

## Quality-assessment expectations and quality-assessment reality in educational interpreting: An exploratory case study

---

Lenelle Foster

Language Centre, Stellenbosch University, South Africa  
Email: lfoster@sun.ac.za

### Abstract

This article focuses on data obtained from three separate studies conducted during a four-year period at Stellenbosch University, a higher education institution in South Africa. All three studies centred on the simultaneous interpretation of undergraduate lectures. Various data sets were used to examine whether there would be a discrepancy between what lecturers in a particular academic department emphasised when they first considered the feasibility of this type of educational interpreting, and what they actually focused on when assessing the interpreters' performance. Discrepancies and correlations in the quality criteria identified by lecturers were examined against a rubric taken from existing literature on interpreter assessment (notably that of Kurz (2002)). Using this information and augmenting it with comments from a similar assessment of the same material undertaken by experienced interpreters, these discrepancies and correlations are briefly discussed. Given the exploratory nature of this case study, few recommendations are made. However, the fact that the data from this study seem – in broad terms – to agree with studies conducted in the field of conference interpreting would seem to indicate that the discrepancy between stated and actual quality assessment criteria is real, and will require much more detailed study in an educational interpreting setting.

**Keywords:** educational interpreting, assessment, quality

### 1. Introduction

Interpreting quality is, as even the briefest glance at the literature will confirm, a complex subject. This article does not deal with quality directly.<sup>1</sup> It does not provide answers to questions such as “What is quality in interpreting?” or “How should one measure quality in interpreting?”. Rather, it takes its cue from work done by Moser-Mercer (2008) – amongst others – which would seem to indicate that different role-players in any particular interpreting scenario may have different definitions of ‘quality’, depending on their role in the process (Moser-Mercer 2008:147). Using this assertion as a springboard, this article seeks to examine the possible dissonance between stated and actual quality criteria among a group of lecturers in a particular

---

<sup>1</sup> For a discussion of quality in educational interpreting, particularly during Study B (section 2.1), see a forthcoming publication by this author.

department. In doing so, this article will draw on work by scholars such as Kurz (2002), Collados Aís (2002) and Bartłomiejczyk (2007).

When initially conceived, the objectives of the exploratory study discussed in this article were: (i) to determine how lecturers in a particular discipline – who professed a preference for subject-specialist interpreters – assessed non-professional interpreters from that discipline, and (ii) to compare that assessment with one done by experienced interpreters. Although Blaauw (2008:305) has expressed concerns about the suitability of subject-specialists as educational interpreters, and Verhoef (2008) has written on the subject of subject-specialist interpreters in an educational interpreting context,<sup>2</sup> anecdotal evidence at Stellenbosch University (SU), where the current study was undertaken, seemed to suggest that lecturers still held the view that such interpreters would almost always be better at educational interpreting than so-called “language-specialist” interpreters.

Once the data had been analysed, however, it appeared that, to a large extent, there was a fair degree of agreement between the two sets of assessments. The focus of the study shifted to examining the dissonance between what the lecturers claimed to value (correct use of terminology), and what they focused on in their assessment. With this approach, the assessments provided by the experienced interpreters were no longer foregrounded as strongly as initially anticipated.

Comments by the lecturers were organised according to a set of criteria used by Kurz (2002)<sup>3</sup> and are discussed in some detail in this article. Although these criteria were originally applied to conference interpreting rather than educational interpreting, they deal with general elements related to interpreting (and interpreting quality), and can therefore be used as a way to examine attitudes to performance.

It should be noted that, although it is possible to draw a few tentative conclusions from this exploratory study, the study does have serious limitations. Firstly, only four experienced interpreters and four lecturers assessed the performance of the subject-specialist interpreters. Secondly, only two recordings (of two different interpreters) were assessed by each group. The conclusions presented in this article should therefore be considered against the backdrop of this extremely limited scope.<sup>4</sup>

---

<sup>2</sup> Although not working with educational interpreting, Garwood (2002:268) points out that interpreters are seldom subject-specialists. They may specialise in interpreting a particular field, but that is not the same as a subject-specialist.

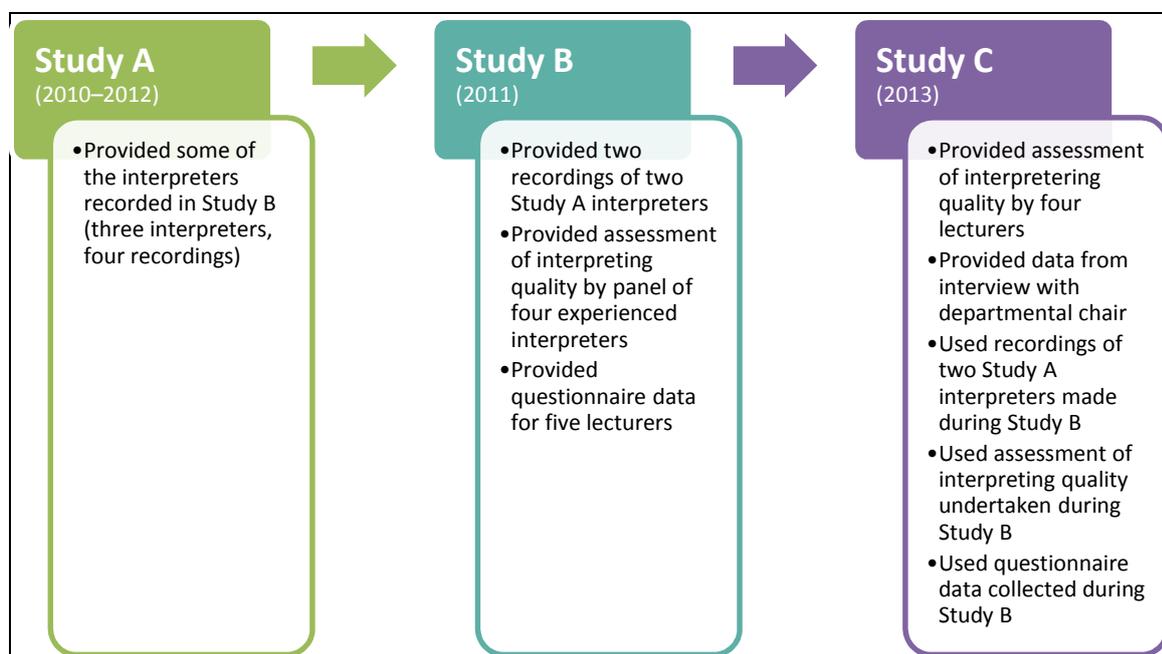
<sup>3</sup> Kurz herself adopted these criteria from a study performed by Bühler (published in 1986), using “the first eight of Bühler’s [15] quality criteria” (Kurz 2002:316).

<sup>4</sup> The data are published at this relatively early stage as SU officially established educational interpreting as an instructional aid in 2013, and has plans to increase the number of classes in which an interpreting service will be available. However, the possible discrepancy between quality-assessment expectations and quality-assessment reality found with one particular group of lecturers identified in this study should raise a red flag to policy makers, indicating that educational interpreting is not merely about placing interpreters in classrooms, but that managing and marrying expectations – whether presumed or actual – could have an important influence on the perceived success of such a service. (Obviously this study would have to be repeated on a much larger scale to determine whether there is indeed a red flag, or whether this exploratory study has simply discovered a red herring.)

## 2. Description of contributing studies

This article uses data collected at three different times and associated with three different studies that were all undertaken on the Stellenbosch campus of SU. To understand the sources of these data and the methodology employed in the study on which this article is based, it is necessary to deviate slightly and discuss the various studies that provided the data.

The first two rounds of data collection occurred during the course of one pilot study (Study B).<sup>5</sup> The third round occurred almost two years later (for the sake of convenience it will be called “Study C”); the data collected during this round form the backbone of this article. As there is some overlap between studies A, B and C, they will be described briefly to provide some contextual background. The information or resources gleaned from studies A and B, and their use in Study C (in particular), are summarised in Figure 1 below.



**Figure 1.** Schematic representation of data sources from three studies

### 2.1 Study A: Faculty of Science

This study was initiated in 2010 as an ad hoc project by the then-Dean of the Faculty of Science at SU to determine the feasibility of a simultaneous interpreting service in the Faculty (Swart 2012a, 2012b). The Department of Biochemistry was identified as the site for this study, mainly because of the Department’s willingness to participate.

The second-year biochemistry course is divided into two semester-long modules. The two constituent modules are compulsory for a number of degree programmes presented by the faculties of Science and AgriSciences. Given the number of students registered for the course each year, the class is split into two groups, with duplicate lectures being presented in Afrikaans

<sup>5</sup> Studies A and B (both pilot studies initiated as a result of managerial decisions) overlapped to some extent, as will be made clear in the description provided further on.

and English – despite the fact that the *SU Calendar (Part 1)* lists the language of teaching and learning as Afrikaans (Stellenbosch University 2013:63). The *de facto* parallel instruction presented an opportunity to gauge interest among students regarding the use of an interpreting service – due to timetable clashes, students who were not proficient in Afrikaans may well have been placed in the Afrikaans-medium group.

The lecturers in the Department of Biochemistry were adamant that they did not want anyone who was not a subject-specialist interpreting their lectures (Swart 2013). They identified an individual who was knowledgeable and had a facility for languages – one of their doctoral students who had worked at MFM (the campus radio station) and therefore had broadcasting experience. Her broadcasting experience, combined with the fact that they knew her personally, proved to be deciding factors for the lecturers. The departmental chair stated that if she had not been available to interpret, the Department would not have participated in the study (Swart 2013). In fact, he was quite candid about the fact that this “broadcasting person”, as he described her, was of a quite different calibre compared to some of the postgraduate students used later on in the pilot study (Swart 2013).

Study A ran from the first half of 2010 to the end of 2012 (Swart 2013). In total, five interpreters were used, all of them postgraduate students in the Department (Swart 2012b). None of these interpreters received any interpreting training and no quality assessment was undertaken by either the Department or the Faculty.

As far as can be established, no reports on this pilot study have been published to date and, besides the performance assessment undertaken as part of Study B, no quality control was carried out to determine whether the interpreting was of a suitable standard. Feasibility appears to have been notionally based on usage but, as no records were kept of the actual number of users, usage (and therefore, presumably, feasibility) remains a matter of conjecture.

## 2.2 Study B: SU Language Centre<sup>6</sup>

The SU Language Centre undertook its own official pilot study in 2011, having received instructions from the University Council to this effect. The live phase, in which interpreting services were provided in the faculties of Engineering and Education, occurred in the second semester and involved four interpreters and one interpreter-in-training.

The overlap between studies A and B involves quality assessment. As part of Study B, two lectures per interpreted subject were recorded and sent to a panel of four experienced interpreters, some of whom were also involved in interpreter training. The panel was not provided with any information on the interpreters or on the study itself, other than that it was a pilot project and that quality assessment was considered an essential component. The assessment panel was not informed that they were assessing both professional and non-professional subject-specialist interpreters.

At the end of the semester, lecturers teaching in the various modules where an interpreting service had been provided – for both studies A and B – were requested to complete a

---

<sup>6</sup> The feasibility study undertaken by the SU Language Centre and its impact on interpreting at SU will be discussed in detail in a forthcoming article. This section serves as a basic introduction in order to contextualise the work described in the current article.

questionnaire compiled by the Language Centre and the SU Centre for Statistical Consultation. The primary purpose of this questionnaire was to establish the attitudes toward and perceptions regarding interpreting. As the lecturers had not listened to the interpreters, the questions covered two broad areas. The main focus was on the impact of interpreting in the classroom, which involved investigating whether or not it created a disturbance, required a change in lecturing style, and to what extent it aided or hindered teaching. A predominantly negative response here may have led the team compiling the report for Council to urge caution in the use of educational interpreting at SU. The second focus area was the lecturers' perceptions of the interpreting, which involved determining whether or not the interpreting was accurate, the lecturers' ratings of the interpreters' language proficiency, and to what extent the lecturer trusted the interpreter.<sup>7</sup> These matters were considered important as it would be difficult to implement an educational interpreting service at SU if it were not possible to count on some measure of support from lecturers. Determining lecturers' perceptions of the interpreters and the service would also be important in devising a strategy to allay possible fears related to such quality matters as accuracy and veracity. (More detail on the content of the various questions will be provided in section 4.)<sup>8</sup>

While the aim of this study was to determine the feasibility and desirability of an educational interpreting service at SU, it was conducted on such a small scale that it could be possible to question some of the recommendations contained in the report to Council. Secondly, the study was undertaken while two separate pilot studies,<sup>9</sup> investigating the feasibility of educational interpreting at SU, were occurring elsewhere on the Stellenbosch campus. One could therefore argue that the findings of the study were either irrelevant or the result of managerial window dressing as the educational interpreting horse had already bolted and was frolicking on the Stellenbosch campus.

### **2.3 Study C: The comparative project**

Study C provides the basis for the present article and was undertaken in 2013 at a time when educational interpreting had already been established as a language management policy at SU, and had changed in status from pilot study to established language service. As was summarised in Figure 1 and will be described in the methodology section below, this study used data obtained during Study B and two recordings made during Study A. To some extent, Study C is a reproduction of a portion of Study B – using different subjects – as it required lecturers in the Department of Biochemistry to assess recordings generated and assessed during Study B.

However, given the difference in the objective of Study C, the data were used differently than was the case in Study B. In Study B, the quality assessment data provided by the four panellists were used to indicate that (i) it was possible to interpret lectures simultaneously, and (ii) the

---

<sup>7</sup> Alexander, Edwards and Temple (2004), Hadziabdic et al. (2010), and Napier (2011) all note the value of trust when using an interpreter: individuals would be more likely to make use of an interpreting service if they trusted the interpreter. As the Department of Biochemistry, rather than the students, would foot the bill for an educational interpreting service, the attitudes of lecturers could be crucial for the implementation of such an educational interpreting service at SU.

<sup>8</sup> A similar questionnaire was sent electronically to all students registered for the interpreted modules, regardless of whether they used the interpreting service. Those results were not pertinent to the current study, and will therefore not be discussed in this article.

<sup>9</sup> The second pilot study, undertaken at the Faculty of Engineering during 2011 and 2012, will not be discussed in this article.

quality of the interpreted product was of such a standard that students using this type of service would not be unfairly disadvantaged. In Study C, the aim was to establish (on a very small scale) whether lecturers who apparently viewed terminological exactness as the most important measure of interpreting quality, actually used that measure when they assessed interpreting products. As secondary aims, the assessments done by the four lecturers who were willing to participate were compared to the assessments provided by the four panellists in 2011, and were checked against existing literature on user perceptions of interpreting quality.

### **3. Methodology**

#### **3.1 Data sources**

Four main sources of data were used in writing this article. Firstly, the chairperson of the Department of Biochemistry was approached for an interview, during which the Department's attitude to educational interpreting and its participation in Study A were discussed. The basic premise of Study C – the examination of subject-specialist interpreter performance by lecturers – was also explained. The chairperson agreed that the staff in his Department could be approached and asked whether they wished to participate in a project to assess the performance of subject-specialist interpreters.

Secondly, lecturers who had taught the second-year course at the Department in 2011 were approached and requested to rate two interpretation performances. The lectures had been recorded in 2011 and it was thought that lecturers who had themselves taught various portions of the course at the time would be able to detect nuances, differences or deviations in the interpreted content that may have eluded other listeners. As anecdotal evidence from the period 2010 to 2012 suggested that more students typically used Afrikaans-to-English interpreting (rather than English to Afrikaans), and since problematic interpreting would potentially have more serious consequences for students who were unable to follow the Afrikaans lecture at all, recordings of interpreting from Afrikaans to English were used for evaluation purposes.

The lecturers were requested to listen to two recordings (an mp3 file with the original and interpreted lectures playing simultaneously), using a marking grid to rate the performance of each of the two interpreters according to various categories. Although four recordings (of three subject-specialist interpreters) were available, and had previously been analysed by the panel of four interpreters (see the elaboration of the fourth data source below), the likelihood of the lecturers listening to and evaluating approximately 200 minutes of interpreting performance seemed rather slim. The options seemed to be to use only 100 minutes in the hope that more lecturers would participate in the study, or to use the full 200 minutes available and risk a lower level of participation from the lecturers. Either option would limit the scope of the study and potentially undermine the validity of the results.

Thirdly, the marks and comments gleaned from the process described above were compared with comments made by the same group of lecturers in 2011 when they completed a questionnaire compiled by the Language Centre.

Finally, the two sets of data obtained from the lecturers were compared to assessments of the subject-specialist interpreters' performance by a panel of four interpreters, provided in 2011 and using the exact same recordings and marking grid.

By comparing data from data collection phase 2 (assessment by lecturers) with data collected in step 3 (questionnaires completed by the same group of lecturers), it was possible to establish, on a very limited scale, whether the lecturers used the criterion for quality assessment that they claimed to value, namely terminological exactness.

The third data source was used as a cross-check to determine whether the lecturers had been unusually lenient (relatively speaking) when assessing the performance of the two subject-specialist interpreters. As already indicated, this comparison was done using quality indicators identified by Kurz (2002).

### **3.2 Marking grid**

The marking grid used as a tool for obtaining data in the second and fourth steps described above, was an adapted version of the grid used by the South African Translators' Institute (SATI) in their interpreter-accreditation exams.<sup>10</sup> The original SATI grid had four categories: accuracy and coherence of message; target language (TL) vocabulary and register; TL grammar, idiom and purity; and interpreting technique (SATI s.a.). Each of these categories provided a list of constituent components in brackets after the main category heading. Although the content was presented quite succinctly, the layout of the SATI grid made quick reference rather difficult. It was mainly with this benefit in mind that the grid was adapted by the Language Centre. It differed from the SATI original in that it divided performance into three broad categories: content, form and interpreting skills. Each of these categories had at least one major component, with examples (which could be considered subcomponents) provided as bullets below the major component.

As Figure 2 shows, TL vocabulary and register, and TL grammar, idiom and purity were grouped together in the adapted grid under "Form", with bullets providing the clarification that TL vocabulary and register should be evaluated with the applicability of the vocabulary, terminology and register in mind, and that TL grammar, idiom and purity should be assessed on the basis of correct use of concord, tense and syntax as well the use of prepositions being on par with that of a mother-tongue speaker. The content, therefore, differs very little from the original SATI grid. It is simply presented in a format that may be easier to use while listening to an interpreter, particularly for a listener who is not used to evaluating simultaneous interpreting.

---

<sup>10</sup> The SATI grid was supplied by Mr Johan Blaauw, current SATI vice-chairperson and head of language practice at North-West University's Language Affairs Directorate.

Interpreting: [Course name] (Afr>Eng)			
Interpreter 1			
Performance		Mark out of 10	Comments
<b>Content</b>	<b>Message accuracy and cohesion</b> <ul style="list-style-type: none"> <li>equivalent meaning conveyed fully</li> <li>names, dates, numbers, abbreviations, acronyms, etc. conveyed accurately</li> <li>ability to deverbilise the message (not interpret literally)</li> <li>cultural/subject knowledge</li> </ul>		
<b>Form</b>	<b>TL vocabulary and register</b> <ul style="list-style-type: none"> <li>applicability of vocabulary, terminology, register</li> </ul>		
	<b>TL grammar, idiom and purity</b> <ul style="list-style-type: none"> <li>correct use of concord, tense and syntax</li> <li>use of prepositions the same as a mother-tongue speaker</li> </ul>		
<b>Interpreting skills</b>	<b>Listening skills</b>		
	<b>Analysis</b>		
	<b>Concentration</b>		
	<b>Problem solving</b>		
	<b>Paraphrasing</b>		
	<b>Presentation</b> <ul style="list-style-type: none"> <li><b>fluency</b> (little or no hesitation or repetition; ability to vary <i>decalage</i> ["following distance"])</li> <li><b>voice quality</b>, e.g. voice and breath control</li> </ul> <b>Correct use of equipment</b> <b>Conduct</b> , e.g. professional interaction with colleagues, way with which documents and information are dealt		

**Figure 2.** The adapted marking grid

## 4. Results

### 4.1 Marks out of 10

Considering the emphasis the Department placed on subject knowledge as essential for an educational interpreter, it would be logical to suppose that terminological exactitude would be important when assessing interpreting performance. The position articulated by the Department is similar to attitudes noted by Kurz (2002), who examined attitudes expressed at three large international meetings.<sup>11</sup> On average, delegates ranked "use of correct terminology" as the third most important in a list of eight criteria by conference attendees – preceded only by "[s]ense consistency with original message" and "logical cohesion of utterance" (Kurz 2002:317).

When completing the questionnaires in 2011, the lecturers reacted because of previous interaction with the interpreters as students in the Department. When asked whether they trusted

<sup>11</sup> In her results, Kurz (2002:317) includes results reported by Bühler (1986) for the purposes of comparison .

that the lecture content had been interpreted accurately, four out of the five lecturers who completed the questionnaire responded “Yes, more or less” and one answered “Yes, definitely”. As can be seen in Table 1, the lecturers’ responses (when invited to elaborate on their answers) focused on the students’ subject knowledge rather than on their linguistic competence. However, knowledge of correct terminology, which a postgraduate student would undoubtedly have, may not translate into use of correct terminology while interpreting.

**Table 1.** Lecturer comments on the extent to which they trusted the correctness of the interpreting<sup>12</sup>

Lecturer	Comment
1	I have never heard the interpreting but I accept the knowledge of the interpreters.
2	The interpreting was done by postgraduate Biochemistry students. I trust their levels of knowledge and I know these particular students.
3	The interpreter is a Biochemistry postgrad student who knows the work well.
4	We use our best senior postgraduate students and we know that they are strong academically and well-prepared. If they were not people who know the subject, I would have had serious reservations.
5	Because I know the interpreter as a good student. Other than that, all I could do was to trust.

In addition, a listener – while valuing correct terminology highly – may actually use other quality criteria when determining what is “good” interpreting, and may be influenced by a range of factors when assessing quality. The listener’s home language (Pöschhacker 2004:155, Kalina 2005:38), gender (Kurz 2001:399 *passim*), and experience in using interpreting services (Pradas Marcías 2006:38) may all influence what he or she considers important when it comes to “good” interpreting.<sup>13</sup> The quality components of “sense consistency”, “logical cohesion” and “correct terminology” may not be as important as listeners believe them to be. Collados Aís (2002:335) has shown that an error-free but monotonous interpretation is considered less accurate than an error-ridden but lively interpretation. Illustrating this quite clearly, Wu (2010:322) cites an interpreting examiner who claims that if the listener does not understand the source text, good interpreting is probably identified based on delivery quality.

One of the revised aims of this study was thus to determine whether the assessment performed by the lecturers would support the claims made in the literature that listeners may believe they prefer “sense consistency”, “logical cohesion” and “correct terminology” but that they are actually swayed by elements of interpreting delivery, or whether the faith the lecturers expressed in the interpreters’ subject knowledge would emerge as a decisive component in the lecturers’ assessment.

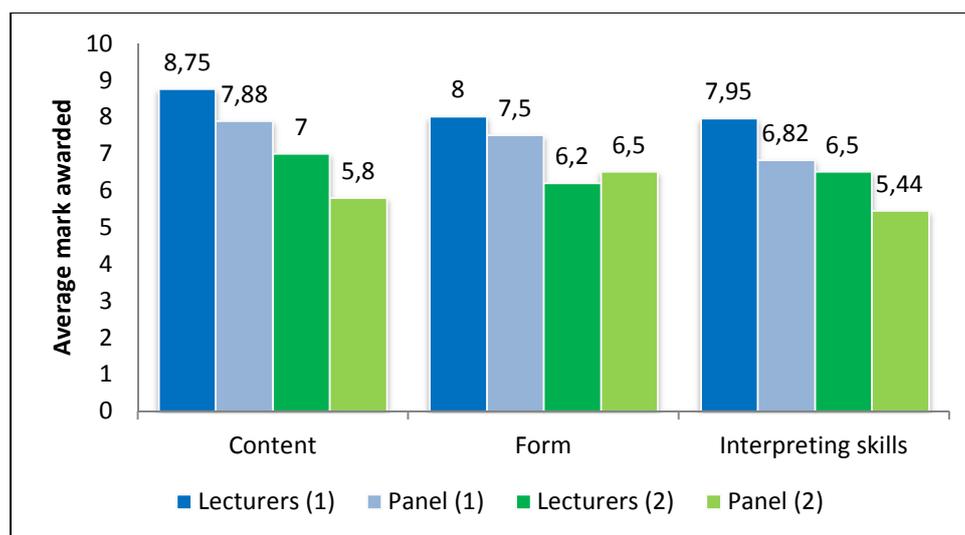
When looking at the scores provided by the lecturers, it would seem that the mean scores awarded by the lecturers and the panel, respectively, do not differ significantly. Figure 3 illustrates that, when placed side by side, the mean scores awarded by the lecturers and the panel for each of the two interpreters differ only by a single point (the largest difference is 1,2

<sup>12</sup> All comments originally provided in a language other than English have been translated into English by the author.

<sup>13</sup> Grbić (2008:248) notes that the audience using an interpreting service is frequently much less homogenous than is generally assumed, while Bartłomiejczyk (2007:251) points out that the assessment of interpreting output could also be more heterogeneous than may be initially supposed.

and relates to how well Interpreter 2 dealt with the lecture content). With one exception, the panel was slightly more critical in assessing performance than the lecturers, but as the discrepancy between the two assessing groups remains more or less constant, it could be discounted for the present.<sup>14</sup>

The reasonably small difference in assessment between the lecturers and the panel would also suggest that the lecturers were not as biased in favour of their students as may have been supposed, and that they were quite critical of their performance (as will be illustrated below).



**Figure 3.** The performance of the two interpreters as assessed by the lecturers (2013) and the assessment panel (2011)

What is clear from Figure 3 is that both groups believed that Interpreter 1 performed better than Interpreter 2. Interpreter 1 was the “broadcasting person” favoured by the Department. It would seem that their proviso that Study A could only continue if she were used as interpreter was justified. She performed noticeably better than Interpreter 2.

## 4.2 Comments

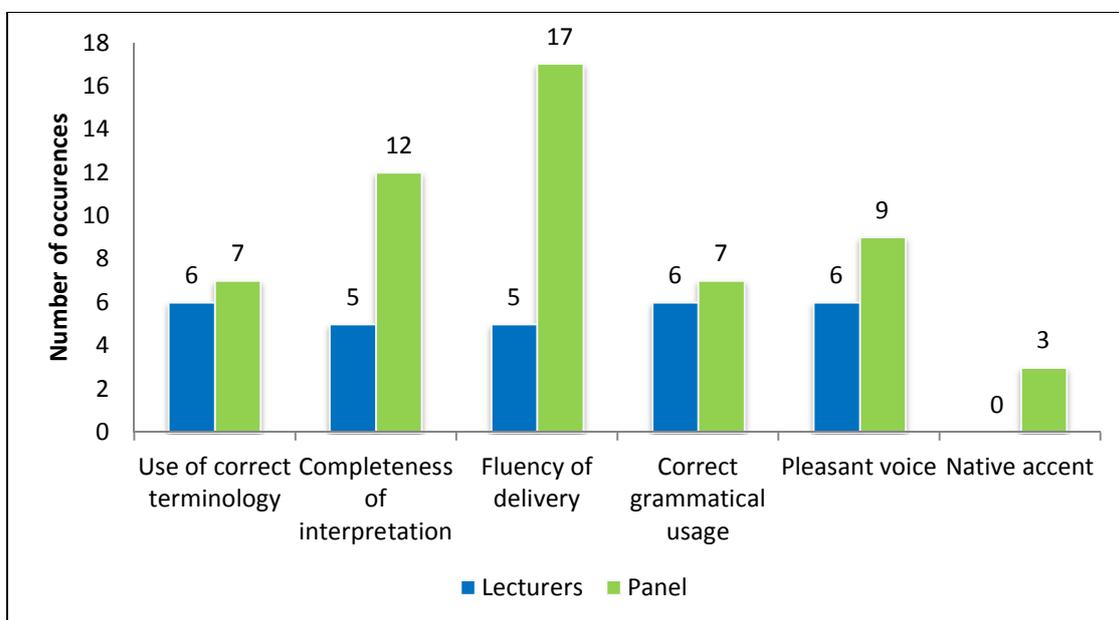
The comments supplied by the lecturers and the assessment panel offer the opportunity to examine their perception of what constitutes “good” interpreting (see Figure 4). Using Kurz’s criteria (2002:317) as a point of reference, the comments made by the lecturers and the panel can be analysed and compared. These criteria are (in descending order of importance, as determined by delegates at three international meetings): sense consistency with original message, logical cohesion of utterance, use of correct terminology, completeness of interpretation, fluency of delivery, correct grammatical usage, pleasant voice, and native accent (Kurz 2002:317). For the purposes of this discussion, the first two criteria will be considered basic expectations for competent interpreting.

<sup>14</sup> This would seem to substantiate findings by Kurz (2002:323) that interpreters (in this case the four panellists) tend to make higher demands on quality than delegates would. It is possible that the lecturers’ perspectives on interpreting would be closer to that of delegates at a conference – even when they actually play the role of quality assessors – than those of working interpreters (also playing the role of quality assessors).

Generally, the more critical approach of the panel is not only evident in the mark allocation, but also in the comments section. Comments by the panel are much longer and more detailed than those provided by the lecturers, and there is often criticism of elements related to interpreting not identified by the lecturers at all. There could be at least two factors contributing to this situation. Firstly, the lecturers had not used the assessment grid before and therefore formulated their comments in terms of the terminology and categories on the grid. Secondly, the panel, all being experienced interpreters, are used to listening to various aspects related to interpreting at the same time, while the lecturers focused on only some of these elements.

A further point to note is the differing importance of the various categories, when viewed by the commenting group. It was unexpected that the lecturers made more references to correct grammar and pleasant voice than to the completeness of the interpretation and delivery fluency, but as these only differ by one mention each, it should not be considered significant. The various aspects mentioned by the panel do not follow the order of priority detailed by Kurz (2002). Fluency of delivery received much more attention from this group than did the correct use of terminology or even the completeness of the interpretation.<sup>15</sup> (It is possible that an assertion by Pöchhacker (2001:419) – citing Shlesinger – that “normal” intonation aids memory while typical “interpreter” intonation does not, may have had some influence on the importance attached to fluency by the interpreting panel, particularly those involved in interpreter training.<sup>16</sup>)

In Figure 4, the number of references by the lecturers and the panel to various quality aspects is grouped according to six of Kurz’s (2002) eight criteria. The figure shows quite clearly which aspects the two groups focused on in their comments. The various criteria are also discussed in the following sections.



**Figure 4.** Focus of comments according to categories identified by Kurz (2002)

<sup>15</sup> Kurz (2002:317) also lists the relative importance of the various criteria as determined by interpreters. The interpreters involved in her study valued fluency of delivery above completeness of interpretation. The panel’s comments would therefore seem to be in line with this trend.

<sup>16</sup> See also Bartłomiejczyk’s (2007) take on Shlesinger’s (1994) study in section 4.2.3.

#### 4.2.1 Use of correct terminology

The lecturers referred to “terminology” six times, while the panel referred to it seven times. Four of the six references to terminology by the lecturers related to whether the terms were used correctly or not, while the two remaining references related to incorrect pronunciation. The panel referred to the correct use of terminology four times. Other references occurred where the relationship between terminological overload and interpreter fatigue was discussed.

#### 4.2.2 Completeness of interpretation

Words used to identify the completeness of the interpretation include “whole”, “completeness”, “gap”, and “omit”. The panel seems much more critical of a lack of completeness, particularly with Interpreter 2. One lecturer comments that “roughly 80% of the lecture is translated [sic]”. The panel, however, had the following comments: “Difficult to follow as much is omitted”, “Lots of information omitted”, “[O]nly the core ideas are conveyed” and “[A] whooping [sic] 34 seconds of silence means the English listeners would have lost crucial information”.

The tremendous difference in opinions may be related to the panel valuing “details or nuances” and an “exchange of [...] experience”<sup>17</sup> (Kurz 2002:321), rather than the transfer of a set of facts, while the lecturers may have been satisfied with “an intelligent, logical, terminologically correct summary of the original”, as Kurz (2002:321) describes in relation to the preferences of medical and engineering conference delegates.

#### 4.2.3 Fluency of delivery

As already indicated in section 4.2.2, the lecturers did not make much of Interpreter 2’s silences. In fact, they made only one reference to “silence”, with other key words including “flow”, “jerky”, and “hesitate”. Meanwhile, the panel focused on “silence”, “breathing” (which they considered far too loud for both interpreters), and used words such as “fluent”, “jerky”, “falter” and “hesitate” in their comments.

Bartłomiejczyk (2007:249) notes that “confident delivery” (which could be presumed to include fluency) “may be crucial for the perceived quality of any interpretation”. She bases this assumption on research done by Shlesinger (1994), which indicates that comprehension and recall are negatively affected by “pauses within grammatical constituents, stress incompatible with semantic contrast, [...] non-final voice pitch in final positions and non-standard alterations of speed” (Bartłomiejczyk 2007:249) – in short, by a lack of fluency in delivery. It may be that the panel, of which some of the members were involved in interpreter training, were exceptionally critical of hesitations precisely because they were aware of the impact it could have on listener perception.

#### 4.2.4 Correct grammatical usage

The correct use of grammar is the only point where there are striking similarities between the comments made by the lecturers and those made by the panel. The lecturers identified errors related to concord, tense, prepositions, and literal interpreting. The latter often leads to the

---

<sup>17</sup> One could argue that interpreting in a lecture setting should not only focus on conveying the facts contained in the lecture itself, but all the phatic nuances and affective components that accompany those facts.

reproduction of source-language syntax in the TL (resulting in unidiomatic language use), and could therefore be interpreted as syntactic errors. The panel identified concord, syntactic and prepositional errors, as well as errors related to *what/which* and *a/an*. It may be that grammatical errors are easy to identify with some degree of confidence, even for someone with a moderate mastery of a particular language.

#### 4.2.5 Pleasant voice

Although one lecturer described Interpreter 1's voice quality as "excellent", others thought she sounded tired or monotonous. Comments on Interpreter 2's voice ranged from "[p]leasant" to monotonous, and one lecturer even stated that "[t]he interpreter's voice also tends to go higher at the end of each sentence or phrase, which becomes irritating later on".

The panel, however, focused on other elements. While one panellist described Interpreter 1's voice as "very clear and sounded good", others detected nervousness and a voice that betrayed fatigue. There were very few descriptions of Interpreter 2's voice, one exception being that she sounds "bored". Rather, there were comments on her extremely long lag time and omissions.

#### 4.2.6 Native accent

None of the lecturers commented on the interpreters' accents. One panellist referred specifically to accent, stating that Interpreter 1's "English accent is good". Two others commented that Interpreter 2 "sounds as if she is not a mother-tongue speaker of the target language", and that her "pronunciation is not good". The lack of attention given to this criterion would seem to justify the fact that it is the least important of those identified by Kurz (2002).

### 5. Conclusion

Based on their comments, the lecturers regarded five of the six criteria examined (native accent being the exception) as almost equally important, rather than singling out correct use of terminology as a particular focus area. There are at least two possible reasons for this. Firstly, the lecturers may have been guided by the format of the assessment grid, which forced them to focus on other aspects of delivery associated with "interpreting quality". Secondly, there does seem to be a slight discrepancy between what they ostensibly valued (terminological exactitude) and what they actually valued (a fluent, terminologically and grammatically correct product). Pradas Marcías (2006:38) argues, however, that new and experienced users of interpreting services focus on different aspects of an interpreted product. If the same lecturers were to use the grid repeatedly to assess interpreters, their assessment practice may change over time, and it is conceivable that a repetition of this study at a later date could yield different results.

The assessment provided by the panel certainly did not follow the order of importance for quality criteria, as established by Kurz (2002). However, their strong focus on fluency and on the hesitation in Interpreter 2's delivery seems to substantiate a view expressed by Pradas Marcías (2006:36-37) that silence has a negative impact on the perception of quality.

As there seems to be a discrepancy between the expectations expressed by this particular group of lecturers and their actual assessment of the interpreting product, and as there is some

difference between the aspects highlighted by the panel of interpreters and those raised by the lecturers, those responsible for implementing and rendering an educational interpreting service should take note that quality may mean different things to different people. This makes it difficult to answer the apparently simple question “Does the educational interpreting at SU work?” since “work” – even if one were to focus on quality rather than *modus operandi* or financial management as its definition – could be evaluated differently by various groups.

As pointed out in section 1, the study on which this article mainly relies is very limited in scope, and it would be irresponsible to jump to conclusions on possible dissonance between expectations among clients of an interpreting service and their actual assessment of interpreting performance, particularly if such conclusions were to be translated into policy (informing interpreter training, negotiations with academic staff, and so forth). What this study has highlighted is that a similar, adequately funded, and certainly much more extensive study should be undertaken to properly investigate such potential discrepancies between expected performance and actual assessment.

Once that has been done, it could be possible to properly customise interpreter training and quality assessment at SU, to properly manage the expectations of both academic staff and SU administration, and, as Kopczynski (1994:90) puts it, to negotiate about terminology.

## References

- Alexander, C., R. Edwards and B. Temple. 2004. *Access to services with interpreters: User views*. York: Joseph Rowntree Foundation.
- Bartłomiejczyk, M. 2007. Interpreting quality as perceived by trainee interpreters: Self-evaluation. *The Interpreter and Translator Trainer* 1(2): 247-267.
- Blaauw, J. 2008. Sourcing and maintaining a pool of suitably skilled interpreters for educational interpreting at a tertiary institution. *South African Linguistics and Applied Language Studies* 26(3): 301-313.
- Collados Aís, Á. 2002. Quality assessment in simultaneous interpreting: The importance of nonverbal communication. In F. Pöchhacker and M. Shlesinger (eds.) *The interpreting studies reader*. London and New York: Routledge. pp. 327-336.
- Grbić, N. 2008. Constructing interpreting quality. *Interpreting* 10(2): 232-257.
- Garwood, C.J. 2002. Autonomy of the interpreted text. In G. Garzone and M. Viezzi (eds.) *Interpreting in the 21st century: Challenges and opportunities*. Amsterdam: John Benjamins. pp. 267-276.
- Hadziabdic, E., B. Albin, K. Heikkilä and K. Hjelm. 2010. Healthcare staff's perceptions of using interpreters: A qualitative study. *Primary Health Care Research & Development* 11: 260-270.

- Kalina, S. 2005. Quality in the interpreting process: What can be measured and how? *Communication & Cognition* 38(1-2): 27-46.
- Kopczynski, A. 1994. Quality in conference interpreting: Some pragmatic problems. In S. Lambert and B. Moser-Mercer (eds.) *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam and Philadelphia: John Benjamins. pp. 87-99.
- Kurz, I. 2001. Conference interpreting: Quality in the ears of the user. *Meta* 46(2): 394-409.
- Kurz, I. 2002. Conference interpretation: Expectations of different user groups. In F. Pöchhacker and M. Shlesinger (eds.) *The interpreting studies reader*. London and New York: Routledge. pp. 313-324.
- Moser-Mercer, B. 2008. Constructing quality. In G. Hansen, A. Chesterman and H. Gerzymisch-Arbogast (eds.) *Efforts and models in interpreting and translation research*. Amsterdam and Philadelphia: John Benjamins. pp. 143-156.
- Napier, J. 2011. "It's not what they say but the way they say it." A content analysis of interpreter and consumer perceptions of signed language interpreting in Australia. *International Journal of the Sociology of Language* 2011(207): 59-87.
- Pöchhacker, F. 2001. Quality assessment in conference and community interpreting. *Meta* 46(2): 410-442.
- Pöchhacker, F. 2004. *Introducing interpreting studies*. London and New York: Routledge.
- Pradas Marcías, M. 2006. Probing quality criteria in simultaneous interpreting: The role of silent pauses in fluency. *Interpreting* 8(1): 25-43.
- South African Translators' Institute (SATI). s.a. Interpreting assessment grid for simultaneous interpreting.
- Stellenbosch University. 2013. *Calendar 2013 (Part 1)*. Stellenbosch: Stellenbosch University.
- Swart, P. 2012a. *Re: Tolking: Vrae*. E-mail to P. Swart. [Online], 12 April. Available e-mail: [lfoster@sun.ac.za](mailto:lfoster@sun.ac.za).
- Swart, P. 2012b. *Re: Tolking: Vrae*. E-mail to P. Swart. [Online], 13 April. Available e-mail: [lfoster@sun.ac.za](mailto:lfoster@sun.ac.za).
- Swart, P. 2013. Interview. 18 February, Stellenbosch.
- Verhoef, M. 2008. Assessing the quality of specialist versus non-specialist educational interpreters at the North-West University. In M. Verhoef and T. du Plessis (eds.) *Multilingualism and educational interpreting. Innovation and delivery*. Pretoria: Van Schaik. pp. 114-134.

Wu, S.C. 2010. Some reliability issues of simultaneous interpreting assessment within the educational context. In V. Pellat, K. Griffiths and S.C. Wu (eds.) *Teaching and testing interpreting and translating*. Bern: Peter Lang. pp. 301-325.