Quality in Interpreting: interpreters' VS listeners' perception

----

Murr, Katy Eloise

KATY ELOISE MURR

Quality in Interpreting:

interpreters' VS listeners' perception

Mémoire présenté à la Faculté de Traduction et d'Interprétation

Pour l'obtention du MA en Interprétation de Conférence

Directeur de mémoire : Prof. Kilian Seeber

Juré : Prénom Nom

Mois 2017

STUDENT INFORMATION:

Katy Eloise Murr

Ecole de Traduction et d'Interprétation

University of Geneva

40, boulevard du Pont-d'Arve,

CH-1211 Genève 4, Switzerland

CONTENTS

1. Introduction

1.1 Quality: a sticking point

In the wake of the Nuremberg trials the profession of conference interpreting enjoyed more limelight than it previously did (Gaiba, 1999; Bowen & Bowen 1985; Baigorri, 2014). This is in part due to the development of simultaneous interpreting which attracted media attention (Baigorri, 2015) The constant preoccupation with quality was increasingly documented too. In part this was due to the greater visibility of simultaneous interpreting. Questions regarding what characterises a good interpretation have frequently arisen. Subject knowledge, terminological accuracy, fluency, native accent, and, last but not least, professionalism are among the factors often considered (Grbíc, 2015). Research into interpreting often covered issues related to quality, including ethics and norms, an interpreter's role and working conditions (sometimes referred to as 'context') (Grbíc, 2015).

1.2 Development of conference interpreting

The advent of interpreting schools across Europe, a result of the profession's prominence on the international stage, led to more research in interpreting studies. The development of both competitive entrance examinations (written and/or oral), commonly known as aptitude tests and final examinations (Timarova & Ungoed, 2008), supports the idea that a barometer of standard has been developed. The debate about aptitude testing became a key issue in the 1960s as the discussion about interpreters' skills came to the forefront (Sawyer & Roy, 2015). This may also suggest that testing became more rigorous manner as a result. Some interpreting professionals suggest screening is more rigorously developed during the initial aptitude tests rather than the examinations which see aspiring interpreters awarded their professional diploma (Tiselius, 2013). Despite developing interpreter curriculum and assessment, one sole conception of quality in interpreting does not, however, appear to exist (Schlesinger 1997, p. 122).

The creation of professional organisations has also had a significant impact on how quality in interpreting is perceived (Keiser, 1999). The International Association of Conference Interpreters (AIIC) put quality – developed via practical experience – squarely at the forefront of the profession. This is reflected in the association's accreditation process. AIIC requires interpreters to have sufficient professional experience to become a fully-fledged member (minimum 150 days) and interpreter peers are able to vouch for every language combination after they have worked alongside you (AIIC, 2012). In other words, AIIC differentiates on the basis of professional experience and proficiency: whether an interpreter has had a large enough taste of 'real-life' interpreting to be deemed a qualified interpreter and whether their peers consider them to be proficient enough, having heard them at work, and are willing to vouch for them. Until the 1970s formal certification did not exist, and although this might seem relatively late at first glance, in comparison to the birth of the profession – post-WWII – it is actually not that long a time. During this time, however, the AIIC accreditation was the only quality benchmark for conference interpreters. Accreditation is to be distinguished from certification. Certification commonly takes the form of a test.

AIIC has a Research Committee which has backed a number of projects such as Moser's 1996 user expectation survey as well as quality criteria research (Bühler's, 1986; Zwischenberger 2010). AIIC's Research Committee's activities actively feed into one of the organisation's key aims: improving working conditions (AIIC, Thiéry, Pöchhacker 2015, pp. 13-15).

Technical equipment (or lack of) is, as Kirchhoff explains (1976/2002, p. 113), key to the whole interpreting process. Interpreting depends on 'fully functional' equipment which means that any disruptions (incorrect switching, speakers not speaking into the microphone) can and does have a direct impact on the quality of the interpreting performance. AIIC strives to improve interpreters' working conditions and a significant part of this is related to technical equipment.

Large international organisations that frequently call upon interpreters' services have also developed their own certification examinations. The UN and the EU provide non-exhaustive lists of marking criteria (UN, n.d. and EU, n.d.). These are respectively comprehension, accuracy, complete sentences, style and register, appropriate speed, suitable editing, coping with difficulties, diction and delivery for the UN; content, presentation/delivery, analytical skills and technique for the EU. The marking criteria provide insight into how employers evaluate interpreters' performance. Since one of the main aims of interpreting schools is to equip students with the suitable skills for today's market (Moser-Mercer, 1998, p. 45), this information can feed into interpreter training.

Interpreters, interpreting schools and professional organisations (be it AIIC or regional organisations such as ASTTI in Switzerland) have naturally desired to find out what constitutes quality in interpreting. This includes how aspiring interpreters acquire the skills to produce quality interpretation, and how to evaluate the quality of an interpreter's performance. Investigating quality is not merely an academic exercise: it is used to predict future progress, to evaluate, to set realistic aims during training, to assess one's own performance and to improve the standing of the profession.

1.3 Definition of conference interpreting in all its shapes and forms

It is important to define exactly what is meant by 'interpreting'. There are two main modes of interpreting: consecutive interpreting, which entails 'speaking after someone had spoken in another language' (Andres, 2015), at intervals if the presentation goes on for a long time, and simultaneous interpreting. Simultaneous interpreting involves interpreting at the same time as the speaker is speaking: it consists of a person being able to 'listen, translate, and speak at the same'[1] (Gaiba, 1999, p.12). Barik describes the process which allows interpreters to be able to do this: 'orally converting a message from one language into another as the message is being received' (1973, p. 237). This is usually done from within a booth with the technical equipment being in a booth requires (headsets, microphone, etc.). This has several advantages, namely, it allows the audience to listen to a translation of a speech 'while it is being delivered', as Paneth (2002, p. 32) points out, thus cutting down the necessary time needed for a meeting or sitting to be conducted.

---

[1]   It is important to recognise that doubts persist as to whether these cognitive tasks really are indeed simultaneous or whether the tasks are in fact rapidly accomplished one after the other exploring. Going into greater detail on this matter is, however, far beyond the scope of this project.

Alexieva (2002, p. 220) concisely sums up Salevsky's explanation of six 'varieties' of interpreting techniques (1982). The fundamental distinction made is between consecutive interpreting and simultaneous interpreting. Alexieva notes that:

1. consecutive interpreting a. can include note-taking

        b. but not necessarily (e.g. if you are walking around a site/doing a factory visit it is highly unlikely you can take notes while walking)

2. Simultaneous interpreting a. can be in a booth with a written source text

        b. in a booth without a written source text

        c. without a booth but with technical equipment ([2]bidules are commonly used today)

        d. or in chuchotage (without technical equipment)

Chuchotage (or 'whispering') interpreting consists of the interpreter sitting behind their one client and whispering a translation to them (Paneth, 2002, p. 39). Whispering interpreting can be carried out in either the simultaneous or the consecutive mode, although generally the simultaneous mode is favoured due to time constraints. The bulk of conference interpreting today, on the private or the institutional market, is conducted in the simultaneous mode. The sheer amount of current research focussing on the simultaneous technique bears out the demands of today's interpreting market.

Early studies such as Gaiba (1999) or Shveitser (1998) help explain why simultaneous interpreting has become the most prominent mode of interpreting. In 1945-46 consecutive interpreting would 'slow down proceedings unbearably' (Gaiba, p.11). Had consecutive interpreting been implemented throughout the Nuremberg trials, the already long and complex trials would have lasted about twice the time they did. Time was of the essence during these meetings and consecutive interpreting simply no longer fit the bill. If this were true at that point in history, time constraints during meetings have only increased since. Interpreter colleagues frequently report both conference duration and individual meetings being radically slashed. Rather than isolated events, this appears to be part of a whole overarching trend. For the purposes of this study, simultaneous interpreting is thus what we will generally be referring to when we speak of interpreting.

---

[2] Alexieva references technical equipment but does not mention the bidule. The bidule system is comprised of a radio pocket transmitter and a number of headphones. It is frequently used for short meetings with a limited number of participants and is ideal for events such as site visits where interpreters can see what is happening. It is not, however, a substitute for a normal booth set-up. The disadvantages are notable: sound quality, both input and output, suffers considerably.

The implication of this is that the focus will be on literature examining the simultaneous mode. However, since consecutive interpreting still exists and is in many ways simultaneous interpretation's predecessor we ought to acknowledge it. It has been shown that interpreters existed as early as ancient Egypt, serving key figures such as leaders and traders (Hermann, 1956/2002, p. 84). Aside from temporarily preceding simultaneous interpreting, consecutive interpreting and simultaneous interpreting rely on two similar skills: the ability to listen attentively and to fully grasp what is being said in the foreign language; the ability to quickly convey that information into your mother tongue in a natural way. Prior to developing the technology required for simultaneous interpreting consecutive interpreting was frequently used.

Paneth mentions interpreters' ability to rapidly analyse information and reprocess information in a more digestible form (Paneth, 2002, pp. 36-37). By speaking of 'reprocessing' Paneth assumes that the information received in the foreign language is first of all understood (i.e. 'processing') then (very quickly) converted into the interpreter's mother tongue and rendered in a clear idiomatic way ('reprocessing'). Paneth's note that this ought to be provided in a form which is even more 'digestible' relates to the idea that interpreters seek to add clarity where possible. We strive to make the information as easy to 'receive' as we can for our listeners. There are times of course when this is more difficult, such as when it is difficult to fully understand what the speaker is saying (strong accent, mumbling, poor microphone usage, several delegates speaking at the same time) if the conference sudden goes off-topic, or for a myriad of other reasons.

Given this overlap in the skill-set, it would be foolish to ignore findings on consecutive interpreting simply because the current climate generally calls for simultaneous interpreting. Some earlier research has stood the test of time and, since there is this overlap in the skill-set required for consecutive and simultaneous interpreting, can nevertheless contribute to the current debate.

1.4 Definition of terms: interpreters and listeners

By 'interpreter' I refer to professional conference interpreters[3]. In other words, interpreters who are able to conduct themselves in the most demanding of situations and embody professional behaviour (AIIC, 2004 & AIIC, 2015). This includes and extends beyond professional attitude, punctuality, upholding confidentiality (AIIC, 2014) and suitable dress code. Professional ethics includes understanding what an interpreter is not: as Herbert (1952, p. 2) points out, an interpreter is not supposed to be the centre of attention, but rather facilitate the international meeting they work at. Herbert also stresses how important good manners are. One

---

[3] Other 'forms' (Moser-Mercer, 2008, p. 143) of interpreting such as community interpreters or informal interpreters are also highly valued activities in their own right, however they will not be the focus of this investigation.

might even say, manners maketh interpreter. Herbert takes manners to mean interpreters must wear appropriate attire and give the impression they are interested in the topic being discussed. Furthermore, they ought to respect all opinions expressed, regardless of whether their own position aligns with them or not. Herbert points out how important it is that interpreters uphold neutrality and never show 'any sign of boredom, skepticism, opposition or irony' (p. 76-77). Professional ethics also includes aspects such as being aware of the conditions necessary in order for a contract to be accepted and adhering to the rules regarding the minimum number of interpreters per number of languages used.

A pertinent question is what the picture looks like for those on the receiving end of an interpreter's performance. Today the most common form of listeners is delegates listening to simultaneous interpretation in the conference room. Kremer (2007) draws our attention to the paradox at the heart of interpreting: those who are listening to us are often anything but language professionals. Most frequently their professional activities (and interests) center on technical matters, be it political, scientific or other. In this respect, Kurz (1993) is right to insist on assessment by 'delegates', not just unspecified listeners (who may happen to be other interpreters!).

The inherent need for an interpreter means these listeners generally do not understand the foreign language well enough and thus are not able to evaluate the quality of the content which is being interpreted from one language into another. Unlike examiners, trainers or fellow interpreting students, listeners (delegates) more likely than not only listen to us (this is especially true for simultaneous interpretation whereby our audience listens with a headset set to only one language). This means that listeners only judge the execution of interpretation services on the output produced in the language which we interpret into rather than the original content of the foreign language.

1.5 Listeners: (head)set on one language only?

This paradox between interpreters and their comprehension as well as those listening to them sheds light on one of the key challenges interpreters face. Interpreters have to juggle the demands of their listeners (the active language into which they interpret, generally their A language[4]) and what they regard as hallmarks of a good interpretation from the passive language. The ingredients interpreters frequently regard as making up a 'good interpretation' are generally those they have amassed during interpreting school as well as the skills

---

[4]   The A language is widely accepted as being the interpreter's native language, a B language is a language which is not the interpreter's native language but which the interpreter has a 'perfect command' of, and a C language is a passive language which the interpreter works from thanks to their excellent understanding (AIIC, 2015).

necessary for the interpreting sector which the training program does not or cannot provide (Timarova & Ungoed, 2008, p. 31), such as broad general knowledge. Yet, as is often the case during the transition from a training environment to the real-life professional world, a gap between what interpreting students are taught is important (e.g. completeness) and what listeners deem important (e.g. a voice which is pleasant to listen to for long periods of time) can arise. Any potential gaps which appear in the analysis will then be able to be thoroughly explored in the discussion section which follows.

A conference interpreting student soon learns during practice sessions that there is often a discrepancy between different listeners' perceptions of quality. Different takes on an interpreter's decisions frequently arise: i.e. was the omission a wise decision in order to convey the main message more succinctly or did the student involuntarily drop the information since they were concentrating too much on the figure which followed? It has been shown (Barik, 1971, p. 207) that more experienced interpreters make fewer mistakes than less experienced colleagues. This methodological framework of this study could, however, be improved considerably and examine qualified interpreters who are split into two groups depending on their hours of experience rather than including students in the population. Interestingly, Barik points out (1971, p. 208) that more experienced interpreters tend to add more material than their less experienced counterparts. In other words, their expertise frees them from sticking to the original too closely and, provided the added material is not context-inappropriate and does not misinform the listener, could add to the interpreter's style.

The gap in how quality is perceived is often the largest between the interpreter herself and the listener (frequently our direct client on the private market) which is what makes comparing and contrasting users' and interpreters' expectations so fascinating. Moreover, on numerous occasions interpreting tutors seek to underline the current market demands might call for succinct, to-the-point interpretation. Sometimes this might come at the price of greater detail which the original might have contained. These comments are often met with astonishment in the early stages of interpreting training, albeit confirmed by interpreters' peers who have already graduated, tested the waters, and found these comments to be backed up by their own experience in the booth.

1.6 Quality in general and as applies to interpreting

Quality as a general definition is an inherently slippery concept: depending on your point of view and perspective its meaning can change. The ideas of quality and quality control (QC) are highly prevalent in the professional environment, and a quick search throws up myriad results (e.g. Evans & Lindsay, 2010). The

very fact that the acronym (QC) and the surrounding terms (quality assurance[5], total quality management) exist also underpins this. Indeed, the workplace is frequently the place we encounter them the most. 'Quality' is often used to refer to a benchmark, standard, goal, or an expected level. Quality exists as a standard (a certain 'level'), as a means to assess or evaluate 'measure', as a comparison. This may not really tell us, however, what quality is.

1.7 Discussion of quality in interpreting

In order to meet clients' demands we need to have a sound understanding of what the client, customer, or delegate is expecting. Moser-Mercer cuts to the chase: she underlines that interpreting is not a cheap service and users' expectations are in line with the price (1998, p. 39).

Interpreting is not a cheap service first and foremost because its service providers (interpreters) are highly trained professionals. Like any other highly trained professionals who are accredited by a professional organisation (medical doctors or engineers, for instance), their initial training and continual lifelong training is reflected in their fees. An interpreter entering the market today is expected to have at least five years' study under their belt, if not more (three-four years undergraduate degree, one and a half to two years postgraduate specialist conference interpreting degree). On top of this they must have a wealth of professional and cultural know-how, acquired via living, studying and working in at least one country per each working language – without this extensive knowledge, they would not be able to keep afloat in the booth. In practice many interpreters have a great deal more under their belt too, be it before starting their career as an interpreter or via lifelong language and specialist subject training (e.g. courses in legal affairs or humanitarian aid).

As quality research has increasingly pointed out, in order to deliver a high-quality service, we need to have a precise idea of what our specific client on that very day and at that very meeting is expecting. This is linked to understanding the interpreting process as a context-orientated service, or a 'holistic' process (Behr, 2013), a concept which will be explored further in the analysis chapter.

It is no easy feat to apply the concept of quality to interpreting. For instance, one listener might give a preferable evaluation to an interpreting performance where the interpreter is a native speaker (and thus speaks with a native accent) despite less precise terminological choices whereas another listener might

---

[5] Moser-Mercer (2008) highlights how crucial piloting studies is in order to meet quality assurance; she demonstrates how QA can be concretely applied in interpreting studies.

experience this the other way around. This study thus calls for including an overview of existing attempts to define quality in interpreting (see later discussion in 2.3 and Collados Ais & Becerra, 2015). Nevertheless, the elusive nature of the concept, as stated, suggests that it would be overly optimistic to expect one single, conclusive definition of quality. The main focus of this project is a comparative discussion of quality. Although there are no illusions regarding achieving one single definitive definition of quality it is worth being ambitious and trying to find some common denominators: if the market principles above really exist and apply 'to all services' then there must be such common traits in interpreting, too.

'Quality' is certainly the sticking point but it is not impossible to propose a broad definition on the basis of researchers' consensus (Collados Ais & Becerra, 2015). This definition can be pictured as being a 'framework' definition. It will nonetheless require more time and attention than the other notions such as A, B and C languages or the distinction between different interpreting modes. Collados Ais and Becerra's recent study (2015) on quality provides a succinct overview of the concept (pp. 386-370). They point out that many researchers have previously referred to quality as a slippery concept and it is difficult to find one single, comprehensive definition (p. 286). Yet despite individual preferences prevailing to some extent, it appears there is a broad consensus. Most researchers base their definition at least partially on Bühler's pioneering quality research and the criteria she developed. To name just a few, Zwischenberger (2010), Chiaro & Nocella (2004), Kurz (1993), Pöchhacker (2005).

Behr (2013) highlights the fact that despite there being a great deal of talk about how difficult it is to pinpoint one definition, a broad consensus has gradually been established. In her words 'norms' have been developed and are being actively applied (p. 87). One salient example she provides is interpreting schools. For this very reason, it is important not to gloss over interpreting schools' criteria or evaluation processes. Although it would be salient to focus our attention on the leading and most pioneering interpreting schools such as FTI Geneva, in order to avoid bias EMCI[6] will inform the interpreting schools part of this. Professional accreditation criteria must also underpin this analysis, e.g. UN criteria or EU criteria.

1.8.1 EMCI

EMCI does not appear to provide a definition of quality nor is it mentioned in its core curriculum. It does, however, reference quality assurance standards which have been created with the aim of a commitment to quality maintenance. A thorough document outlines how these standards can be assessed (notably peer assessment) as well as the criteria involved. Finally, the document concludes with recommended good practice. It is hard to know how much this good practice actually translates into reality, however.

---

[6]   EMCI: European Masters in Conference Interpreting

An updated version of the quality assurance standards would be a welcome improvement, seeing as the document dates from Spring 2015. An updated version might take into account technological changes such as the rise of remote interpreting or the increasing use of bidule technology.

1.8.2 Accreditation: how the institutions perceive quality

As mentioned, international organisations that frequently call upon interpreters' services have their own examinations, known as accreditation tests (EU) and language competitive examinations (UN). The UN and the European Commission provide lists – albeit non-exhaustive – of marking criteria (UN, n.d. and EU, n.d.). These are as follows:

UN: comprehension, accuracy, complete sentences, style and register, appropriate speed, suitable editing, coping with difficulties, diction, and delivery.
EU consecutive (see Fig. 1 for full details): content (completeness, accuracy, coherence/plausibility), presentation/form (active language, communication skills), analytical skills.
EU simultaneous (see Fig. 2 for full details): content (completeness, accuracy, coherence/plausibility), delivery/form (active language, communication skills), technique (interpretation strategies).

The order of the list provides insight into the hierarchy of different elements and suggests skills are weighted accordingly. It is clear that in both cases, content is the number one priority. The EU's lists also group different skills, as seen in Fig. 1 and Fig. 2, and provides questions to prompt examiners' responses.

Access to marking criteria full stop is to be welcomed but greater clarity would be a positive development. The Commission's criteria for consecutive interpreting mentions dress code under 'presentation/form', for example, without specifying exactly what is expected. Interestingly, for simultaneous interpreting under 'delivery/form', the equivalent category, 'presentation' is substituted by 'delivery' and dress code no longer is mentioned. Although interpreters are generally less visible when working in simultaneous mode because they are in the booth, they are not necessarily invisible and there are times that they do come into contact with delegates. It would thus make sense to either maintain 'dress code' for the simultaneous criteria if it is to be found in the consecutive criteria or to delete it altogether.

It could be argued that dress code does not necessarily affect the quality of interpreting output itself but rather how 'professional' a service is deemed. Dress code remains wrapped up with professionalism, yet there are cases of highly trained professionals (medical doctors in a GP practice or the tech sector) which

reflect that more casual attire does not necessarily hinder one's professional performance at all. It is true, however, that dress code could affect quality.

The counter-argument would be the classic example of a lawyer who dons their robes before entering the courtroom: the robes are more than decoration, they can also provide a buffer of sorts between the person as an individual outside of their workplace and the lawyer they are while at work representing their client(s). It could also be argued that professional clothing helps provide a better work/home balance. Despite this argument that clothes can act as a buffer, as we have seen, they do not necessarily, and the work/home argument is rather tenuous for interpreters given that a great deal of preparation goes on at home anyway and it is unlikely that the interpreter will don a suit in order to sit in their office and read through the preparation documents.

In a similar vein to the comment on dress code, for consecutive interpreting the delivery time is mentioned yet there is no mention for simultaneous interpreting. Delivery time concerns whether the interpreter's version is shorter/longer than the original and if it was longer was the 'overrun excessive'. Unfortunately, further indication is not provided. Unlike other categories (e.g. omissions etc.) time cannot possibly have a true zero (there is no zero-time interpretation). Nevertheless, it is unclear how long/short the interpretation must be in order to complete the criterion. This is one example of how limiting the marking criteria are. It is difficult to see how the criteria are actually applied in reality.

Is there no mention of delivery time for simultaneous interpreting simply because it is simply expected that simultaneous interpreters keep the same pace as the speaker? Pace concerns how 'fluent or staccato' a rendition is (see Fig 2). Lag is about keeping up with a speaker, not falling too far behind and not running too far ahead. If the interpreter has a long lag it is almost impossible to render all elements and there is a greater probability that more content will be lost. There is no mention of lag at all in the EU guidelines which make it hard to determine what the expectations are.

Other criteria are particularly hard to evaluate due to their subjective nature: what exactly is meant by conveying the content in 'full' as is expected under the Commission's content guidelines? Omissions are notoriously subjective. Often an interpreter is required to slim down a delegate's speech, as more experienced interpreters and examiners point out (Buhler, 1986, p. 232).

Likewise, what exactly is meant by the criterion 'appropriate speed'? Does this mean summarising, is it linked to diction and a desire for interpreters to enunciate quickly yet accurately? What if appropriate speed comes at the price of content, who is to say which content is most important and when hurrying up is a salient choice and when not?

There is a more fundamentally problematic issue related to reverse-engineering based on interpreters' output. 'Comprehension', the UN's number one criteria, poses significant problems – how can one know and assess comprehension in an interpreting context? Use of correct terminology and sound grammatical structure does not mean that one has truly understood, so how can examiners assess this? Does comprehension simply equal correctness of output? When evaluating an interpreter's production examiners only look at the effect. It seems difficult to reverse engineer and get back to the initial criterion such as 'comprehension'.

This suggests that even highly experienced institutions also struggle to pin down what exactly is meant by quality in interpreting; the issue is one the profession as a whole struggles with and thus requires a joint effort by interpreting programmes, professional bodies (be it AIIC or regional associations), individual interpreters as well as the institutions for whom we provide our services, both private market and international organisations.

1.9 Roadmap

This paper aims to map out interpreters' and listeners' takes on quality by comparing and contrasting them. To date most research has focused on examining quality in interpreting either from the interpreter's perspective: 'the other actors in the socio-communicative constellation [have received] relatively little attention' (p. 295, *The Interpreting Studies Reader*, 2002, Examining Expectations and Norms: Introduction). Moreover, were listeners to be consulted, the 'listener' is often taken to be a professional interpreter/examiner rather than an audience an interpreter may frequently encounter such as business professionals or diplomatic delegates (Kurz, 1993, underlines this). In order to address this common discrepancy between the interpreter and listener's perception of quality in interpreting both parties' angles will be examined side by side.

This paper's findings will inform interpreters themselves, interpreting schools, interpreting associations (AIIC or regional associations) as well as interpreters' clients (both the institutional market and the private market). Interpreters will hopefully be able to better put themselves in their listeners' shoes. In doing so, interpreters might see a shift in how they perceive their performance and in particular how they prioritise different skills. Taking a step back is rarely fruitless and often a highly productive activity. Actively comparing listeners and interpreters' take on quality in interpreting can in this respect be considered something which can contribute to interpreters' deliberate practice (Motta, 2013).

The methodology will center on a review of current literature in this field. Given the scope of this research project as a Masters thesis, it would not be a judicial decision to undertake quantitative survey research. Rather, existing scholarly work will be examined in a new light by adopting the Harvard Method of comparison.

The next chapter (2) will contain a thorough review some of the cornerstones in interpreting studies research on quality. Chapter 3 will outline the specific method and show how data from the different studies were pooled. Chapter 4 will present the findings and, as such, will make up the largest segment of this research project. Chapter 5 will present an analysis of the findings and draw a conclusion. This chapter will also contain potential further avenues which could be explored. Finally, all references and appendices are to be found in chapter 6.

2. Literature review

The main intention here is not to provide insight into the individual papers' purpose and content in isolation. Rather, by grouping the papers into meaningful categories this will help establish major trends and patterns as well as the relationships between studies. A shape of the research will be fleshed out by understanding which projects are key landmark moments in the field of quality research. This chapter will also point to existing gaps in current research and may help indicate what future research could look at. Moser-Mercer's work (1996, 1998) is a good springboard for examining interpreting studies with a critical, constructive eye.

2.1 Past research: sketching out the first qualitative criteria

In 1986 Bühler conducts a ground-breaking survey on how AIIC members evaluate the performance of a potential AIIC candidate with a view to establishing criteria for evaluating conference interpretation as a service and interpreters as professionals (Bühler, 1986, p. 231). Her methodology (survey) as well as the fact that she was one of the first to look at quality in interpreting means she is generally considered a pioneer in the field. Bühler investigated how experienced interpreters perceive 'quality' in interpreting. Her population drew on experienced AIIC members as well as CACL members (AIIC's Committee on Admissions and Language Classification), although she does note that others could be potential assessors (teachers, professional organisation, chief interpreter, end user; p. 231). By including these two population groups Bühler was able to distinguish between the two.

One overarching criterion stood out: 'sense consistency with the original message'. It was an 'absolute and relative leader' (p. 231). This same criterion has remained so ever since. 98% of respondents considered it 'of high importance'. In other words, it is the criterion which was rated the most highly by both groups of respondents (AIIC interpreters and CACL members).

Bühler groups the criteria into two categories: 'linguistic criteria' and 'extra-linguistic criteria' (p. 233).[7] Her results show that the former were deemed highly important: fluency of delivery, logical cohesion of utterance, sense consistency with the original message (the key) completeness of interpretation, correct grammatical usage, and use of correct terminology. There were clear leaders among the extra linguistic criteria: reliability, thorough preparation of conference documents, and the ability to work in a team. Finally, some criteria were described as desirable albeit not essential: native accent, pleasant voice, use of appropriate style, endurance, poise, and pleasant appearance.

In her conclusions, Bühler highlights the determining role the context plays when establishing criteria for an 'ideal' interpretation or an 'ideal' interpreter (p. 233). Here Bühler pre-empts much of the current research which focusses on a 'context-based' or 'holistic' approach (see Collados Ais & Garcia Becerra, 2015). Bühler also stresses the importance of conducting oneself in an agreeable manner and working well as a team with other colleagues. Bühler's comments on decent professional manners re-join and underpin much of what Jean Herbert explains (1952).

2.2 Past research: building on qualitative parameters

Many researchers have drawn inspiration from Bühler's first survey: almost all from her criteria, especially the overarching leader, and some who decided to investigate the interpretive perspective on quality. Many surveys sprung up in the aftermath of her initial project and continue to do so today. Bühler's survey-based methodology clearly had a huge impact. Chiaro and Nocella (2004) and Zwischenberger (2009, 2010, who drew on Chiaro & Nocella's web-based approach) are a couple of notable examples. Some researchers also decided to contrast interpreters' and users' take on quality.

Of particular merit is Kurz's survey-based study (1993). Kurz also adopts a survey-based methodology, and her innovative approach meant she was the first to compare interpreters and users (delegates). Kurz' research question aimed to determine what, if any difference, would result from the quality criteria as perceived by interpreters and listeners (delegates). In other words, whether Bühler's findings from an interpreter's perspective could be extrapolated 'to the expectations of users' (TISR, p. 296, 2002). She also

---

[7] Here I take on Bühler's exact wording

intended to find out whether differences would result in the way groups of users rate the individual quality criteria. Moreover, she attempted to empirically reflect the 'alleged importance' of context (p. 15). Kurz' population includes the AIIC interpreters from Bühler's original study as well as three different types of listeners (different types of delegates at a Council of Europe meeting as well as professionals such as doctors and engineers). Given the time lapse between Bühler's study and her own, as well as the fact that interpreting is a rapidly-changing profession, it would perhaps have been salient to re-administer the questionnaire to AIIC members. Kurz does not state why she chose not to do so. Kurz's population included listeners whose native language differed (English and German) and the survey's language differed accordingly.

Again, 'sense consistency with original message' came out on top. Overall Kurz findings showed that interpreters were consistently more demanding on all the different criteria than listeners. Another take on this would be that interpreters are not only more demanding but that listeners are more realistic. Interpreters may also feel the need to protect the profession and thus actually provide a response which does not represent the reality of the situation on the conference floor. There were also discrepancies between different listeners with engineers being the least demanding and Council of Europe delegates the most (doctors were in between). Kurz concludes that the three different user groups evaluated quality differently and that these differences are worth investigating. Kurz further notes that despite a broad consensus on the importance of some criteria, opinions within user groups also differ to a significant extent. In short, delegates are less demanding than AIIC interpreters in Bühler's study (1986).

Chiaro and Nocella (2004)

Chiaro and Nocella (2004) examine the results of a study on how 286 interpreters perceive quality in interpretation. Their research question centers on examining interpreters' perception of factors (linguistic and non-linguistic) which might affect quality. The most innovative aspect of their study is their methodological framework: they were among the first to incorporate IT to administer the survey via the Internet. This facilitated the data collection and analysis steps of conducting a study. It is hence unsurprising that several other researchers (notably Zwischenberger whose work will be discussed in this study) decided to follow in their footsteps. IT can facilitate research by allowing much more data to be collected more rapidly. Provided it is collected in a sound manner, a larger sample size is generally a positive aspect of a survey. Chiaro and Nocella also affirm that quality control is important both in 'every imaginable walk of life' including interpreting (Chiaro & Nocella, 2009, p. 278). Indeed, the whole project is anchored against the backdrop of

quality control: their very first sentences begin by introducing 'systems of quality control' (Chiaro & Nocella, 2009, p. 278).

First, perhaps the most glaring aspect of Chiaro and Nocella's work is the fact is the fact that despite stressing the reasonably sizeable sample, it remains unclear exactly which population they drew on, i.e. which interpreters they recruited for their study. The authors stress that they have incorporated the results of '286 interpreters across five continents' (Chiaro & Nocella, 2009, p. 278) which they compare with Kurz' 'small, uneven sample' (Chiaro & Nocella, 2009, p. 282). Kurz' sample is as follows: 47 interpreters, 47 medical doctors, 29 engineers, 48 Council of Europe delegates. Vague or lacking information on population is a recurrent issue among the more empirical-based questionnaire or survey research. This engenders serious implications for the studies' results and conclusions: should the interpreters recruited be less well-qualified and less experienced than AIIC interpreters, it is highly probable that they would also be less demanding. In other words, a non-AIIC group of interpreters might well have lower expectations than a more highly-qualified and more experienced AIIC cohort.

Second, another point which deserves our attention regarding Chiaro and Nocella's population is the fact that the majority of the interpreters who participated in Chiaro and Nocella's work do not interpret into their A language. This is somewhat surprising due to AIIC's industry recommendations (reference?) and further suggests that the population are not AIIC members. Chiaro and Nocella describe the interpreters who participated as conference interpreters – without precisely defining what they mean by that – but it remains unclear what kind of interpreting these interpreters actually conduct and whether they actually work as conference interpreters. Overall little information is provided on how the respondents were selected. Were the respondents to work primarily as a community interpreter, for instance, the findings may not hold true for our field. An inadequate initial selection procedure for respondents could have a disastrous impact on the results' pertinence in the end.

Nevertheless, in line with Bühler's original study, the three most important factors regarding the perception of criteria affecting quality were consistency with the original, completeness of information and logical cohesion (Chiaro & Nocella, 2009, p. 287). Fluency of delivery, correct terminology and correct grammatical usage followed as a second group. Finally, appropriate style, pleasant voice, and native accent were deemed the least important. In their conclusion, Chiaro and Nocella highlight the importance of information technology - both as a way to collect data, as they did, and as a way to improve training and specialisation via more terminological databases (Chiaro & Nocella, 2009, p. 291). Chiaro and Nocella saliently underlined one of the most significant technological advances which would significantly change the way in which data is collected, analysed and discussed.

Pöchhaker (2005)

Rather than conducting a survey, Pöchhaker's more recent study (2005) focuses on examining existing questionnaire-based surveys. He suggests a critical, methodology-focused approach in order to tighten up the conclusions researchers may draw. As Moser-Mercer notes (2008, p. 144), 'literature on quality in interpreting is fairly rich in examples of survey type studies but lacks fundamental unprincipled guidance regarding survey methodology'. Pöchhaker's research question aims to (re)examine conceptual and methodological decisions in order to put quality research models 'to the test' (Pöchhaker, 2005, p. 143).

Pöchhaker adopts statistical analysis (rating scales, chi-square tests, Kruskal-Wallis test, Mann-Whitney *U*-test) in order to put survey data under the microscope. As for his population, Pöchhaker focuses principally on the study by Chiaro and Nocella (2004) but also draws other studies (Bühler 1986, Kurz 1993).

Although Pöchhaker praises Chiaro and Nocella's statistical methods (Pöchhaker, 2005, p. 158) he highlights some issues regarding imprecision. For instance, the use of the term 'perception' clouds the main distinction between two types of QE research: that is, quality research on generic expectations (notably Bühler and Kurz) and 'direct assessment' (Pöchhaker, 2005, p. 158), i.e. judging an actual interpreting performance (introduced by Gile 1990).

Pöchhaker also highlights the mystery surrounding Chiaro and Nocella's sampling procedure (or lack of). Sampling is indeed critical to the research process quite simply in order for the results to be relevant to the conference interpreter. The overview (1.3) of different shapes and forms of interpreting shows how important it is to ensure that the interpreters sampled and the population is indeed conference interpreters. In conclusion, Pöchhaker notes that quality assurance (QA) in interpreting studies ought to be more systematic and reliable and thus encourage greater attention to detail in the design and presentation of papers (Pöchhaker, 2005, p. 163).

Zwischenberger (2010)

Zwischenberger (2010) compares two web-based studies in order to this compare and contrast members of AIIC and members of the German Association of Conference Interpreters (VKD). Her research question aims to investigate whether members affiliated to national and an international professional organisation differ in the way they perceive quality (Zwischenberger, 2010, p. 127). Zwischenberger explains that the different admission requirements might go some way in explaining why opinions diverge (Zwischenberger, 2010, p. 129). Unlike Chiaro and Novella, Zwischenberger clearly defines the respondents' population

(Zwischenberger, 2010, p. 127). The assertion that there is 'no consensus on how to define quality in the field of interpreting studies' (Zwischenberger, 2010, p. 128) seems slightly imprecise in light of our previous discussion on the matter. As previously discussed, there is no single definitive definition although there does appear to be at least a minimum consensus regarding Bühler's top criteria.

Zwischenberger foregrounds her population and method by explaining her rationale: according to Zwischenberger, the interpreter's perspective has long been neglected, despite it playing a key role in being able to enter the profession (Zwischenberger, 2010, p. 128) due to AIIC membership conditions. It remains questionable whether the interpreter's perspective has truly been neglected to this extent, given the difficulty of collecting results from delegates.

Moreover, Zwischenberger's method centers on a larger research project ('Quality in Simultaneous Interpreting', conducted by the Center for Translation Studies (ZTW) at the University of Vienna) which lends her project more weight (Zwischenberger, 2010, p. 131). Zwischenberger is quite clear about the AIIC sample and the VKD sample: interpreters with a double membership who had already filled out the AIIC survey were not invited to participate in other (later) survey (Zwischenberger, 2010, p. 132). Ample information is provided regarding Zwischenberger's sample, including the average age of AIIC and VKD respondents (52 and 40 respectively), formal education, average working experience, and the main sector which they work in (non-agreement sector, i.e. private sector, or agreement sector i.e. UN family, EU institutions and so on). Here there was a significant difference between AIIC and VKD members. Just 4% of VKD respondents work for the institutional market, whereas the number reaches 33% for AIIC respondents. Zwischenberger highlights the fact that AIIC interpreters tend to have significantly more working experience (24 years as opposed to VKD respondents' 12 years): this is due to VKD also accepting conference interpreter candidates (Zwischenberger, 2010, p. 133).

Zwischenberger's results show that content-related criteria were held in similar esteem by both groups, with *sense consistency with the original* leading overall. As for form-related criteria, *correct terminology* scored highest, followed by *correct grammar* and *appropriate style* (Zwischenberger, 2010, p. 136). Rather than just providing raw data, Zwischenberger's comments help us to better gauge why respondents responded as they did (Zwischenberger, 2010, p. 136). On the notion of synchronicity, for example, one respondent provided examples of when it is more important (e.g. humorous speeches or number-heavy speeches). The discussion of native accent is worth noting: native accent is perhaps not weighted so heavily, but one respondent draws the salient link to native intonation, which is indeed 'essential to meaning' (Zwischenberger, 2010, p. 136).

In her conclusions, Zwischenberger highlights the general trend: content-related criteria are given greater precedence over form- and delivery-related parameters. This holds true for both population groups and supports the results of earlier studies (Zwischenberger, 2010, p. 140). Interestingly, Zwischenberger points out that both groups view the various quality criteria in a similar light; the single exception is *correct grammar* (Zwischenberger, 2010, p. 140). Despite the current emphasis on a context-based discussion of quality, Zwischenberger draws our attention to the fact that not all parameters vary depending on the type of meeting, notably the two main content-related criteria - *sense consistency with the original* and logical *cohesion* - do not (Zwischenberger, 2010, p. 140). Although Zwischenberger's study is not an evaluation of one specific performance, she also highlights that should a performance be evaluated, it is important to bear in mind the type of professional assignment and adjust the quality expectations accordingly (Zwischenberger, 2010, p. 141). In short, Zwischenberger's findings do not vary radically from previous findings, but the context she provides is extremely helpful.

2.3 Past research: implementing qualitative parameters

Michaela Albl-Mikasa (2012) investigates interpreter competence via a process- and experience-based model. Her population experienced conference interpreters working in the German market, although her sample (10 interpreters) is not the most extensive. Albl-Mikasa draws upon a 90,000-word corpus of detailed interviews with these respondents. The 10 interviewees were recruited at random from the 32 respondents who had previously filled out a questionnaire used in a former paper of Albl-Mikasa's (2012, p. 59). Albl-Mikasa notes that the population sample works above all on the private market. Following a brief discussion of different models of translator and interpreter competence (Pöchhaker, Göpferisch, Kutz and Kiraly - pp. 59-60), Albl-Mikasa opts for Kalina's (2006, p. 257) model of the interpreting process which divides it into pre-, peri-, in- and post-process dimensions.

Albl-Mikasa's method centers around Kalina's model (Albl-Mikasa, 2012, p. 63) which Albl-Mikasa extrapolates on (pp. 63-90) and adapts. Albl-Mikasa's results highlight an 'intriguingly differentiated picture' (p. 89) depending on the interpreter in question - in other words, despite the relatively homogeneous group of interviewees, the way in which they deal with the interpretation process varies a great deal. Albl-Mikasa thus concludes that interpreter 'competence' ought to be considered as part of the bigger picture, i.e. the personal requirements that the individuals have.

Conducting research on quality in interpreting the logical next step in order to fully respond to our clients' demands (Moser-Mercer 1996, 1998). Albl-Mikasa (2012) also rightly draws our attention to the

issue of quality control in the larger world. Albl-Mikasa argues in favour conference interpreting participating in this process in order to improve the quality of our service. According to her thesis, the key skill of an interpreter is a 'self-critical openness' (p. 85). That is to say, being able to take a step back. Just as a painting is not best observed from ten centimeters' distance, neither can interpreting be fully observed without a certain distance. Albl-Mikasa suggests interpreters regularly record their performance and listen to their bad habits. The very suggestion that 'bad habits' exists assumes that quality parameters are known. Alternatively, or in addition, she suggests asking trusted colleagues to 'inform each other later about what they noticed and what needs to be corrected' (p. 85). For those who have been through a conference interpreting course of good standing, this may go without saying. The key perhaps lies in having the tenacity to continue applying this skill rather than resting on their laurels as they begin to acquire more experience.

Moreover, an additional difficulty is the very assumption that every interpreter is fully aware of what was not conform to a certain quality standard and what these quality standards actually are. In light of this, it is important to bear in mind the history of interpreter training (many interpreters working today did not have any formal training although demographics suggest that even the older ones, at 65 and up, would have had ample choice of training programs). It is equally important to be aware that discrepancies between different interpreting programmes exist and that not every interpreter graduate will graduate with the same awareness and expectations. This latter point is a sensitive matter but ought nevertheless to be mentioned and to be taken into account.

One potential solution to overcoming these issues is for interpreters (and training interpreters) to pool their resources in order to improve their awareness of what can be improved or done differently. Being able to do this does, however, require a certain amount of openness and willingness to take constructive criticism on board. Fresh graduates might be more used to going through this process as part of their training programme and be open to continuing it as part of their professional development. It would be interesting to look at ways of encouraging this beneficial practice which perhaps wanes too quickly and too easily once interpreter's get into the swing of professional life.

Interpreter competence, Albl-Mikasa (2013)

Albl-Mikasa (2013) follows up on the process- and experienced-based model of interpreter competence. Albl-Mikasa draws on Kurz's 'Dolmetschkompetenz'[8] (2010): it is a 'general term for everything an interpreter needs to know and be able to do to perform a professional task' (p. 19). This study is underpinned by the same corpus (90,000-word semi-structured in-depth interviews with 10 professional conference

---

[8] 'Dolmetschkompetenz' basically means 'interpreting ability or skill' (my translation)

interpreters) and the same aims. Albl-Mikasa returns to investigate how the interpreters developed throughout their career and how they honed their interpreter competence (p. 17).

Albl-Mikasa's population consists of AIIC participants, the sample is the same as detailed in Albl-Mikasa (2012). Albl-Mikasa clearly details her method of conducting the interviews (p. 18). The vast majority of the interviews took place at the interpreter's home although three were conducted at the interviewer's home. This raises important questions about whether such surroundings are optimal. It would not be a surprise if the interviewees and interviewer knew each other in a personal capacity if the interviews were not conducted in a public space. It is debatable whether this is an ideal way to proceed in order to encourage critical acumen and honest responses. Research does indeed at least suggest that interviewer behaviour affects the results (Ian Brunton Smith, Patrick Strugis & George Leckie, 2016, p. 4), and arguably, interviewer behaviour varies depending on where the interview is conducted. It remains to be seen to what extent and exactly how it varies, but this is a likely conclusion.

Albl-Mikasa's results show that interpreters feel some requirements can be learned (lexical and grammatical knowledge in the foreign language(s), entrepreneurial skills, preparation skills, oral production skills and fluency, interpreting strategies). More than this, however, the interpreters stressed the importance of being interested in language; Albl-Mikasa describes this as an 'eagerness' to communicate (p. 20). In other words, it is important to have an appetite for learning the necessary skills in order to truly acquire them.

The main competence which stands out is *language competence(s)*. It is both a prerequisite prior to their university interpreting degree and their main concern throughout their career (p. 22). Other competencies are *interpreting competence proper* and *business competence*. It is interesting to note that continuous professional development seems to wear off as interpreters gain experience (p. 23) - it would be interesting to compare interpreters who do continue with continuous professional development as well as those who only incorporate on-the-job training and see if what, if any, effect this has on their interpreting performance.

Albl-Mikasa notes that the interpreters interviewed hardly stressed post-process skills (p. 85-86). She categorises these much-neglected skills are grouped into two main categories: terminological wrap-up and quality control. Given that post-process skills (terminology wrap-up and quality control) are an integral part of producing good quality interpretation, Albl-Mikasa suggests interpreters ought to dedicate more time to this part of the process. This is a question of mindset: it is easy to slip into thinking that the interpreting process 'over' once the microphone is switched off. Our interpreting practice is, however, part of a larger continuum which should encompass post-process skills such as taking stock of our performance and actively listening to our recordings to gauge which areas require closer attention. Within quality control, interpreters

ought to be able to take that crucial step-back but all too often fail to do so. The question arises why this is the case and if this is indeed the case, why measures are not implemented which encourage interpreters to regularly participate in auto-feedback. Research on quality here re-joins with research on expertise and deliberate practice. Reflective comments on quality control methods could help develop structures within interpreting associations such as AIIC or the institutions.[9] Post-process skills are thus key to producing a good quality interpretation and, Albl-Mikasa suggests, deserve more of our time and attention.

Albl-Mikasa concludes by suggesting that self-monitoring ('self-regulation') might replace continuous, deliberate practice as time goes on (p. 32). Albl-Mikasa underlines that interpreters are always on the lookout, partially 'on duty' or 'standby' mode (p. 31). This may go some way to explaining why there is this shift in how interpreter competence is developed. Albl-Mikasa describes this as a 'transitional process' (33) in which on-the-job and assignment-based activity takes precedence over time.

The suggestion is that in order to produce a good interpretation it is necessary to master language competence first and foremost which is a prerequisite for mastering interpreting competence proper. These two competencies can be considered essential to producing a quality interpretation.

2.4 The relationship between different parties' perspectives on quality

Without going into all the details of Behr's study, it is worth pointing out that Behr discusses the issue of examining different parties' perspectives on interpreting. She suggests it is a reasonable question which we ought to take into account. Although research including several stakeholders in feedback or quality assessment has been around long before Behr (2013), Behr clearly underlines that incorporating the relationships between client and interpreter is integral to producing a good quality interpreting service. Behr includes this relationship as one of the three main aspects of quality in interpreting (p. 86), i.e. one of the three key ingredients.

For Behr, the first 'aspect' which helps provide a good quality interpretation is the equivalence between the source language and the target language. The second is being aware of (and taking into account) the potential different perspectives on quality: client, interpreter, and 'examiner'. Finally, the third is the holistic approach, in other words examining the situation in which interpreting takes place. It is important to note that these three aspects are interlinked and interdependent. For instance, if the situation in which

---

[9] Although it is unlikely that the private market would implement such measures, interpreters working on the private market could nevertheless avail themselves of institutional support to help them effectively carry out auto-feedback.

interpreting takes place incurs poor external conditions (e.g. no view of the speakers, inadequate ventilation) then this would likely have an impact on all stakeholders' perception of quality, or at least all stakeholders who are aware of the poor conditions.

This is not a given, however, as it assumes that all stakeholders are aware of the certain factors' potential to influence quality. Even stakeholders who are aware of the poor conditions may not make the connection between the poor conditions and how they impact the quality of interpretation. Behr's argument relies upon the image of delegates who are relatively well-informed about the interpreting process. Just as an example, the very fact that interpreters are often still referred to as 'translators' reflects widespread ignorance about what our profession consists of. If people are unable to name the profession correctly there is scant chance of them fully understanding what interpreting is about.

Here Moser-Mercer's (1996) term 'optimum quality' is salient and re-joins Behr: 'the quality an interpreter can provide if external (working) conditions are appropriate to the task at hand' (1998, p. 40). It is interesting to think about how clients might work to improve our working conditions if they understood what this term means and how it impacts the final service interpreters are able to provide. We might liken interpreters' poor working conditions to a chef being given food which is unripe, off or simply not the right ingredients for the dishes which need to be cooked. Without the right initial ingredients or conditions, it is nigh on impossible to provide an excellent service. In short, interpreters working conditions (good or bad) have a direct impact on the quality of the service they can provide.

Similarly, Pignatoro and Velardi's study (2013) stresses that above all, the interpreter must provide a service 'in line with users' expectations' (p. 133). They note that this depends on the 'specific communicative context and culture' (129), in other words, they also advocate a context-based approach for analysing quality in interpreting.

As Bacigalupe also succinctly sums up: 'It is in the interaction between (a) (users')[10] needs and expectations, (b) constraints, and (c) skills that quality can be achieved' (p. 19). If Bacigalupe's assertion is correct, users' needs (delegates) are at the very heart of quality in interpreting and achieving quality. This means that research must be conducted to find out about users' needs and expectations and work must be done to take account of this when providing interpreting services. Only once the findings of this research are applied can the loop between a, b, and c be completed.

2.5 Interpreting as a process, service or product?

---

[10]   My addition

How we examine quality depends on our conception of interpreting - whether it is a 'good', a product, a service or a process. Most researchers today in a particular line of research which Moser-Mercer (1997) and Gile (1997) are part of trend to lean towards describing and conceptualising interpreting as a cognitive process. Several processed-based models exist to set out to demonstrate, from the macro to the micro, how it is possible to interpret in a conference setting. Process-based models, including using experimental data, have indeed existed for a considerable amount of time (Barik, 1972 & 1973). These process-based models break down interpreting into different processes which sometimes occur at the same time, often in order to better evaluate the product the process led to. Bacigalupe (2013), for instance, distinguishes between different 'processing levels'. He examines how these interact over time during simultaneous interpretation. This attempts to break-down simultaneous interpreting into smaller units. Such models can also be used for educational purposes to allow students to better grapple with distinct units.

Chiaro and Nocella's study (2004) is based on more of a service-based model, but also argues that interpretations cannot be considered 'goods': 'interpreting is a service and according to economics a service is an intangible and non-transferable economic good and thus quite distinct from a physical commodity' (281). A service- or process-based take on interpreting is without a doubt the current dominant manner in which interpreting is perceived. Moser-Mercer describes the research community's focus on the process of interpreting (1994) and suggests that we also ought to look at the interpreting product. Perhaps rather than asking whether interpreting is a process, service or a product, it is possible that it is all three of these, depending on the point during the interpreting process. It might be possible to bring these approaches together under one roof: interpreting is a (complex) process which interpreters provide as a (professional) service which results in a product (the final rendering) at the end.

2.6 Research on quality: 4 categories

It is possible to group the research on quality into four main categories depending on research methodology. This categorization is a result of this MA project itself. These four broad categories are not necessarily clear cut but serve as a model. The interconnectedness between the different categories ought to be welcomed - it suggests that researchers are open to engaging in different methodologies.

The first category mainly consists of questionnaire or survey-based research which adopts an empirical approach to examining quality in interpreting. Bühler's study (1986) and the studies it inspired (Chiaro and Nocella (2004) and Zwischenberger (2009, 2010)) can be described as participating in this category. This group constitutes a key part of quality research. These studies most frequently examine

interpreters' point of view. This is often the case even under the guise of examiners' perspective since the examiners too are generally interpreters. Some studies have examined the users' (delegates') point of view. Some studies combine both in a contrastive approach. Kurz (1993) set the foundation for this kind of analysis. A pertinent question which is often skirted or simply ignored is the issue of the small body of research engaging with users' experience of quality - the question remains whether this is a deliberate oversight or perhaps down to logistics (more difficult to find willing delegates who have spare time to dedicate to interpreting research).

The second category is the critical analysis of category one's research. This type of work is where researchers critically engage with existing studies' methodology. It appears to have been developed later, which seems logical. There appears, however, to be have been a significant – prolonged? – time delay between the first category of research and the second. It almost appears as if researchers were reluctant to engage in this kind of critical analysis. We are currently seeing more of this kind of research. Moser-Mercer's efforts (1996, 1998, and 2009) are in large part to thank: on numerous occasions Moser-Mercer sought to bring this segment of quality research to the forefront. This is reflected in comprehensive recent research such as Becerra and Collados Aís' (2015) sound overview of research to date. More of this type of research would be a welcome addition: it helps contextualise individual papers within interpreting studies and highlights where there is work to be done. Critical analysis of research methodology is not research on quality strictly speaking. It is indirectly related to quality and principally related to methods, but it has its role to play within research on quality.

The third category consists of corpus-based research which underpins a discussion on quality. Since this is from the author's perspective, who often is an interpreter, this category generally shifts towards examining the process of interpreting from an interpreter's point of view. Much of the research in this group is somewhat interdisciplinary, such as Vuorikoski's (2013) recent study of speech act theory in interpreting. Pignatorio and Velardi's study on quality assessment in media interpreting (2013) bridges rhetorical skills, researchers who conduct projects straddling interpreting studies and another field (in this case, linguistics) and provides the reader with a refreshing angle. Surveys try to elicit quality criteria from users whereas in a corpus-linguistic approach researchers' discussions are underpinned by criteria found in the first category (questionnaires/surveys).

Finally, the fourth and last category is a discussion on research without what today's researchers would describe as research which is not underpinned by empirical data. This fourth category is generally most prevalent among earlier research. It is understandable that this methodology differs so starkly from the other three since these studies were conducted prior to interpreting studies developing as its own 'field'. This

can be explained by the researchers' lack of a set of methodology of which to avail themselves, nor a distinct research area as is the case today. Likewise, among this category, there may be no explicit mention of 'quality' in so few words, but this does not mean we ought to discard this research for our purposes. Rather than digging out obscure references, I will turn to those who have stood the test of time (e.g. Herbert 1952) and demonstrate just how relevant these foundational writings remain.

A closer analysis of several papers on quality in interpreting is possible now that the key terms have been clearly defined and the outline of the four-category methodological approach drawn. In order to contextualise the issues surrounding interpreters' versus users' perspectives, we have briefly touched on a few of the wider questions in the field of interpreting studies, such as deliberate practice, to contextualise the debate surrounding different perspectives on quality in interpreting.

3. Harvard Method of comparative analysis

3.1 Comparative approach based on a point-by-point analysis

The method used for data analysis is a comparative approach. In order to efficiently compare and contrast interpreters' and listeners' takes on quality, the Harvard Method of comparative analysis (Harvard College, n.d.) ensured this paper would be clearly organised.

3.2 Frame of reference

The frame of reference, in other words, the context in which the two groups' opinions are placed (interpreters and listeners) centers around the question which many researchers have asked: to what extent do interpreters' and listeners' perspectives on the quality of an interpretation vary? Sub-questions include: do different groups of listeners (e.g. diplomatic delegations, technical professionals, profession by profession) differ accordingly in their expectations and is this reflected in their evaluation? If differences exist, what kind of differences are they – can they be grouped? Are we able to point to possible reasons or explanations for the gap?

3.3 Grounds for comparison

This is the reasoning which underpins the choice of two groups, i.e. interpreters and listeners in this case. The rationale behind the decision is reasonably clear if we consider interpreting as a process and as an

output/service. Interpreters are active in that they are busy during the process of interpreting, working on producing the output. Interpreters listen to both the passive language and the active language. They listen to the whole process, albeit they do not necessarily listen to themselves in the same way that an outside listener does. This is particularly true during simultaneous interpreting given that brain resources are 'split' between listening and speaking (see Moser-Mercer 1997 or Gile 1997 for cognitive models of the different processes during simultaneous interpreting). Listeners, on the other hand, are generally pure listeners (especially the case during simultaneous interpreting). They only hear the output and thus, as outlined (section 1.5) can only judge the quality of our interpretation on our performance, not the original speech.

3.4 Thesis

Interpreting studies research on quality has often examined interpreters' perspectives and seldom true listeners' (delegates, not interpreting tutors) take on quality. Due to the inherently distinct positions of both parties within the interpreting process from the onset (i.e. the interpreter hears both languages, the listener hears only one and is uniquely the recipient of the performance), it is expected - and has been demonstrated - that a gap exists. A comparative overview of the cornerstones of current interpreting studies research on these issues is much needed. It will allow us to gain a greater understanding of the differences which exist. Such an overview will provide a gateway to further research which could look at bridging the gap in practical terms, for instance during interpreter training either within university programmes or as part of ongoing training within the workplace (perhaps organised by AIIC or another main professional organisation). This overview will thus sketch out the most salient points and point to further avenues to be explored. It will be of relevance to all participants within the interpreting community (educators, interpreters, and clients).

3.5 Organisational scheme

The two main ways to organise a comparative analysis according to the Harvard Method are either text-by-text (analysis of the whole of A then B) or point-by-point (alternating analysis of A with B). In order to fully underscore the differences and similarities between interpreters' and listeners' takes on quality, it makes sense to conduct a point-by-point analysis to the extent that this is possible. Since we will intend to focus equally on both parties it also makes sense to proceed in this manner.

It must be acknowledged that material investigating interpreters' take on quality is far easier to obtain than the material on listeners'. This may in part be explained due to the sheer availability of interpreters. If most researchers in interpreting studies are practicing interpreters it is logical that they know many interpreter colleagues. Seeing as conferences often involve different groups of delegates it may be more difficult to find the time to inform delegates about a project and gather results than asking one's interpreter colleagues. Many researchers have however pointed out that more observational data of delegates would be welcome (Kurz, 1993, p. 20).

In order to fit within the remit of this project, it was necessary to focus on a select number of studies for the Harvard analysis, especially studies consisting of questionnaire or survey-based research which adopt an empirical approach to analysing quality in interpreting. Given the size of the project, five studies were selected to be compared in a tabular form: from Bühler's early contribution (1986), including Kurz's study (1993) which builds on Bühler's earlier research, through Peter Moser's investigation of users' expectations (1996), via Chiaro and Nocella's web-based study (2001) heralding a new era of research, to Albl-Mikasa's recent contribution on Interpreter Competence (2013).

4. Findings

4. 1. Introduction and criteria selected

For the empirical part of the project studies which focus on a range of different criterion were selected in order to be able to compare and contrast efficiently. Other studies focusing on meta-discussion, for instance, discussing the exact methodology of IS (Moser-Mercer, 1996 & 1998) will feed into the analysis of these studies and are pertinent but ought not to be listed. Vuorikoski's study (2012), for instance, which asks exactly what 'accurate' and 'faithful' mean when it comes to SI transmission of the original message and how an interpreter may grasp the speaker's intent (p. 156), is useful in that it underpins the discussion. It is not, however, suitable for analysing a broad range of content and form parameters. Researchers tend to break the parameters down into two main categories: either form and content (Moser, 1996), verbal and non-verbal or linguistic and non-linguistic (Bühler's, 1986; Chiaro & Nocella, 2001). There is a lively discussion (for instance regarding criterion 7, fluency of delivery, see Bühler p. 232) as to which criteria belong in which category. This suggests that the distinctions between the different groups of categories are not clear-cut but rather porous. The decision was made to select the following parameters which are covered in the vast majority of studies and broach both form and content:

1. sense consistency with the original message

2. logical cohesion of utterance

3. completeness of interpretation

4. correct grammatical usage

5. use of correct terminology

6. use of appropriate style

7. fluency of delivery

8. thorough preparation of conference documents

9. ability to work in a team

10. native accent

11. pleasant voice

12. endurance

This means that not all criteria represented in the individual studies are shown here in the chart. Reliability, for instance, is not included because only one study (Bühler, 1986) mentions it. In fact, reliability received the highest possible rating by the CACL group (100% rated it 'highly important') and a strong rating by the AIIC group overall (89% rated it 'highly important').

Nowadays reliability might appear to go without saying; most likely it is presumed to be a key part of thoroughly preparing oneself for upcoming conferences. However, one could argue that the first three criteria listed (1. sense consistency with the original message; 2. logical cohesion of utterance; 3. completeness of interpretation) could be considered good descriptors of reliability. In this case, reliability does not go without saying. Rather, reliability is not mentioned as a term itself but is nevertheless included in the criteria and is split into more specific sub-components.

Poise, too, was not included in this table because it was only mentioned by Bühler (1986). The word itself may sound outdated but most likely it is still part of what is expected when talking about the importance of a professional demeanour, although today it might be couched in different terms, for instance 'confidence' or 'self-assurance'. Bühler also makes the point (p. 233) that some of her criteria such as poise and pleasant appearance, can only be directly judged by the client when working in consecutive mode.

Likewise, it would be highly surprising to hear 'pleasant appearance' (Bühler, 1986) mentioned today. Stipulating specific expectations regarding professional appearance is a minefield, in particular when gender issues are taken into account. Despite interpreters' appearance not being explicitly in more recent

studies, experienced colleagues say that it certainly has its role to play, including in simultaneous mode when you generally work in a booth, for instance, if you are asked to retrieve documents from the meeting room.

In consecutive mode, chuchotage or when working with a bidule[11] it is clear that the highest standards of professional appearance are expected and ought to be upheld since interpreters are clearly in the public eye at all times. This still holds true for interpreters who are not necessarily in the public eye however, since wherever and whenever interpreters work they may come into contact with delegates and thus ought to represent the profession with the highest standards of professional appearance at all times.

Native accent is a criterion which is ranked differently according to the respective booth. It is said, for instance, that in certain booths, such as the French booth, the audience (and recruiters) have stricter expectations regarding accent (Bühler, 1986, p. 232). On the other hand, some booths are renowned for generally having a more open policy. Native accent is also a criterion which, over time, has clearly been regarded as gradually less important (studies).

Teamwork is at the heart of what interpreters do and this is borne out during interpreter training: students are widely discouraged from working in the booth on their own and emphasis is placed on the importance of strong teamwork. In practice, however, this depends on traditions within specific organisations and varies. In some institutions it might rare that colleagues leave the booth during their half-hour not interpreting, while in others it might be normal practice.

Like athletes or professional musicians, endurance (or stamina, as we may opt for today) is often what distinguishes average interpreters from good interpreters. It is inherently linked to the idea that we must be able to provide a good quality service at all times – even when tired or nearing the end of a thirty-minute session it is essential that a minimum quality is upheld. Working largely behind the scenes does not necessarily mean our appearance goes unnoticed. Moreover, just as medical doctors and lawyers often wear scrubs or robes respectively, how interpreters dress does influence their outward demeanour. Interpreters may need to acquire documents from the conference room; it is risky to hedge one's bet on being able to hide away in the booth.

4.2 Findings

Legend

---

[11]   A bidule is a headset which both the interpreter and delegates wear. The interpreter speaks into the microphone which is then transmitted to delegates' headsets.

not reported	NR
Bühler, 1986	
highly important	HI
important	I
less important	LI
Kurz, 1993	
medical doctors	MD
engineers	eng
Council of Europe delegates	CE
Moser, 1996	
large technical	LT
small technical	ST
large general	LG
small general	SG

Fig. 1

	Bühler, 1986	*Kurz, 1993*	*Moser, 1996*	*Chiaro & Nocella, 2001*	*Zwischenberger, 2013*
Criterion grouping	2: linguistic (6) and extra-linguistic (5) [12]	*NR*	2: content and formal criteria (part 2 of study)	2: linguistic (9) extra-linguistic[13] (3)	3: content (3); form (3); delivery (3)[14]
How criterion	Ranking: highly important-important-	Ranking: order of importance	Part 1: open-ended questions	Ranking: linguistic	Ranking: very important-important-

---

[12]   Fluency of delivery (7) is noted as a criteria which straddled both linguistic and non-linguistic criteria.

[13]   The full ranking of all extra-linguistic criteria included in the study is as follows (p. 288): concentration, preparation of conference documents, ability to work in team, endurance, physical well-being, mnemonic skills, encyclopaedic knowledge, absence of stress.

[14]   Two additional delivery-related criteria are not included in this table although Zwischenberger does include them (p. 135): lively intonation and synchronicity.

analysed	less important-irrelevant			Part 2: ranking (content- and formal-based criteria)		criteria Score: extra-linguistic criteria	less important-unimportant[15]
Population (N)	AIIC members (41)	CACL[16] members (6)	Interpreters (47)	Delegates: medical doctors (47), engineers (29), Council of Europe delegates (48)[17]	Delegates (201) at a range of conferences[18]: large technical, small technical, large general, small general (SG).	Interpreters (286)[19]	AIIC members and members of the German Association of Conference Interpreters (VKD).
Criteria							
1. **sense consistency with original message**	HI 96% 1st for AIIC and CACL		1st	1st: MD & Eng 2nd: CE	1st, expectation mentioned most frequently	1st linguistic	1st content

---

[15]   Data only provided in terms of relative importance; must follow structure. Moreover, data is lacking in order to examine all the criteria together based on the criteria used here: for instance for the delivery-related criteria, several criteria (lively intonation, synchronicity) are not included here which would alter the full results were they also taken into account.

[16]   AIIC commission for the admission and language classification of applicants

[17]   Kurz' delegates were solely English- and German-speaking. As Kurz suggests, other language groups may weigh criteria differently (p. 20).

[18]   Delegates were interviewed in English, French, German Italian and Spanish and presumably used interpretation services of this language; the majority of interviews were conducted in English and French (p. 150).

[19]   'Experienced working professionals'. On average these interpreters had at least 16 years' experience and interpret for around 66 hours per month. Chiaro and Nocella explicitly tried to avoid surveying trainee or novice interpreters. Most interpreters do not, however, interpret into their mother tongue (p. 285), contrary to AIIC recommendations ( ).

2. **logical cohesion of utterance**	overall 2nd[20] HI 83%	1st HI 100%	2nd	1st: MD 3rd: Eng 4th: CE	*NR*	3rd linguistic	2nd content
3. **completeness of interpretation**	6th HI 47% I 49%	7th HI 17% I 83%	5th	3rd: MD & CE 5th: Eng	34% mentioned it, with (7) & (4).	2nd linguistic	3rd content
4. **correct grammatical usage**	5th HI 48% I 50%	3rd HI 67%	6th	6th: MD & CE 8th: Eng	See (3.)	6th linguistic	2nd form
5. **use of correct terminology**	4th HI 49%	2nd HI 83%	3rd	1st: CE 2nd: MD & Eng	Very important: >35% all, >65% LT 'most important'[21] All groups: >25% very important; ST >50%; LT: most important.	5th linguistic	1st form
6. **use of appropriate style**	10th HI 7%; I 68%; LI 15%	5th HI 33%	*NR*		*NR*	7th linguistic	3rd form

---

[20] CACL members rated 100% highly important.

[21] p. 163.

7. fluency of delivery	4th HI 49%	1st HI 100%	4th	4th: MD & Eng 5th: CE	Preferred term: 'regular delivery', see (3.)[22]	4th linguistic	1st delivery
8. thorough preparation of conference documents	3rd HI 73% I 22%	3rd HI 67% I 33%	*NR*		*NR*	2nd extra-linguistic	*NR*
9. ability to work in a team	6th HI 47% I 49%	2nd HI 83% I 17%	*NR*		*NR*	3rd extra-linguistic	*NR*
10. native accent	9th HI 23% I 47% LI 28%	4th HI 50%	8th	7th: MD & Eng 8th: CE	*NR*	9th linguistic	2nd delivery
11. pleasant voice	7th HI 28% I 61% LI 9%	4th HI 50% I 50%	7th	5th: MD 6th: Eng 7th: CE	Frequently mentioned (p. 155)	8th linguistic	3rd delivery

---

22   Synchronicity is also a factor which is a concern (p. 155).

12. endurance	8th HI 24% I 59%	6th[23] HI 24% I 59%	NR	NR	5th extra-linguistic	NR

4.3.1 Sample: Kurz (1993)

Fig 2.

Interpreters' ranking *(criteria number in overall table, fig. 1)*	delegates agree	delegates disagree	
		deem criterion more important than interpreters (ranking per delegate group)	deem criterion less important than interpreters (ranking per delegate group)
1. sense consistency with overall message *(1)*	MD Eng		CE (2)
2. logical cohesion of utterance *(2)*		MD (1)	Eng (3) CE (4)
3. use of correct terminology *(5)*		CE (1) MD & Eng (2)	
4. fluency of delivery *(7)*	MD Eng		CE (5)
5. completeness of interpretation *(3)*	Eng	MD & CE (3)	
6. correct grammatical usage *(4)*	MD CE		Eng (8)

---

[23]   Information not provided for the two groups separately.

7. pleasant voice *(11)*	CE	MD (5)	
		Eng (6)	
8. native accent *(10)*	CE	MD & Eng (7)	

4.3.2. Kurz' sample: commentary

It makes sense to use Kurz' study as an example since this study provides us with both interpreters' and listeners' perspectives, which actually proved relatively difficult to find. All of Kurz' initial eight criteria are included here, however four criteria included in other studies (25% of the criterion listed here) were not examined in Kurz' study: use of appropriate style (6), thorough preparation of conference documents (8), ability to work in a team (9) and endurance (12). As suggested earlier, this may be because over time these characteristics are now deemed par for course; no interpreter could truly survive without embodying them.

Kurz chose to use the first eight of Bühler's original criteria in order to make the process of comparing and contrasting with Bühler's study easier (p. 15). Aside from the extra-linguistic criteria 1 (native accent) and 2 (pleasant voice), Kurz chose to retain solely linguistic criteria - this may be due to Bühler's emphasis on these criteria or because Kurz believes that these criteria are easier for users, too, to evaluate.

As Bühler and Kurz underline, interpreters' voice, in particular, appears to be of great interest and, as one of Bühler's informants commented, 'may have a disproportionate effect either way' (p. 232). Kurz notes (p. 15) that both Gile (1990) and Kurz (1988) found that quality of voice is held in higher esteem by interpreters than by end users.

Kurz' hypothesis that delegates' expectations may vary depending on their professional group could thus be tested by analysing different groups of listeners.

It made more sense to compare delegates' views with interpreters' ranking given that delegates' rankings were separated into professional categories. While this provides more in-depth detail and is thus welcome, it is simply clearer to use interpreters' ranking as a baseline. Kurz breaks delegates' views down into three different groups: medical doctors, engineers, and Council of Europe delegates.

In order to facilitate comprehension, the rankings were broken down into the following categories: agree (in other words, the criterion received the exact same ranking as the interpreters gave it); disagree and deem the criterion more important (the criterion thus received a higher ranking than interpreters gave it); disagree and deem the criterion less important (the criterion thus received a lower ranking than interpreters gave it).

No criterion enjoyed the full agreement of all users. For three out of eight criterion the majority of users (two thirds) agreed entirely with the interpreters' ranking: sense consistency; fluency of delivery; correct grammatical usage. For three out of eight criteria one group of listeners agreed: completeness of interpretation; pleasant voice; native accent. For the remaining two out of eight criteria (a mere 25%) none of the delegates agreed with the interpreters: logical cohesion of utterance, use of correct terminology.

At first glance it may sound as if delegates' and interpreters' rankings vary wildly. However, a closer look at the results reveals that this is not the case: at most any given criteria is ranked one or two numbers higher or lower than interpreters'. There is also a relative homogeneity among the different groups of delegates: all agreed with interpreters on three criteria.

Criterion 1 (sense consistency with overall message), as Kurz explains, remains the top criterion, as expected based on previous studies (Bühler, 1986). Kurz mentions that the Council of Europe delegates' ranking (they ranked criterion 5, correct terminology, first in the overall table) is an exception (p. 17). It is possible to go even further; the Council of Europe delegates' ranking is an outlier both within Kurz' study and when compared with ranking across a range of studies: sense consistency with the original message is repeatedly number one.

As has been noted, experienced users of conference interpreting services –which Council of Europe delegates are likely to be due to the diplomatic nature of their work – frequently stress the importance of correct terminology (Bühler, p. 232). This does not usually, however, occur to the detriment of criterion 1, sense consistency.

Interestingly, the criteria interpreters ranked most highly: sense consistency (1), logical cohesion of utterance (2) and use of correct terminology (3) proved to be the most contentious; none of the groups of end users ultimately agreed. Council of Europe delegates, as noted, ranked 'use of correct terminology' first, but this is considered an outlier.

It is important to note that joint rankings skew results generally. Here, medical doctors' joint first is an excellent example of this. Medical doctors attributed a joint first to two criteria – sense consistency and logical cohesion of utterance – which skews the results. Thus, in five cases medical doctors deemed a criterion more important than interpreters.

Had medical doctors ranked logical cohesion (2) just after sense consistency (1) rather than awarding it a joint first, their ranking of the top three criteria would have been identical to the interpreters' rankings in Kurz' sample. Yet the fact that medical doctors awarded a joint first changes the results and gives the impression that this professional group consistently ranks quality criteria higher than interpreters. The figures themselves – five criteria out of eight ranked higher than interpreters – express this.

It is reassuring to see a high degree of homogeneity across the board, even if details regarding one or another specific criterion vary: more specifically, the three professional groups each agree with interpreters on three counts. Medical doctors rate five criterion more highly than interpreters, as explained; for engineers, the figure is three; for Council of Europe delegates, two. Medical doctors (unsurprisingly, due to the joint first which produces skewed results) thus do not rank any criteria less important than interpreters; engineers deem two less important; Council of Europe delegates three.

Kurz's responses from both interpreters and different groups of delegates (assuming the Council of Europe delegates' results is an anomaly) confirm previous results: sense consistency with the original message is the resounding number one quality criterion for interpreters and their users.

Interpreters rank logical cohesion of utterance (criterion 2) second, the same as Zwischenberger's interpreters who rate it the second content-based criteria. Zwischenberger's study is easiest to compare due to the original study adopting linear numbered rankings. Bühler's study also rates this criterion second overall, underpinning the general consensus; 83% of AIIC interpreters rate it highly important, although 100% of CACL members deem it highly important, for whom it is a first.

5.1 Discussion of individual criteria

There is a broad consensus regarding which criteria were chosen and reported. This is probably due to the criteria having been established by Bühler (1986), yet the question remains how Bühler decided on these criteria. Bühler does not specify exactly but the most likely hypothesis is a combination of her own personal experience, colleagues and CACL and AIIC members.

5. 1.1 Sense consistency with original message (criterion 1)

Bühler, 1986	Kurz, 1993 interpreters	delegates	Moser, 1996	Chiaro & Nocella, 2001	Zwischenberger, 2013
HI 96% 1st for AIIC and CACL	1st	1st: MD & Eng 2nd: CE	1st	1st linguistic	1st content

This criterion enjoys very broad consensus both among interpreters and listeners and across the different studies. Given it is the number one criterion this suggests interpreters' and users' expectations align.

The only exception is Kurz' Council of Europe delegates. This result is clearly an outlier, however, and does not negate the findings as a whole. Council of Europe delegates ranked criterion 5 (use of correct terminology) first – possibly due to the fact that it is indeed impossible to accurately render the original message without first grasping the terminology in the source language, and second, being able to locate the correct terminology sufficiently quickly in your target language and produce an appropriate interpretation.

Zwischenberger's decision to split the criteria into three different groups (content-related, form-related and delivery-related) also suggests a hierarchy of sorts since the results are presented in that order. It would appear that Zwischenberger deems content-related criteria of utmost importance since these are the criteria presented first, followed by form-related and delivery-related criteria. Although Zwischenberger separates form-related and delivery-related criteria, it remains questionable to what extent these two categories are actually different. It might be the case that 'form' is what contains the 'delivery'. The two words might not be exact synonyms, but they do appear to be closely connected in any case. Chiaro and Nocella follow Bühler's division into linguistic and extra-linguistic criteria.

Although Moser's results are somewhat more difficult to compare in a linear fashion, delegates' responses to his open-ended questions (part 1 of the survey) also support the ranking results. When asked what makes a good interpreter, sense consistency with the original is clearly the expectation mentioned the most frequently and hence of utmost importance to listeners. It is important to note that Moser surveyed a wide range of delegates at a vast array of conferences which are classified as follows: large technical; small technical; small general; large general.

This can be seen in the bigger picture of a broader trend towards focusing on context-based expectations and towards methodology which aims to take this into account. Moser-Mercer (1996) elaborates the concept of 'optimum quality' which depends on external (working) conditions.

Moser elaborates, explaining that external conditions include (but are not limited to) aspects such as the physical environment within the booth and how comfortable it is, the complexity of subject matter, shifts between different subjects, how adversarial the atmosphere is, the type of discourse (for instance, is it highly emotionally charged or is it incoherent at times?), the delivery, how well-prepared the interpreter is (are the documents made available? If so, was there any time to read and prepare them?), are delegates talking over each other, and so on (p. 44-45).

It is logical that there are significant differences between meetings of a more general nature and highly technical meetings, and these affect issues such as how much specific terminology is required (criterion 5), the type of preparation the meeting calls for (criterion 8), as well as how tiring a meeting is (endurance, criterion).

Differences also occur depending on the size of a meeting: larger meetings mean that interpreters are often more likely to be placed further away from delegates and it is often difficult to see who is actually speaking. Delegates are also more likely to speak over each other, often the fact that there are more participants means more differing views and people wanting to bring up a wide range of issues, thus the subject matter changes both frequently and sometimes radically.

5.1.2. Logical cohesion of utterance (criterion 2)

Bühler, 1986	Kurz, 1993 interpreters	delegates	Moser, 1996	Chiaro & Nocella, 2001	Zwischenberger, 2013
Overall 2nd HI 83%	1st HI 100%	2nd	1st: MD 3rd: Eng 4th: CE	NR	3rd linguistic

Overall this is the second most highly rated criterion and this criterion also enjoys broad consensus. Unfortunately, Moser's study did not report on this criterion, but Kurz' findings, in particular, the medical doctors' results, back up interpreters' rankings. There is slight variation among interpreters: Chiaro and Nocella's interpreter respondents rank logical cohesion of utterance third place and criteria 3 (completeness of interpretation) second place, whereas Zwischenberger's interpreter respondents rank logical cohesion of utterance second place. Bühler's respondents rank this criterion very highly, with AIIC respondents ranking it second place and CACL members a joint first: the former rank logical cohesion of utterance as 'highly important' by 83% and 100% for the latter.

In Kurz' study, medical doctors ranked this criterion as a joint first with the criterion 1 (sense consistency with original message). This makes sense since if an interpretation is lacking logical cohesion it cannot make sense: sense consistency is heavily dependent on some extent of logical structure which underpins the utterance. The medical doctors' results can thus shed light on the interaction between criteria.

Kurz interpreters' results also support Bühler and Zwischenberger's results, providing broad consensus on all levels for this criterion.

5.1.3. Completeness of interpretation (criterion 3)

Bühler, 1986	Kurz, 1993		Moser, 1996	Chiaro & Nocella, 2001	Zwischenberger, 2013
	interpreters	delegates			
6th	7th	5th	3rd: MD & CE	34% mentioned it, with (7) & (4).	2nd linguistic
HI 47%	HI 17%		5th: Eng		
I 49%	I 83%				

This is a criterion which appears to be important but hierarchically is significantly less important than the previous two criteria. The CACL members' rankings say a great deal about this: only 17% of CACL members ranked criterion 3 as highly important, whereas 83% ranked it as important. This suggests that criterion 3 is lower down the priority list. Although AIIC members are experienced interpreters they are typically less experienced than those on AIIC's special CACL committee; 47% AIIC members ranked completeness of interpretation as highly important, 49% as important. This response clearly ranks completeness of interpretation lower (7 overall) but it is a far fewer nuances response than CACL's. Nevertheless, the difference is not so large: criterion 3 is ranked sixth by AIIC members and seventh by CACL members. As Bühler suggests (p. 232), experienced professionals may be more aware of the impossibility of conveying all of the original message and of the need to edit on the one hand and on the other hand due to a large among of superfluous information in the average speaker.

Overall this criterion is ranked significantly higher, however. In Kurz' study medical doctors and Council of Europe delegates rank it third. This ranking also enjoys the support of Zwischenberger's interpreter respondents. In Kurz' study, however, engineers and interpreters rank completeness of interpretation lower, in fifth place.

Interestingly, Moser notes that along with criterion (7) and (4) as well as 'absence of hesitation' and 'clarity of expression', completeness of interpretation is mentioned spontaneously by 34% respondents when asked what, along with faithful rendition to the original, makes a good interpreter (p. 155).

Chiaro and Nocella's interpreters rank completeness of interpretation higher than the majority of interpreter respondents: second place. Similarly to how in Kurz' study medical doctors perceived (2) logical

cohesion of utterance as being part of (1) sense consistency with the original message and vice versa, it appears that Chiaro and Nocella's interpreter respondents took the same approach to (1) sense consistency with the original message and (3) completeness of interpretation. In other words, they perceive faithfully conveying the original utterance's meaning as being heavily dependent on how much of the original information/content is effectively conveyed: the how good is closely linked to how much.

5.1.4. Fluency of delivery (criterion 7)

Bühler, 1986	Kurz, 1993 interpreters	delegates	Moser, 1996	Chiaro & Nocella, 2001	Zwischenberger, 2013
4th HI 49%	1st HI 100%	4th	4th: MD & Eng 5th: CE	Preferred term: 'regular delivery', see (3.)[24]	4th linguistic

The general consensus is that this criterion remains of relatively high importance: both Chiaro and Nocella's intepreters as well as Kurz' intepreters and delegates (medical doctors and engineers, not Council of Europe delegates) rank it fourth place. Bühler's CACL respondents place this among their top criteria: fluency is ranked first along with sense consistency with the original message (criterion 1) and logical cohesion of utterance (criterion 2).

For Zwischenberger it is the first delivery-related criterion (seventh place overall among all criteria). It is logical that this criterion be classified as such, given that is precisely what the name indicates. Moser mentions that a 'regular delivery' (a synonym for 'fluent delivery') was mentioned by 34% respondents when asked what, along with faithful rendition to the original, makes a good interpreter (p. 155).

This is in stark contrast to Bühler's original results, whereby AIIC members rated fluency fourth and 100% of CACL members rated fluency as highly important – according to Bühler, reflecting users' needs. As we have seen, Kurz' delegates also rank fluency highly and for Moser's respondents this criterion is of great importance.

Although we are dealing with delivery-related aspects, fluency is a big part of first impressions. It is thus closely linked to providing a professional and polished service. A fluent interpretation enables the listener to concentrate on the message's content rather than being distracted. By speaking with a pleasant voice the interpreter ensures that they do not detract from the meaning. Content is certainly important when

---

[24] Synchronicity is also a factor which is a concern (p. 155).

it comes to quality in interpreting, but it is equally important to remember that the aim is to sound as much like a natural native speaker speaking normally as is possible.

It is reassuring to see that on the whole, broad consensus is achieved: there are no two ways about it, fluency is very important.

5.1.5. Use of correct terminology (criterion 5)

Bühler, 1986	Kurz, 1993 interpreters	delegates	Moser, 1996	Chiaro & Nocella, 2001	Zwischenberger, 2013
4th HI 49%	2nd HI 83%	3rd	1st: CE 2nd: MD & Eng	Very important: >35% all, >65% LT 'most important'[25] All groups: >25% very important; ST >50%; LT: most important.	5th linguistic

Three groups (Bühler's CACL respondents) as well as Kurz' medical doctor and engineer delegates) ranked this in second place; Kurz' Council of Europe delegates produced an outlier by ranked it first place. Kurz' interpreter respondents ranked it slightly lower (third place), Bühler's AIIC members settled for even lower (fourth place). Chiaro and Nocella's interpreter respondents ranked this criterion in fifth place; overall, Zwischenberger's interpreter respondents plumped fifth place (first in the form-related criteria).

This was one of the criteria which was consistently ranked higher by delegates than by interpreters. All of Kurz' groups of delegates ranked it above third place: for medical doctors and engineers correct terminology usage was a joint second (although medical doctors' joint first means that otherwise, it would have been third) and Council of Europe delegates went so far as to rank it in first place. The logical

---

25   p. 163.

conclusion is that interpreters underestimate the importance of terminological accuracy for their users and more emphasis ought to be placed on it in interpreter training.

Moser notes that correct terminology was ranked very important by over 35% of all participants and over 65% for technical meetings. It does not come as a surprise that in technical meetings using correct terminology takes on a greater importance. Perhaps even in meetings which interpreters may classify as 'general', the terminology is of greater importance than interpreters estimate.

In large technical meetings terminology is noted as being of higher importance than in small technical meetings. This is likely because the larger the meeting, the more the meeting tends to be formal and thus more is at stake.

Regarding how terminology related to the other criteria, Bühler notes that due to CACL members rating 1 (sense consistency with original message) and 5 (use of correct terminology) the same (83% rated them highly important), the two are interlinked. In other words, it is necessary to use correct terminology in order to render the original message faithfully (p. 232). Bühler points out that if this is not the case interpreters create additional work for their listeners, who are expected to make the right inferences – despite, of course, this not being guaranteed by any means.

5.1.6. Correct grammatical usage (criterion 4)

Bühler, 1986	Kurz, 1993		Moser, 1996	Chiaro & Nocella, 2001	Zwischenberger, 2013
	interpreters	delegates			
5th	3rd	6th	6th: MD & CE	See (3.)	6th linguistic
HI 48%	HI 67%		8th: Eng		
I 50%					

Zwischenberger classifies correct grammatical usage among form-related criteria which reflects its status as of secondary importance compared to content-related criteria. This is borne out by the other results. The overall consensus (shared by Kurz' interpreters, medical doctors, Council of Europe delegates as well as Chiaro and Nocella's interpreters) ranks criterion 4 in sixth place.

Bühler's interpreter respondents rate grammar slightly higher (fifth place) and Bühler's CACL group puts it even higher (third place), as well as Zwischenberger's interpreters (fifth place). A possible

explanation for CACL's emphasis on correct grammatical usage may be that over time experience has shown CACL members that correct grammatical usage is a key component of the logical cohesion of utterance.

Along with criterion 3 (completeness of interpretation), criterion (7) 'absence of hesitation' and 'clarity of expression', completeness is mentioned spontaneously by 34% respondents when asked what, along with faithful rendition to the original, makes a good interpreter (p. 155). This bundle of criteria which respondents mention spontaneously suggests that content- and form-related criteria certainly are closely interlinked and that it is not possible to achieve one without the other. Correct grammatical usage is also part of what provides a polished interpretation – interpreters' aim not only to convey the information provided but also to do so in a way which makes it understandable and, if possible, easily digestible.

5.1.7 Pleasant voice (criterion 11)

Bühler, 1986	Kurz, 1993 interpreters	delegates	Moser, 1996	Chiaro & Nocella, 2001	Zwischenberger, 2013
7th HI 28% I 61% LI 9%	4th HI 50% I 50%	7th	5th: MD 6th: Eng 7th: CE	Frequently mentioned (p. 155)	8th linguistic

All studies included in the chart examined voice quality and overall, having a pleasant voice is ranked slightly higher than native accent (criterion 10), in seventh place – according to Bühler's AIIC interpreters, Kurz' interpreters and Council of Europe delegates as well as Chiaro and Nocella's interpreters.

Zwischenberger's interpreters ranked it slightly lower (ninth). On the contrary, Kurz' engineer delegates ranked it sixth, while the medical doctors ranked it fifth, a rating supported by Bühler's CACL team.

A pleasant voice can be compared with fluency of delivery (criterion 7): both are likely related to first impressions. Regardless of how precise an interpreter is, how much content is conveyed, how well the

terminology flows, if listeners detect stress, strain, boredom or fatigue while listening to an interpreter, it does detract from the message and sometimes considerably. Often, during meetings where interpreting services are provided, the stakes are already high, tensions can be running high and listening to a relatively enthusiastic pleasant voice does make a difference.

It is nevertheless one of the more difficult criteria to pin down: what exactly makes someone's voice 'pleasant'? It is easier to say what is not pleasant to listen to – an interpreter who sounds rushed or bored, for instance, than saying what a pleasant voice sounds like. At first glance, a steady delivery may spring to mind, but again, there is the risk of sounding too pedestrian and not conveying sufficient enthusiasm or emotion. For criteria which are difficult to define it would be interesting to ask more open-ended questions in order to elicit responses from users and gain a better idea of what these criteria may mean to them.

5.1.8 Thorough preparation of conference documents (criterion 8)

Bühler, 1986	Kurz, 1993		Moser, 1996	Chiaro & Nocella, 2001	Zwischenberger, 2013
	interpreters	delegates			
3rd HI 73% I 22%	3rd HI 67% I 33%	NR	NR	2nd extra-linguistic	NR

Three of the five studies do not report this criterion. Interestingly, neither Kurz, Moser nor Zwischenberger report on criteria 8 (thorough preparation of conference documents), 9 (ability to work in a team) or 12 (endurance).

Despite a lack of results, this in itself is telling. A likely hypothesis is that these skills are expected to 'go without saying'. In other words, they are so important in order to provide a good interpretation that it would be impossible to survive without them. Taking thorough preparation (criterion 8) as an example, it is clearly impossible to use the correct terminology for a specific organisation and a specific topic without

being familiar with both the organisation's terminology on a broader level and the specific terminology relevant to the topic(s) dealt with. Thorough preparation is also ingrained in all new interpreters who undergo interpreter training.

Another plausible explanation would be that this is a very indirect criterion. As a result, researchers might struggle to find a reliable way to detect it in the interpreter's performance. Researchers'' inability to detect the criterion would thus lead to them leaving it out.

The results available support the hypothesis that this criterion was not forgotten or neglected due to insignificance, on the contrary: Bühler's interpreters and CACL members ranked it third; for Chiaro and Nocella's respondents it was the second extra-linguistic criteria, just below concentration. More documentation on users' expectations would be welcome in order to compare and contrast efficiently across a broader range of criteria.

It must be said that documents can only be thoroughly prepared should they be made available to interpreters. This is not always the case, for a number of reasons such as documents not being ready until last-minute, documents being so confidential they cannot be taken out of the organisation's building or sent via email. Herbert draws our attention to the fact that a lack of documents mean interpreters cannot prepare adequately (1952, p. 79). It is true, however, that in reality documents are not always made available. Herbert also underlines the importance of ordering the documents so that the relevant information can be accessed rapidly. There is scant information available on this and it would be interesting to see different methods of classification (and annotation, too).

5.1.9. Use of appropriate style (criterion 6)

Bühler, 1986	Kurz, 1993		Moser, 1996	Chiaro & Nocella, 2001	Zwischenberger, 2013
	interpreters	delegates			
10th	5th			7th linguistic	3rd form
		NR	NR		
HI 7%; I 68%; LI 15%	HI 33%				

Kurz and Moser chose not to report this criterion. There are two potential reasons which spring to mind: perhaps Kurz and Moser did not consider this criterion to be sufficiently important, or 'appropriate style' is such as slippery, elusive concept that they decided not to attempt to tackle it.

What good style means depends on the language in question. Good style in French meets French-speaking listeners' expectations rather than English-speakers' expectations, and vice versa. Good style is generally considered to be unobtrusive in English, in order to ensure that the 'sense can be immediately apprehended' (Dummett, 2003, p. 62). This implies the following: having a clear message, avoiding ambiguity and using a structure which makes the message as easily digestible as possible for the listener (Dummett, 2003, pp. 62-63). Often it is easier to point out examples of poor style rather than to teach good style. Style is, as expected, not necessarily easy to teach.

Rather than 'appropriate style', perhaps it would be easier to aim for clearer goals which are part of good style, for instance, clear syntax or grammatical correctness. Grammar is itself a sticking point. When faced with speeches delivered at a quick rate, interpreters have to resort to survival methods which might include staccato statements and individual clauses rather than whole sentences. This would not be considered good grammar in other situations, but, as brought to our attention by early researchers (Herbert, 1952, p. 85), above all, interpreters need to adapt to the circumstances and know how to tailor their approach.

Zwischenberger counts this among form-related criteria; Bühler as an extra-linguistic criterion, Chiaro and Nocella as a linguistic criterion. Style is notoriously difficult to evaluate: unlike grammatical structure, a style which one delegate enjoys might not be enjoyed by all: it is clearly among the most subjective of all criteria.

In terms of rankings, style clearly ranks lower than using terminology correctly: the highest ranking it receives is from Bühler's CACL respondents (fifth place), then Zwischenberger's interpreters (sixth place; third form-related criteria), followed by Chiaro and Nocella's interpreters (seventh linguistic criteria) and finally Bühler's AIIC interpreters (tenth). It is a shame that there are no rankings based on users' expectations and that little is said in Moser's discussion and analysis. Even if this criterion is considered less important it would nevertheless be helpful to gain a better idea of how style links into the bigger picture and where precisely users place it.

5.1.10. Ability to work in a team (criterion 9)

Bühler, 1986	Kurz, 1993 interpreters	delegates	Moser, 1996	Chiaro & Nocella, 2001	Zwischenberger, 2013
6th HI 47%	2nd HI 83%	NR	NR	3rd extra-linguistic	NR

I 49%	I 17%

As indicated, Kurz, Moser, and Zwischenberger did not include this criterion in their studies and as such no data is available. Since only two studies surveyed users' expectations (Kurz and Moser) it is thus not possible to provide a point-by-point analysis for this criterion either.

However, the available data suggest that this criterion is of great importance. Interpreters always work in teams and thus it is in each interpreter's interest that the team as a whole performs well. Bühler's AIIC interpreters ranked teamwork sixth, the CACL group second. This is a significant difference in ranking – as Bühler suggested regarding fluency of delivery (criterion 7), quite possibly CACL users, due to their expertise and extensive experience, are able to better gauge users' expectations.

Were this the case, it is likely that users would rank this criterion highly. As ever, however, issues surrounding users' (often insufficient, incomplete) understanding of the processes involved in conference interpreting and what really goes on in the booth would call for users to gain a better understanding of the process in order to evaluate criteria such as teamwork. Users certainly may hear interpreters handing over to their colleague when their thirty minutes stint has elapsed but it remains highly unlikely that they grasp the full extent to which teamwork can make a significant difference.

Due to the fact that interpreters today primarily work in booths in the simultaneous mode, users do not necessarily witness the true teamwork which does go on, for instance, a colleague providing a document which is out of reach for their colleague who is interpreting; looking up a specific term which is unfamiliar; even pouring a glass of water so that when the channel briefly switches to the interpreter's mother tongue they can quickly sip a bit of water before carrying on.

Chiaro and Nocella's interpreter respondents rank the ability to work in a team ninth overall and third among extra-linguistic criteria, following shortly behind thorough preparation of conference documents (criterion 8) which is ranked eighth[26].

5.1.11. Native accent (criterion 10)

Bühler, 1986	Kurz, 1993		Moser, 1996	Chiaro & Nocella, 2001	Zwischenberger, 2013
	interpreters	delegates			
9th	4th	8th	7th: MD & Eng		9th linguistic

---

[26]   In terms of this table; but Chiaro and Nocella actually include more extra-linguistic criteria.

			8th: CE	*NR*	
HI 23% I 47% LI 28%	HI 50%				

On the whole, speaking with a native accent is clearly among the criteria deemed less important. The general consensus is that it is ranked eighth (by Kurz' interpreters and Council of Europe delegates; by Zwischenberger's interpreters), yet a few differences do exist.

Kurz' medical doctors and engineers rank it slightly higher, in seventh place; Bühler's CACL members much higher, in fourth; Chiaro and Nocella's interpreters, however, put native accent slightly lower down the list, in ninth place. The lowest ranking is provided by Bühler's AIIC respondents, at tenth place.

Although Moser does not report on native accent, voice quality is frequently mentioned (see criterion 2, pleasant voice) and it seems understandable that some users would correlate having a pleasant voice with speaking with a native accent.

It is unsurprising that some groups rated this criterion significantly lower. Zwischenberger's population, for instance, consists of both AIIC and VKD members. While AIIC always recommends (source) interpreting into one's native language, VKD members work primarily on the private market (p. 134) where one more frequently works into a language which is not one's native language (known as a 'B language'). Chiaro and Nocella's interpreters also ranked this criterion extremely low but it, again, is unsurprising given that themselves most do not interpret into their native language (p. 285).

The importance of native accent is also widely acknowledged as varying depending on the booth in question, with some booths, for instance, the French booth, being notoriously less tolerant about foreign accents, and the English booth quite the opposite (Bühler, p. 232). It is important to note that Bühler does not corroborate this claim; it is unsubstantiated.

5.1.12. Endurance (criterion 12)

Bühler, 1986	*Kurz, 1993* interpreters	delegates	*Moser, 1996*	*Chiaro & Nocella, 2001*	*Zwischenberger, 2013*
8th HI 24%	6th[27] HI 24%	*NR*	*NR*	5th extra-linguistic	*NR*

---

[27]   Information not provided for the two groups separately.

I 59%	I 59%				

It made sense to retain the term Bühler uses, 'endurance' despite the term 'stamina' most likely being what we would use today. Along with criteria 8 and 9, this last criterion was not reported by Kurz, Moser or Zwischenberger. Once again, it is impossible to conduct a point-by-point analysis.

The lack of data on this criterion may be explained by the difficulty in asking users to rank interpreters' stamina – how are users to gauge how much an interpreter has worked in one said day, when an interpreter last had a break, and so on? It seems impractical at best to ask users to gauge this. If surveying users' expectations, it might be best to reword the criterion and ask them about maintaining a minimum quality throughout or ask them how important it is that the interpretation service is consistently good.

Although Bühler only presented the AIIC and CACL members' ranking together, with 24% rating endurance 'highly important' it is thus in 8th place for AIIC respondents and sixth place for CACL members. CACL members' rankings are skewed by them awarding first place to three different criteria (sense consistency with the original, criterion 1; logical cohesion of utterance, criterion 2; fluency of delivery, criterion 7 as well as joint second and third places). In terms of numbers it is sixth place for CACL members but in terms of relative importance, it remains quite far down.

For Chiaro and Nocella endurance is the fourth extra-linguistic criteria and their interpreter respondents rank it close to the ability to work in a team (criterion 9, ranked ninth). Teamwork received a ranking of 1218 whereas endurance was just under, at 1208.

5.2 Summary of findings

Quality criteria, in particular, the top criteria such as sense consistency and logical cohesion, enjoy broad consensus. These findings are to be welcomed and suggest that interpreters have a sound understanding of users' needs and expectations.

Since Bühler's cornerstone study was published in 1986 attitudes have changed a great deal and some aspects such as professional attire have become more controversial or even publicly side-lined due to political correctness. Researchers may hence be hesitant to investigate these criteria. Issues such as this do not mean that we ought not to tackle the issue of appearance, however: it is a question of how it ought to be dealt with (i.e. not split along gender lines), rather than whether it ought to be dealt with. As previously mentioned, some of Bühler's original criteria are also now likely to be couched in more modern terms: for instance, 'stamina' is preferred over 'endurance'.

It is interesting to note clear differences between more and less experienced users of interpretation services. Moser highlights that delegates with vast experience of interpreting services clearly prioritise content over other criteria (p. 157).

The results suggest that terminology is important to readers. In order for interpreters to be able to use appropriate terminology, they first and foremost need to understand how the processes they are talking about actually work. Herbert points out that an interpreter facing a technical discussion on atomic weapons, for example, must at least have a basic grasp of nuclear physics (Herbert, 1952, p. 21). It is understandable that interpreters may struggle to understand every domain and thus appropriately use the necessary terminology. One way in which to overcome this would be by interpreters focusing on developing specific fields of interest or specialisations. This is already the case for some interpreters but could be further developed.

Although content-related criteria are clearly of primordial importance, the way in which the information is conveyed remains very important for listeners. Moser highlights 66 respondents who mention aspects such as the importance of a lively voice, clear enunciation, a natural voice (p. 155). Immediately thereafter Moser comments on a spontaneous reference made to the interpretation reflecting the feeling of the original, thus suggesting a link between these two aspects. In short, a pleasant voice (which hopefully sounds natural, lively and is easy to understand) encourages users to trust interpreters. Interpreters thus ought to remember that voice is important for their clients. As Herbert (1952, p. 60) sums up, 'an interpreter with a beautiful voice gets the benefit of the doubt', whereas an interpreter with a voice which is jerky, too fast, too slow or too modulated is quite simply a pain to listen to.

Voice is closely linked to fluency: it is difficult to provide a fluent delivery if there are major vocal problems and vice versa. Indeed, trying to provide a fluent delivery will likely ease the strain on the voice and make one's voice sound more natural.

Some criteria could benefit from being expressed in a more nuanced way: when asking about how important a native accent is, it first needs to be established if an interpreter's A or B language is being discussed. If an interpreter is working into their A language it makes sense to expect them to have a native accent. If an interpreter is working into a B language, however, when a B language is by definition not a native language, is this criterion not out of the question?

There has also been a backlash against the expectation that interpreters should only work into their A language: Brian Harris, for instance, mentions that he insists his students at the University of Ottawa also interpret into their second (B) language (1990, p. 116). Yet Harris also acknowledges that listening to an interpreter speak in their B language is not as pleasant for the delegates (p. 177). In short, quality suffers when interpreters go against AIIC guidelines and interpret into a language which is not their mother tongue.

Were examples provided along with individual criteria this could also help users' better grasp what the researchers are asking about and ensure that what one understands by 'logical cohesion', for instance, is broadly agreed on.

5.3 Conclusion
Achieving quality: listeners' participation in training

First, survey design proved a significant challenge when trying to pool data in order to provide a succinct, easy-to-read overview. There are so many different types of studies which makes providing tabular data difficult – to name just a few, empirical, questionnaire-based; in-depth interviews; open-ended questionnaires, etc. Moser's survey design (1996), for instance, focusses on inferring information. 'Can an ideal interpreter be interfered from the replies [of different user groups]' (p. 146) is one of his leading questions. Seeing as there is a relative lack of information on users' expectations and Moser does indeed study users' expectations, it seems a pity not to include this study. The non-linear nature of many of the responses, however, is problematic when trying to incorporate the information into one table which is easy to read.

On a related note, impartiality can also arise depending on the type of survey: if researchers survey fellow colleagues or friends in interviews it seems likely that respondents may shy away from bringing up certain issues, especially if they fear this would reflect poorly on them. This is true for any professional but it seems to be even more so the case for interpreting which remains a field where professional ties are close, to say the least.

Second, there still remains relatively little information about how the type of conference affects users' needs, as demonstrated by Kurz (1993) and Moser (1996) this. In particular, there is scant research on how interpreters' feel about working in different environments and how they adapt their services. This remains true despite a significant amount of recent research in interpreting studies which clearly emphasises what Collados Ais and Garcia Becerra call 'quality under the circumstances' ('Quality', 2015, p. 376). This builds on Moser-Mercer's idea of 'optimum quality' (1996) and or Pöchhacker's term (1994), or even Herbert's notion of 'adapting to circumstances' (1952, p. 85).

Questions hence need to be asked about what context-based quality means for empirical research and how what this means for listeners can be effectively investigated. How might information about different expectations in different circumstances be collected in order to be able to compare it? Analysing how

different groups of people – not individuals, since we want to focus on overall trends and recurrent patterns in order to collect data which can substantiate well-founded reasoning – perceive interpreting and what they believe constitutes a good interpretation can enable us to gain a better understanding of their needs. This deeper understanding can, in turn, allow practitioners to tailor interpreting training first and foremost in order to adapt their skills and provide a suitable service which fits users' needs.

Third, lack of data more generally is a big issue: although different studies dating from Bühler's original study in 1986 to the present day can be compared, more data, in particular, more current data would be welcome. On the institutional market, more surveys of delegates' experiences would be welcome and help provide researchers with more data. It is important to have a significant among of recent data. As suggested earlier (see introduction), there is notably a trend towards wanting the information to be provided more quickly. It would be interesting to know if delegates are willing to accept more succinct interpretations even if this means completeness is partially comprised.

Fourth, these concerns highlight the need for interpreter training programmes to work even more closely with end users from both the private and institutional market. On the University of Geneva's FTI course, this is indeed the case. Although the main idea behind inviting guest speakers to master classes is for students to be subjected to different speakers, it also provides the opportunity for students to hear about speakers' expectations. In tandem with tutors' feedback, this is clearly beneficial for students. Students are able to familiarise themselves with a number of meeting room delegates who bring their own individual wealth of experience to the table. By integrating all parties into interpreter training – the trainee interpreter, more experienced colleagues, as well as speakers/delegates, interpreting students get a true feel for the demands of the profession.

Overall, researchers in interpreting studies need to move beyond bemoaning the lack of 'one single comprehensive definition' of what quality means when it comes to interpreting (Collados Aid & Garcia Becerra, 2015, p. 368) and move on to acquiring more quantitative data on users' and interpreters' expectations and needs. Survey design still has a long way to go on becoming more rigorous, as Moser-Mercer has repeatedly called for since many years ago (1994, 1996, 1998). By integrating research on the methodological framework and current context-based theories researchers would be able to ask more precise questions which can provide information on how users' needs change. A comparison of users' needs with interpreters' expectations can help interpreters work on tailoring the service they provide as well as ensuring the profession remains at the forefront of our clients' needs in an ever-changing world.

6. References and appendices

AIIC. (1982). *"Practical guide for professional interpreters"*. Geneva: International Association of Conference Interpreters.

AIIC. (2004). *What is a conference interpreter?* Retrieved 02 August 2017, from http://aiic.net/page/1469/what-is-a-conference-interpreter/

AIIC. (2012). *Code of Professional Ethics*. Retrieved 02 August 2017, from http://aiic.net/page/6724

AIIC. (2015). *Professional Standards*. Retrieved 02 August 2017, from http://aiic.net/page/6746

AIIC. (2012). *Who can join AIIC?* Retrieved 02 August 2017, from http://aiic.net/node/2397/who-can-join/

AIIC. (2015). *AIIC announces changes to language classification procedure.* Retrieved 02 August 2017, from http://aiic.net/page/7149/aiic-announces-changes-to-language-classification-procedure/lang/1 language classification

Albl-Mikasa, M. (2013). Developing and Cultivating Expert Interpreter Competence. *The Interpreters' Newsletter,* (17)1, 17-34.

Albl-Mikasa, M. (2012). The importance of being not too earnest: a process- and experience-based model of interpreter competence. In Ahrens, B, Michaela, A, & Sasse, C, *Dolmetschqualität in Praxis, Lehre und Forschung* (pp. 59-92). Tübingen: Narr.

Alexieva, B. (2002). A Typology of Interpreter-Mediated Events. In *The Interpreting Studies Reader (*pp. 218-233.) London: Routledge.

Andres, D. (2015), in *Routledge Encyclopaedia of Interpreting Studies*. (p. 86). Oxon/NY: Routledge.

Bacigalupe, L.A. (2013). Interpretation Quality: From Cognitive Constraints to Market Limitations in Quality. In *Quality in Interpreting: widening the scope Vol 2* (pp. 9-33). Granada: Comares.

Baigorri-Jalón, J. (2015). The History of The Interpreting Profession. *The Routledge Handbook of Interpreting*, Mikkelson, Jourdenais (eds.), 11-29.

Baigorri-Jalón, J. (1999). Conference Interpreting: From Modern Times to Space Technology. *Interpreting,* 4:1, 29-40.

Baigorri Jalón, J. (2006). Anecdotage of an interpreter: Olivia Rossetti Agresti (1875-1960). *Pliegos de Yuste: Revista de cultura y pensamiento europeos* 1(4), 83-90.

Barranco-Droege, R, Maracena Pradas Macías, E & García Becerra, O (Eds), *Quality in Interpreting: widening the scope (Vol. 2)*. Granada: Comares.

Behr, M. (2013). Zur Einbeziehung kommunikationswissenschaftlicher Erkenntnisse bei der Evaluation von Dolmetschleistungen. In *Quality in Interpreting: widening the scope Vol 1* (pp. 85-99). Granada: Comares.

Bowen, D. & Bowen, M. (1985). The Nuremberg Trials: communication through translation. *Meta*, (30)1, 74-77.

Brunton-Smith, Patrick Strugis & George Leckie (2016). *Detecting and understanding interviewer effects on survey data using a cross-classified mixed-effects location scale method (working paper)*. National Centre for Research Methods. Retrieved 02 August 2017, from http://eprints.ncrm.ac.uk/3772/1/NCRM_workingpaper%202%2016.pdf

Bühler, H. (1986). Linguistic (semantic) and extralinguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua* 5 (4), 231-235.

Chiaro, D. & Nocella, G. (2004). Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the World Wide Web. *Meta*, 49 (2): 278-293

Collados Aís, Á. (1998/2002). Quality assessment in simultaneous interpreting: The importance of nonverbal communication. In *The Interpreting Studies Reader*. London/New York, Routledge, 327-336.

Collados Aís, Ángela; García-Becerra, Olalla. (2015). 'Quality'. Mikkelson, Holly; Jourdenais, Renee (eds.), *The Routledge Handbook of Interpreting*. NY/London: Routledge, 368-383.

Dollerup, C. and Lindegaard, A. (eds) *Teaching Translation and Interpreting 2. Insights, aims, visions*, Amsterdam/Philidelphia: John Benjamins.

Donovan, C. (2005). Teaching simultaenous intepretation into B: a challenge for responsible interpreter training. In *Communication and Cognition. Monographies*. 38 (1-2):147-166.

Dummet, M. (2003). Grammar and style. London: Duckworth.

European Master in Conference Interpreting. (2015). *EMCI Core Curriculum.* Retrieved 02 August 2017, from http://www.emcinterpreting.org/?q=node/13

European Master in Conference Interpreting. (2015). *EMCI Consortium Quality Assurance Standard*s. Retrieved 02 August 2017, from http://www.emcinterpreting.org/?q=system/files/EMCIQA%20_Standards_01042015.pdf

European Union. (2017). Becoming a Freelance Interpreter with the European Institutions. Retrieved 02 August 2017, from http://europa.eu/interpretation/accreditation_en.htm

European Union. (2013). EU marking criteria. Retrieved 02 August 2017, from http://ec.europa.eu/dgs/scic/docs/become_an_interpreter/marking_criteria_final_2013_long_version.pdf

Evans, L. (2010). *The Management and Control of Quality.* Ohio: South Western Publishing.

Gambier, Y, Gile, D, Taylor, CH (Eds). (1997). *Conference Interpreting: Current Trends in Research,* John Benjamins.

Gerver, D. (1974). The Effects of Noise on the Performance of Simultaneous Interpreters: Accuracy of Performance. *Acta Psychologica* 38, 159-167.

Gile, D (2001) Quality assessment in conference interpreting: Methodological issues. In Ángela Collados Aís, Maria Manuela Fernández Sánchez and Daniel Gile (p. 109-123). *La evaluación de la calidad en interpretación: Investigación.* Grandada: Editorial Comares.

Grbić, N (2008) *Kommunaldolmetschen: Probleme - Perspektive - Potenziale.* Berlin: Frank & Timme.

Guo, Y. (2013). Towards a model for evaluating meaning transfer quality in consecutive interpreting. In *Quality in Interpreting: widening the scope Vol 1* (pp. 57-84). Granada: Comares.

Gyde Hansen, Andrew Chesterman, Heidrun Gerzymisch-Arbogast. (2008). *Efforts and models in interpreting and translation research: a tribute to Daniel Gile.* Amsterdam & Philadelphia: John Benjamins.

Harris, B. (1990). 'Norms in Interpretation'. *Target* 2:1 (1990), 115-119.

Harvard College (2017). *How to Write a Comparative Analysis.* Retrieved 02 August 2017, from http://writingcenter.fas.harvard.edu/pages/how-write-comparative-analysis

Herbert, J. (1952). *The Interpreter's Handbook: How to Become a Conference Interpreter.* Geneva: Georg.

Hermann, A. (1956/2002). Interpreting In Antiquity. In *Routledge Encyclopedia of Interpreting Studies*. (p. 84). Oxon/NY: Routledge.

Keiser, W. (1999). Histoire de l'AIIC. *Interpreting 4* (1), 81-95.

Kichhoff (1976). Simultaneous interpreting: interdependence of variables in the interpreting process, interpreting models and interpreting strategies (trans. David Sawyer). In *The Interpreting Studies Reader. London/New York, Routledge, 327-336.*

Kremer (2007), A brief reflection on interpretation and professionalism. *AIIC.* Retrieved July 22, 2016, from http://aiic.net/page/2575/a-brief-reflection-on-interpretation-and-professionalism/lang/1

Kurz, I. (1993). Conference interpretation: Expectations of different user groups. *The Interpreters' Newsletter*, 5, 13–21.

Liu, M., Schallert, D. L. & Carroll, P. J. (2004). Working memory and expertise in simultaneous interpreting. *Interpreting*, 6 (1): 19–42.

Macdonald, P. (2013). It Don't Mean a Thing... Simultaneous Interpretation Quality and User Satisfaction. *The Interpreters' Newsletter*, 18.

Moser, P. (1996). Expectations of users of conference interpretation. *Interpreting,* 1 (2): 145-178.

Moser-Mercer, B. (2008). Constructing Quality. In Gyde Hansen, Andrew Chesterman, Heidrun Gerzymisch-Arbogast (pp. 143-156). *Efforts and models in interpreting and translation research: a tribute to Daniel Gile.* Amsterdam & Philadelphia: John Benjamins.

Moser-Mercer, B. (1998). Measuring Quality in Interpreting. In *Evaluating an Interpreter's Performance* (pp. 39-49). Lódź: Centre For Modern Translation and Interpretation Studies.

Moser-Mercer, B. (1996). Quality in Interpreting: some Methodological Issues. *The Interpreters' Newsletter,* 7, 43-55.

Moser-Mercer, B. (1994). Training and Research: The Foundation for Conference Interpretation. *ATA Chronicle* 23 (4), 14-15.

Opdenhoff, J.-H. (2013). Interpreting quality in light of directionality: A study on the interpreter's perspective. In *Quality in Interpreting: widening the scope Vol 1* (pp. 57-84). Granada: Comares.

Paneth, Eva (2012), An Investigation into Conference Interpreting. In *The Interpreting Studies Reader (*pp. 31-40.) London: Routledge.

Pignatraro, C. & Velardi, S. (2013). The quest for quality assessment criteria in media interpreting. In *Quality in Interpreting: widening the scope Vol 1* (pp. 57-84). Granada: Comares.

Pöchhacker, F. (2012). Qualität die man versteht: ein funktional-kognitiver Ansatz. In Ahrens, B, Michaela, A, & Sasse, C, *Dolmetschqualität in Praxis, Lehre und Forschung* (pp. 19-31). Tübingen: Narr.

Pöchhacker, F. (2010). "The Role of Research in Interpreter Education". *Translation & Interpreting. Vol 2*

(1). 1-10.

Pöchhacker, F. & Zwischenberger, Co. (2008). Survey on quality and role: conference interpreters' expectations and self-perceptions. *AIIC*. Retrieved 02 August 2017, from http://aiic.net/page/3405/survey-on-quality-and-role-conference-interpreters-expectations-and-self-perceptions/lang/1

Pöchhacker, F. (2005). Quality Research Revisited. *The Interpreters Newsletter (*13).143-166.

Pöchhacker, F. (1994). Quality assurance in simultaenous interpreting. In *Teaching Translation and Interpreting 2. Insights, aims, visions*, Amsterdam/Philidelphia: John Benjamins, pp. 233-242.

Salevsky, H. 1982. "Teoreticheskie problemi klassifikatzii vidov perevoda". (Theoretical Problems of the Classification of Types of Translation), *Fremdsprachen* 26 (2): 80-6.

Sawyer, D. (2004). *Fundamental Aspects of Interpreter Education: Curriculum and Assessment.* Amsterdam & Philadelphia: John Benjamins.

Sawyer & Roy (2015) 'Education'. In *Routledge Encyclopedia of Interpreting Studies*. (p. 128). Oxon/NY: Routledge.

Schlesinger et al. (1997.) Quality in Simultaneous Interpreting. In Gambier, Y, Gile, D, Taylor, CH (Eds), *Conference Interpreting: Current Trends in Research,* John Benjamins.

Schlesinger, M. (1989). Extending the theory of translation to interpretation: Norms as a case in point. *Target* 1:1, 111-115.

Seleskovitch, D. (2015). *L'interprète dans les conférences internationales : problèmes de langage et de communication.* Paris: Lettres Modernes Minard.

Seleskovitch, D. (1977). Take care of the sense and the sounds will take care of themselves or why interpreting is not tantamount to translating languages. *The Incorporated Linguist,* 15: 7-33.

Stavrakaki S., Megari K., Kosmidis M.H., Apostolidou M., Takou E. (2012). Working memory and verbal fluency in simultaneous interpreters. *Journal of Clinical and Experimental Neuropsychology.* 34 (6), 624-633.

Tiselius, E. (2013). Expertise without Deliberate Practice? The Case of Simultaneous Intepreters. *The Interpreters' Newsletter,* 18 -.

Tiselius, E. (2013). Research on quality in interpreting. *AIIC.* Retrieved 02 August 2017, from http://aiic.net/page/6318/research-on-quality-in-interpreting/lang/1

United Nations. Examinations Interpreters. Retrieved 02 February 2017, from https://languagecareers.un.org/content/examinations-interpreters

Vuorikoski, A.-R. (2012). Fine-tuning SI Quality Criteria: Could Speech Act Theory be of any Use? in Kellett-Bidoli, C.J (Ed.), *Interpreting across Genres: Multiple Research Perspectives,* Trieste: Edizioni Università di Trieste, 152-170.

Zwischenberger, C. (2010). Quality criteria in simultaneous interpreting: An international vs. a national view. *The Interpreters' Newsletter,* 15. -

Zwischenberger, C. (2009). Conference interpreters and their self-representation: A worldwide web-based survey. *Translation and Interpreting Studies* 4 (2), 239-253.