs of injustice." *The Translator* 5 (2): 221–246.

Brennan, Mary, Brown, Richard K. and MacKay, Brenda. 1997. *Equality Before the Law: Deaf People's Access to Justice*. Durham: Deaf Studies Research Unit.

Cokely, Dennis. 2005. "Shifting positionality: A critical examination of the turning point in the relationship of interpreters and the deaf community." In *Sign Language Interpreting and Interpreter Education: Directions for Research and Practice*, Mark Marschark, Rico Peterson and Elizabeth A. Winston (eds), 3–28. New York: Oxford University Press.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Humphrey, Janice H. and Alcorn, Bob J. 2001. *So You Want to be an Interpreter? An Introduction to Sign Language Interpreting (4th ed.).* Amarillo, TX: H and H Publishers.

Janzen, Terence D. 2010. Pragmatics as start point; Discourse as end point. Presentation given at High Desert Linguistics Conference. University of New Mexico, Albuquerque.

Janzen Terry (ed.). 2005. *Topics in Signed Language Interpreting*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Johnson, Kirstin. 1991. "Miscommunication in interpreted classroom interaction." *Sign Language Studies* 70: 1–34.

Jones, Roderick. 1997. *Conference Interpreting Explained*. Manchester: Saint Jerome Press.

Labov, William. 1969. *The Logic of Nonstandard English*, *Georgetown University 20th Annual Round Table, Monograph Series on Languages and Linguistics*. Washington DC: Georgetown University.

Leeson, Lorraine. 2005. "Vying with variation: Interpreting language contact, gender variation and generational difference." In *Topics in Signed Language Interpreting*, Terry Janzen (ed), 51–68. Amsterdam and Philadelphia: John Benjamins Publishing Company.

———. 2007. 'I know where I'm going wrong': Self-analysis as the first step in improving quality interpreting. Paper presented at the Critical Link 5 Conference, Sydney, Australia.

Leeson, Lorraine and Byrne-Dunne, Deirdre. 2009. "Applying the common European reference framework to the teaching, learning and assessment of signed languages. In *D-Signs: Distance Online Training in Sign Language (UK/08/LLP-LdV/TOI/163_141)*. Dublin: Trinity College Dublin.

Leeson, Lorraine and Grehan, Carmel. 2004. "To the lexicon and beyond: The effect of gender on variation in Irish Sign Language." In *To the Lexicon and Beyond: Sociolinguistics in European Deaf Communities*, Mieke van Herreweghe and Myriam Vermeerbergen (eds), 39–73. Washington DC: Gallaudet University Press.

———. 2009. A common European framework for sign language curricula? D-sign(ing) a curriculum aligned to the common European framework of reference. In *First Applied Sign Linguistics Conference*. Bristol, UK.

Leeson, Lorraine and Saeed, John. 2011. *Irish Sign Language: A Cognitive Linguistic Account*. Edinburgh: Edinburgh University Press.

Leeson, Lorraine, Saeed, John, Leonard, Cormac, MacDuff, Alison and Byrne-Dunne, Deirdre. 2006. "Moving heads and moving hands: Developing a digital corpus of Irish Sign Language. The 'Signs of Ireland' Corpus Development Project." In *IT&T Annual Conference Proceedings*: 33–43.

LeMaster, Barbara. 1990. The maintenance and loss of female and male signs in the Dublin deaf community. Ann Arbor: U.M.I. University of California.

———. 1999–2000. "Reappropriation of gendered Irish Sign Language in one family." *Visual Anthropology Review* 15 (2): 1–15.

Leonard, Cormac. 2005. "Signs of diversity: Use and recognition of gendered signs among young Irish Deaf people. *Deaf Worlds* 21 (2): 62–77.

McNamara, Tim. 2000. *Language testing*. Oxford: Oxford University Press.

Mindess, Anna, Holcomb, Thomas K. and Langholtz, Daniel. 1999. *Reading Between the signs: Intercultural Communication for Sign Language Interpreters*. Boston: Intercultural Press.

Napier, Jemina. 2002. *Sign Language Interpreting: Linguistic Coping Strategies*. Coleford, England: Douglas McLean.

———. (ed.). 2009. *Signed Language Interpreter Education and Training: A World Survey*. Washington DC: Gallaudet University Press.

Napier, Jemina, McKee, Rachel and Goswell, Della. 2006. *Sign Language Interpreting: Theory and Practice in Australia and New Zealand*. Sydney: Federation Press.

Nord, Christiane. 1992. "Text analysis in translator training." In *Teaching Translation and Interpreting: Training, Talent and Experience*, Cay Dollerup and Anne Loddegaard (eds), 39–48. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Roy, Cynthia B. (ed.). 2000. *Innovative Practices for Teaching Sign Language Interpreters*. Washington, DC: Gallaudet University Press.

Sadlier, Laura. 2007. Assessment of sign language interpreters in training: A case study. Report for the Centre for Language and Communication Studies, School of Linguistic, Speech and Communication Sciences, Trinity College Dublin, Dublin.

———. 2009. "Pandora's box: Lifting the lid on issues of testing – A case study of sign language interpreters in training in Ireland." *The Sign Language Translator and Interpreter* 3 (2): 177–202.

Sawyer, David. 2004. *Fundamental Aspects of Interpreter Education: Curriculum and Assessment*. Amsterdam: John Benjamins Publishing Company.

Seleskovitch, Danica. 1978. *Interpreting for International Conferences*. Washington DC: Pen and Booth.

Shohamy, Elana. 1982. *Affective Considerations in Language Testing*. London: Blackwell.

———. 2001. *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. London: Pearson.

Sloane, Pauline. 2005. *A Guide for PBL*. Dublin: Trinity College Dublin.

Stewart, David A., Schein, Jerome D. and Cartwright, Brenda E. 1998. *Sign Language Interpreting: Exploring Its Art and Science (2nd Edition)*. Needham Heights, MA: Allyn and Bacon.

Stone, Christopher. 2009. *Toward a Deaf Translation Norm*. Washington, DC: Gallaudet University Press.

Ushioda, Ema. 1996. *The Role of Motivation*. Dublin: Authentik Language Learning Resources.

Van Herreweghe, Mieke. 2002. "Turn-taking mechanisms and active participation in meetings with deaf and hearing participants in Flanders." In *Turn-Taking, Fingerspelling, and Contact in Signed Languages*, Ceil Lucas (ed.), 73–106. Washington DC: Gallaudet University Press.

Wilcox, Sherman and Shaffer, Barbara. 2005. "Towards a cognitive model of interpreting." In *Topics in Signed Language Interpreting*, Terry Janzen (ed.), 27–50. Amsterdam: John Benjamins Publishing Company.

Websites

www.flexiblelearning.net.au/ (accessed 24 August 2008).

www.tki.org.nz/r/assessment/overview_e.php (accessed 24 August 2008).

www.mqc.gov.mt/pdfs/Grid%20of%20statements.pdf (accessed 25 August 2008).