AUTHOR:
Laurent Roger Azambou Ndgongmo[1, 2]
Sandrine Tatsabong[1, 3]
Séraphine Dougophe[1, 4]

AFFILIATION:
[1]Advanced School of Translators and Interpreters (ASTI), University of Buea, Cameroon
[2]Email: azamboulaurent@yahoo.fr
[3]Email: sandroutchoup@gmail.com
[4]Email: sdougophe@yahoo.fr

A STOCK-TAKING OF THE INTERPRETATION EVALUATION APPROACH AT THE ADVANCED SCHOOL OF TRANSLATORS AND INTERPRETERS (ASTI)

ABSTRACT

This article examines the evaluation methods utilised at the Advanced School of Translators and Interpreters (ASTI) of the University of Buea in Cameroon. The assessment of interpreting trainees has garnered significant attention in conference interpreting research because it is crucial for developing competent interpreters who are capable of navigating the complexities of multilingual communication. As the field of interpreting continues to evolve, it is crucial to evaluate the assessment methods used to gauge the proficiency of trainees. This endeavour documents and appraises the interpretation assessment at ASTI. By reflecting on questionnaires completed by both permanent and visiting lecturers, it seeks to review the assessment criteria applied for the marking of core subjects, namely simultaneous interpreting, consecutive interpreting and sight translation, with the aim of shedding light on the specificities of scoring operationalisation, its possible lacunae and opportunities for the establishment of a more objective system. This work, first, reviews the current assessment methods at ASTI. Then, it points out the strengths and weaknesses of these methods. Finally, it identifies future directions that could help overcome the challenges and ensure greater fairness and objectivity in the assessment process.

Keywords: *interpretation, assessment, interpreting competence, scoring, objectivity*

RÉSUMÉ

Cet article passe en revue l'approche d'évaluation utilisée à l'École supérieure de traducteurs et interprètes (ASTI) de l'Université de Buéa au Cameroun. L'évaluation des étudiants en interprétation est devenue un sujet d'intérêt croissant dans la recherche en interprétation de conférence. Elle joue un rôle essentiel dans la formation d'interprètes compétents capables de répondre aux exigences complexes de la communication multilingue. Alors que le domaine de l'interprétation continue d'évoluer, il est crucial d'apprécier les méthodes d'évaluation utilisées pour mesurer la compétence des apprenants. Ainsi, cette recherche documente et fait le point sur l'évaluation de l'interprétation à l'ASTI. En se basant

sur des questionnaires renseignés par des enseignants permanents et à temps partiel, elle passe en revue les critères appliqués pour l'évaluation des matières fondamentales, à savoir l'interprétation simultanée, l'interprétation consécutive et la traduction à vue, dans le but de mettre en lumière les spécificités de la notation, ses éventuelles lacunes et les opportunités de mise en place d'un système plus objectif. Cet article examine d'abord les méthodes d'évaluation actuelles à l'ASTI, puis souligne les forces et les faiblesses desdites méthodes. Enfin, il identifie les orientations futures qui pourraient contribuer à surmonter les défis et garantir une plus grande impartialité et l'objectivité dans le processus d'évaluation.

Mots-clés: *interprétation, évaluation, compétence d'interprétation, notation, objectivité*

1. INTRODUCTION

To ensure high quality language interpreting services, it is crucial to assess interpreter trainees' competency at different stages of the training process. Assessment also informs decision-making at professional and education levels, for certification, before trainees earn degrees and start working as interpreters.

According to Han (2021), the enterprise of interpreter testing assessment (ITA) seems to have drawn far less attention from language testers than it rightfully deserves, possibly mainly because the developmental trajectory of ITA has not been at the core of the research interests of language assessors/raters working in the well-researched area of language assessment. An explanation could be that ITA is generally mainly shaped by forces originating from the interpreting profession. Such forces may vary across different regions, reflect social, cultural, educational, historical and even institutional differences and be related to a peculiar social situation, such as the global mobility of goods, services, technologies, ideas and people. A case in point would be interpreting as a reaction to waves of immigration of skilled labour, refugees, asylum seekers and so on into countries such as Australia, Canada and the United Kingdom, which gives rise to social challenges such as enabling equitable access to legal, medical and other public services (Han 2021). Chen and Han (2021) refer to developments in China to explain that testing and assessment practice received a major boost in the 2000s around admission testing, formative assessment (self and peer assessment) and summative assessment at the end of a training programme (Su 2019). The results that ensued play a crucial role in informing relevant stakeholders (interpreting clients, practitioners, certifiers, researchers and educators) about matters related to a wide array of contextual decision-making, namely admission into and completion of a degree programme, hiring, certification and so on (Li et al. 2022).

In the case of the African continent, the need for interpreters can be associated with the multiple international organisations that have historically been present on this naturally richly endowed continent. Before globalisation and a cohort of bilateral and international conferences organised throughout the continent, history will remember that this continent has been attracting people from different linguistic backgrounds as a consequence of the drive to exploit its rich natural resources and the ensuing crises. A case in point is the Berlin Conference of 1884–1885, which was staged to share the 'African cake', which could not have been done without carefully selected interpreters. The ensuing colonisation left many countries using the language of the coloniser. In a peculiar turn, Cameroon inherited two languages, namely English and French, at the same time.

These two languages became the languages of administration and education, while over 200 local Cameroonian languages were relegated to secondary, less vital functions. Therefore, after independence, the need for conference interpreters arose immediately. After using

mainly European interpreters for few decades, the need for interpreters led to the creation of the Advanced School of Translators and Interpreters (ASTI) in 1985, to train translators and interpreters for the civil service. Historically, ITA in Cameroon was a decisive step for trainees to obtain postgraduate diplomas before they were absorbed into the civil service. This stage lasted about a decade, until the needs of the civil service of that time were saturated. At this turning point, ASTI started training for the private market at national and international levels. The next historical turning point would be in 2007, when the BMP (Bachelor, Master's, PhD) system was adopted by higher education in the entire Central African subregion, for the sake of the professionalisation of education. More recently, in Cameroon, a five-year recruitment drive for the public service was initiated for graduates of ASTI and other national institutions. All these turns required specific forms of assessment.

Despite all the turning points described above, the situation is what Angelelli (2005, 23) describes as a 'closed circle':

> The circle is closed by the presence of the schools of interpreting, in which the practice and the professional associations have an impact upon the education of interpreters. In other words, the crucial relationship arising from the interaction of theory and research (which normally would inform practice by helping a field move forward) and practice (which normally would inform theory and research by setting new directions in which the field needs to move) is compromised and almost non-existent.

Consequently, publications on interpreting are still at the embryonic stage and, in the case of Cameroon, ITA practice is not well documented.

Whereas much of the literature describes how testing and assessment of interpreting has been conducted in different countries, relatively little is known about recent practice and research on this topic in Cameroon. This dearth of information justifies the choice of an exploratory approach to the study of ITA at ASTI. This work, therefore, describes the actual practices of interpreting tests and assessment, by focusing on how raters evaluate interpreting tasks in the real education context of the Interpretation Division. We start by reviewing the construct of interpreting competence, which is fundamental for assessment, as discussed by theoretical and empirical research. The next step is discussing assessment in interpreting, specifically automatic assessment and rater-mediated assessment. This leads to a discussion of the methodology used in this study, followed by a presentation and discussion of results and the way forward, which involves the design of and experimentation with a contextual scoring tool.

2. LITERATURE REVIEW

In this section, we discuss interpreting competence as a multifaceted construct and the major types of interpreting assessment.

Interpreting competence is considered to be a construct that encompasses cognitive, linguistic, cultural and even ethical dimensions and is at the core of assessment. To understand this construct, research on interpreting competence has various foci and can be broadly subsumed under theoretical endeavours and empirical studies; the latter collecting field data to see how ITA is manifested in the field, whether for educational, certification or research purposes.

Theoretically, researchers have tried to understand cognitive processes and elements and traits underlying interpreting competence. Han (2022) subsumes such attempts under three different angles, namely the *cognitive processing* approach, the *multicomponential* approach and the *interactional* approach. He presents Gerver (1975), Gile (2009), Moser (1978) and

Setton (1999) as proponents of the *cognitive processing approach*, to help researchers understand interpreting competence by modelling essential linguistic-cognitive operations and processes involved in simultaneous interpreting. A familiar model for interpreting processes is Gile's effort model (2009), which conceptualises the different kinds of efforts involved in interpreting.

With the *multicomponential* approach (Pöchhacker 2000; Wang 2007), interpreting competence is grasped by identifying core traits and characteristics that interpreters must possess. *Bilingual language competence* and *linguistic transfer competence* are considered important traits of a competent interpreter, alongside several other components, such as subject matter, topical knowledge and professionalism.

The *interactional approach* (Han 2015, Wang et al. 2020) combines cognitive process modelling and multicomponential modelling. Therefore, interpreting competence is understood by emphasising the role of (meta) cognitive processes that regulate linguistic-cognitive, external, task-specific characteristics. In other words, what is happening inside the interpreter is associated with their response to external influences from the context in which interpreting is occurring. These theoretical attempts to understand interpreting competence are coupled with empirical endeavours.

Empirical studies on interpreting competence have gathered data on the development and progression of interpreting competence, on acquisition of interpreting strategies and the cognitive processes involved. A meta-analytical study by Wen and Dong (2019) found that interpreters have an advantage over other bilingual people at controlling both working memory and short-term memory spans. To understand the cognitive load faced by an interpreter, techniques such as digital pen recording, eye tracking and event-related potential have been applied (e.g., Chen 2020; Tiselius & Sneed 2020). Tiselius and Sneed (2020) found no significant difference in gaze patterns of experienced and inexperienced interpreters, and that interpreting into L2 was less demanding in terms of cognitive load than into L1. Chen (2020) examined cognitive processes during consecutive interpreting with notes through eye-tracking and determined that L2 into L1 interpreting was less cognitively demanding than the reverse direction.

Researchers also examined interpreting competence through the lens of expert performance (Ericsson 2000; Moser-Mercer 2008). The development of this competence was studied by Albl-Mikasa (2013) and Cai et al. (2015), who analysed interviews with professional interpreters to identify how they acquire critical competences and strategies over time. Han (2022) refers to another group of researchers who focused on interpreters' use of strategies and who report that strategy use may ease cognitive burden and improve delivery (e.g., Bartłomiejczyk 2006, Li 2013).

Han (2015) applied an interactional approach to construct the components of interpreting ability, which include knowledge of languages, interpreting strategies, topical knowledge and metacognitive process. Setton and Dawrant (2016) identified four core elements as interpreting competence: bilingual language proficiency, knowledge, skills and professionalism.

Generally speaking, all the perspectives for understanding interpreting competence point out that competence has to do with more than just having major skills, such as bilingual and transfer skills. In addition to these skills, another range of contextual, situational and world knowledge skills must be considered when trainees are assessed, since clearly identifying and agreeing on competencies to be assessed is of crucial importance for the assessment process.

Assessment serves different purposes and refers to the screening of an individual's achievement (Amato & Mack 2022, 464). Research has shed light on a diverse range of topics relating to ITA practices, of which Han (2022) identifies three major practices: first, specificities of interpreting assessment (modes, language combinations, directionality); second, assessment design (assessment task and assessment criteria); and third, scoring and rater training (scoring methods, scoring operationalisation, rater selection and recruitment, training and calibration). In addition to self-assessment by students in a pedagogical environment, empirical research has investigated human-rater-mediated assessment of interpreting (how raters use rubrics or descriptors, allotment of weighting to different assessment criteria in an analytical scale, measuring raters' effects on the assessment, etc.) and a range of matters relating to automatic assessment of interpreting.

Automatic assessment refers to the use of automated scoring engines to assess the quality of an interpretation, with little or no human judgement; this has recently become one of the hotspots of research in computer-assisted language learning (Li et al. 2022). Efforts to research automatic assessment revolve around the following: using an acoustic analysis software to objectively measure prosodic features such as fluency or intonation (Li et al. 2022) and using corpus/computational linguistics tools to assess the overall quality of speech. Automatic assessment makes use of tools such as the Interpreting Proficiency Test, SIQA (Standards in Interpreting Quality Assessment), and Interpreting Quality Rating. The Interpreting Proficiency Test is a standardised tool that can be used to measure a trainee's performance in consecutive and simultaneous interpreting. It is a set of scoring rubrics that provides scores for different sections of a test, such as vocabulary, comprehension, syntax and verbal expression. It has been found to have high inter-rater reliability and validity (O'Brien, Li & Tawil 2019). SIQA is a framework that was developed by the National Council on Interpreting in Health Care (NCIHC). It is a software-based tool that objectively evaluates simultaneous interpreting quality by measuring aspects such as delay, fluency, omissions and additions by comparing the interpreter's output to the source material. The Interpreting Quality Rating was also developed by NCIHC; it is based on SIQA and is a scale that assesses the quality of interpreting based on a set of criteria that includes accuracy, completeness and register (Gile 2009). There are also tools that use technology, such as eye tracking and brain imaging, that can provide valuable insights into the cognitive processes involved in interpreting and can help identify areas in need of improvement.

However, there are weaknesses embedded in the use of objective tools. Objective assessment tools, even if they are standardised, have disadvantages. They can be insensitive to the nuances of interpreting, by not being able to distinguish between a minor error in accuracy and a major error in accuracy. They can also be expensive to develop and administer in contexts such as those where there is still a struggle to build a certain standard of laboratories that are up to the standards for workshops, which makes it difficult to adapt them to different interpreting settings. For example, a tool designed to assess medical interpreting might not be appropriate for assessing legal interpreting. This may explain the prevalence of human-rated approaches to assessment and the tendency to mix this approach with rater-mediated assessment. Han et al. (2024) investigated the scoring process in the assessment of interpreting based on eye tracking and retrospective verbalisation of raters.

Rater-mediated assessment allows for the evaluation of nuanced aspects of interpreting performance that are difficult to capture automatically. Raters play a critical role in the assessment of interpreting and a large proportion of the research is centred on raters (Han

2022). This research sheds light on rater effects (e.g., their severity, leniency and accuracy, the halo effect), behaviour and characteristics displayed by raters (whether they have received formal training, have taught or practiced interpreting and even their language combination) and rater-generated measurements.

Methods used by raters in interpreting can include self-assessment, peer review and expert assessment (Lee 2019) and feedback from trainers, colleagues, or clients who have worked with the interpreter trainee or listened to and observed them. Raters typically make use of rubrics, that is, scoring guides that outline specific criteria for evaluating different aspects of interpreting performance, such as accuracy, fluency, demeanour and cultural competence. Pöchhacker (2007) assessed the impact of subjective assessments on the self-perceived competence of interpreter trainees. He found that trainees who received positive feedback from subjective assessments were more likely to rate themselves as more competent interpreters than those who received negative feedback. Building on these findings, another study, conducted by Lee (2015), examined the impact of formative assessment on interpreter trainees' self-efficacy beliefs. Lee found that trainees who received frequent feedback and were given opportunities to improve their skills through practice were more likely to develop a sense of confidence in their ability to interpret.

Overall, according to Han and Lu (2021), who assessed interpreting quality, human raters use atomistic scoring, questionnaire-based scoring, multi-method scoring, rubric scoring and ranking. Atomistic scoring is based on error analysis and item scoring. This means that raters focus on specific points (such as omission, addition or *contresens*) to detect errors in the rendition while comparing it to the source text. Despite the process being quite time consuming, the major benefit is the possibility of providing a description of faulty renditions, which is useful for pedagogical purposes such as feedback to students or recommendations to work on avoiding specific types of errors.

Item-based scoring seems to be an enhanced version of error analysis because raters can decide which units they want to assess. González, Gatto and Bichsel (2010) describe how raters applied a dichotomous scale (e.g., correct vs incorrect) to assess trilingual interpreters. Despite being reductionist, item-based scoring increases consistency but concerns remain over the validity of items chosen to be scored.

Raters also use questionnaires composed of checklists of assessment criteria grouped in categories (e.g., fidelity, delivery, expression). Lee (2015) identified 21 aspects of interpreting performance, grouped into three assessment categories (content, form, delivery), that can be used to calculate the final score of an interpretation.

When they use rubric scoring, raters assess interpreters by applying rubric-referenced rating scales. Descriptors are created to capture typical features and characteristics for different levels along a performance continuum (Han & Lu 2021). The method involves ranking, whereby raters compare a number of interpretations according to their overall quality and order them (e.g., from worst to best quality).

Scoring methods that use rater-mediated assessment involve different ways of conceptualising interpreting quality or competence. Much research has focused on psychometric characteristics of rater-assigned scores (scoring reliability, accuracy and severity) as a function of various factors, including assessors' professional or language backgrounds (Han, Hu & Deng 2023).

Rater-mediated assessment is a cornerstone of interpreting evaluation. Arguably, assessing interpreting quality is a complex process in which raters need to interact with several materials (e.g., the source text, the target text and the scoring rubric), attend to various aspects of interpreted rendition (e.g., informational, prosodic and linguistic characteristics) and integrate local judgements into a holistic evaluation that captures interpreting quality best (Gile 1999).

 Limitations of both subjective and objective assessment methods underscore the necessity to constantly revisit the assessment approach in a given context, hence this endeavour to capture the essence of what is happening at ASTI in assessing training and to keep revising the existing approach.

The kind of interpreter training, specificities of the context where the assessment is required, as well as targeted objectives and outcomes, are of utmost importance when an assessment approach is selected. Objective assessment tools could, therefore, be used in conjunction with other assessment tools that would be deemed more 'subjective'. Some common subjective approaches include performance-based evaluation by experts, self-assessment, peer evaluation, observation and feedback and standardised tests. They have the common point of resorting to a human agent. The current work is focused on understanding the processes undergone by that human agent in the case of summative assessment in the context of Cameroon (ASTI), with the entailed challenges and ways forward.

3. METHODOLOGY

The target population of this research was permanent and part-time lecturers in Division II (Interpretation) at ASTI. This group of trainers was chosen for their in-depth knowledge and hands-on experience of the assessment procedures at ASTI and their ability to provide informed feedback, especially trainers who possessed experience of assessing end-of-semester exams, re-sit exams and final exams. To achieve the second objective of this research, which was to sketch out the strengths and weaknesses of the current assessment methods at ASTI, members of the target population – permanent and part-time lecturers of interpretation at ASTI – were asked 64 questions.

Data were collected using a web-based survey that had been crafted to address the main objectives of the paper and that was distributed as a Google Forms questionnaire. The questionnaire was designed to elicit valuable insights into the current assessment system, identify potential areas for improvement, and gather and discuss suggestions for enhancing the overall evaluation process.

The Google Forms questionnaire was sent to a sample of 25 trainers, who were selected based on their expertise of conducting assessments, so as to ensure highly relevant feedback on current practices. The return rate for the questionnaire was 56% (14 responses), indicating a satisfactory level of participation and engagement from the target population.

The questionnaire had a combination of closed-ended and open-ended questions. The closed-ended section included yes or no and multiple-choice questions, thereby providing quantitative data on specific aspects of the interpreter evaluation system. The open-ended questions allowed the respondents to express their opinions and provide detailed insights into various aspects of the assessment process. Upon completion of data collection through the questionnaire, the responses were compiled and analysed according to the main objectives of the paper. Quantitative data obtained from the closed-ended questions were processed

to identify trends and patterns. The qualitative data from the open-ended questions were subjected to thematic analysis to extract key themes and perspectives related to the current assessment system, potential deficiencies and suggested improvements.

As an integral step of our research, we diligently requested ethics clearance from ASTI to ensure the ethical integrity of this scientific endeavour. To uphold ethics standards, the confidentiality and anonymity of the participants were ensured. The data would be used solely for research purposes and were kept confidential. Consent was obtained from the participants before their involvement in the study, and they were informed about the voluntary nature of their participation.

Overall, the methodology employed in this study facilitated a comprehensive examination of the interpreter evaluation system at ASTI and provided valuable insights about potential enhancements and improvements.

4. RESULTS: REVIEWING THE CURRENT ASSESSMENT METHOD AT ASTI

The results of our analysis provide a bird's eye view of the current interpretation evaluation system at ASTI. We start by reporting on the strengths of the collegial approach to assessment and discuss text selection and assessment criteria, as reported by raters from different backgrounds. Then, we report on weaknesses, including lack of consistency in marking criteria and the way forward, that is, the need to design and experiment with a context-fit scoring guide.

4.1 Strengths of the current assessment system

The strengths of the current system are related to the composition of the assessment panel and the assessment process.

At ASTI, assessment is done collegially, with a panel of external and internal examiners composed of young and mature professionals. Results of the survey show that, at ASTI, interpreter trainees are assessed by a blend of (permanent and part-time) external and internal professionals from various backgrounds.

Respondents served on the panel that assessed all exams, from entrance examination to end-of-semester examinations through re-sit examinations and final examinations. They recorded the involvement of permanent lecturers from ASTI and part-time lecturers from other national institutions and foreign institutions, who are all generally professional interpreters (90.9%), as Figure 1 shows.
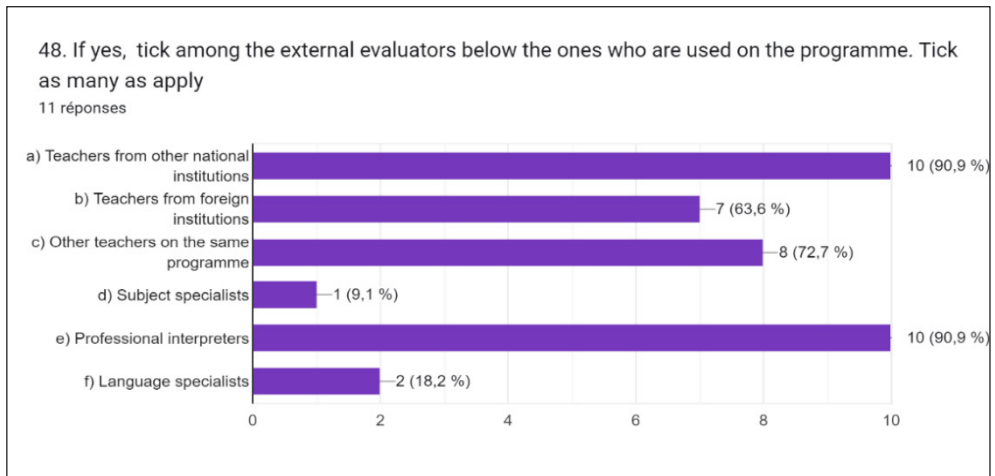
Figure 1:   Background of evaluators

Concerning the backgrounds of evaluators, the data show that all lecturers involved in the interpretation programme at ASTI were generally professional interpreters with some involvement in academic work. Regarding qualifications, a quarter of the respondents were PhD holders, half had postgraduate diplomas in interpretation; others were PhD candidates and holders of MA degrees in conference interpreting.
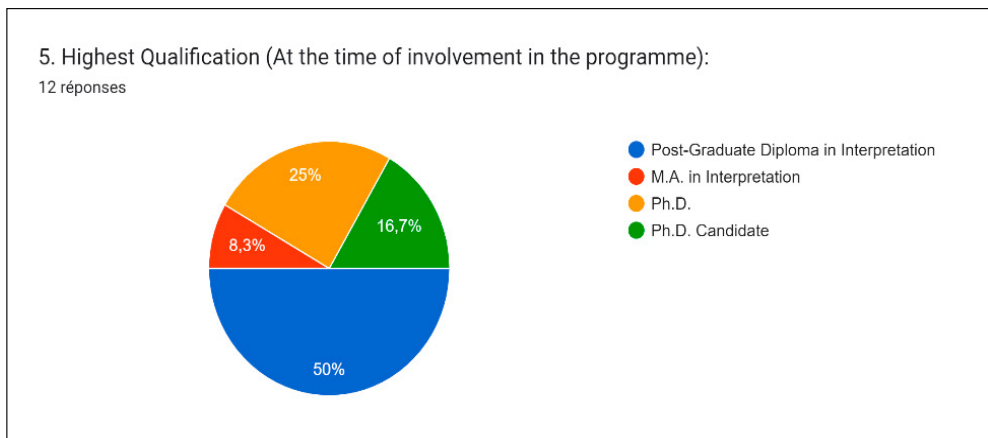

Figure 2:   Qualifications of evaluators

Such a variety in backgrounds can be considered an asset, as it suggests that assessors have good understanding of the requirements of the profession and of the market. This premise could be solidified by ongoing training.

4.2 Evaluators from different backgrounds and generations

The survey reveals that the majority of respondents are well prepared for the task of assessment given their professional/academic backgrounds, but also their continuous capacity building.

Though few respondents (25%) reported having gone through a formal training as interpreter-trainers, they reported having attended seminars focused on teaching interpreting, teaching interpreting practice and assessing interpreting performance; 75% of them reported attending seminars on teaching interpreting, whether in Cameroon or abroad, by ASTI or jointly by ASTI and a foreign institution. All those who reported having attended such seminars (100%) believed this experience contributed to making them better interpreting lecturers and assessors, as shown in Figure 3.
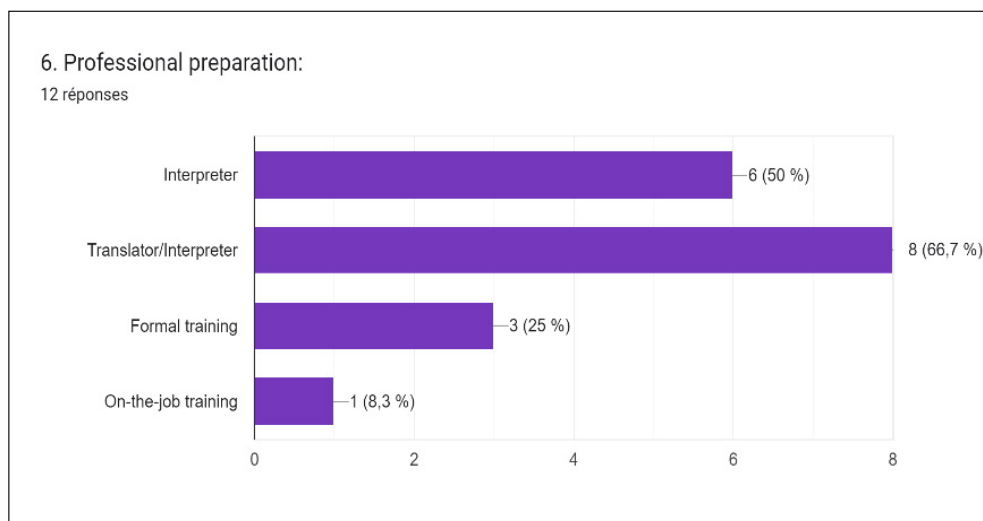

Figure 3:   Professional preparation

Their belief that they were prepared for assessment also derived from their experience as either translators/interpreters (65%), interpreters or professional service providers.

Transgenerational training and transfer of knowledge can also be considered as taking place during the assessments, as a broad range of experienced lecturers were involved, with years of experience ranging from 5 to more than 20 years. The largest group (33.3%) had between 10 and 15 years of experience, followed by the 6–10 years of experience group (25%), the 20+ years of experience group, the 0–5 years of experience group (16, 7%) and, finally, the 15–20 years of experience group (8.3%), as illustrated in Figure 4.
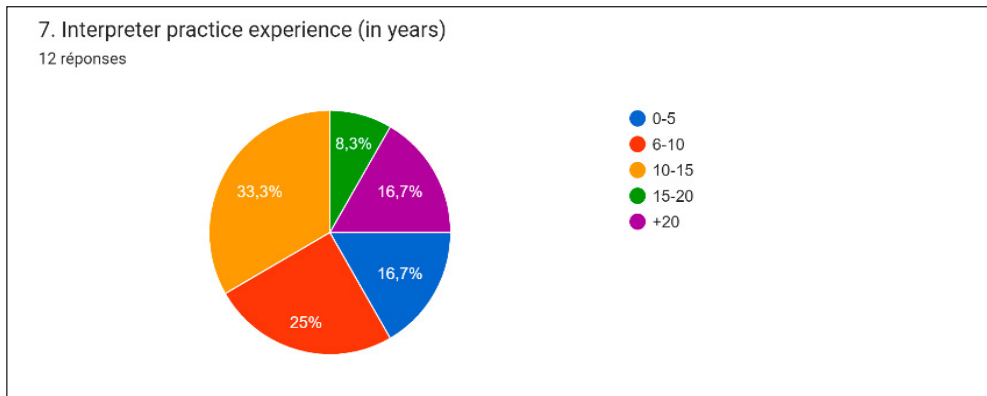
Figure 4:   Interpreters' experience

Figure 4 shows that, having been trained on the job, the practice-experience gap between young professionals (up to 5 years) and the most senior professionals (20+) involved in the assessment process can be seen as an asset. With such synergy, those with the most experience can be considered as the guarantors of the teaching/assessment system, while those with less experience are being groomed and mentored in the process.

This can also be considered as a strong point, as it guarantees an intergenerational transfer of knowledge and constant injection of new life into the programme, while keeping the fundamentals of teaching/assessment, which are the backbone of interpreter training at ASTI. Other positive aspects of the current assessment system at ASTI are linked to the assessment approach.

4.3 Choosing a text collegially

Despite there not seeming to be a strictly established process for choosing texts for assessment, respondents report a clear approach, starting with the panel deciding on texts to use for the assessment, inviting candidates and unfolding the assessment process.

Concerning the material to be interpreted during the exam, respondents explain that evaluators are presented with a varied set of texts proposed by some of them. The texts originate from a variety of sources, for instance, newspaper excerpts (91.7%), materials modified or oralised by a lecturer to serve an evaluation purpose (75%), and even authentic written speeches (83.3%). Authentic video speeches are also used, though to a lesser degree (41.7%), as illustrated in Figure 5.
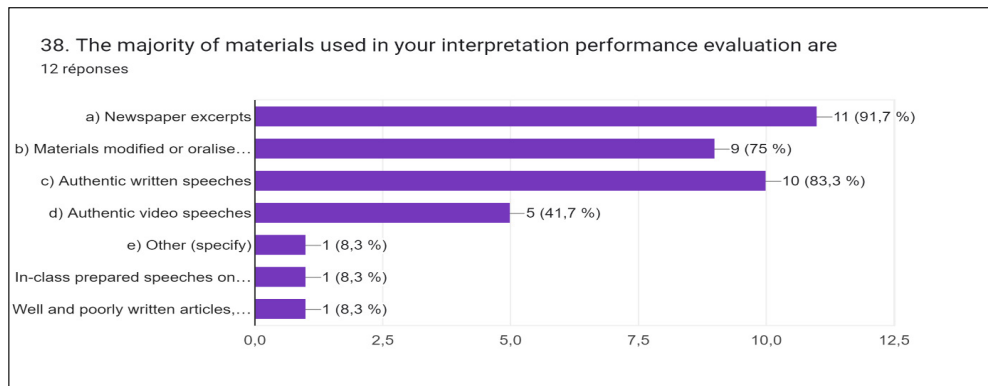
Figure 5:   Evaluation materials

Evaluators are given time to read through the various texts that are proposed. First, they indicate their individual choices, after which the choice of the majority is drawn, with a possibility for argumentation. The fact that exam texts are selected collegially by panel members means that there is a preliminary review of all the texts proposed by lecturers. The final choice is dictated by a collective appraisal of the level of difficulty of the text and a judgement on its appropriateness for the said level and semester (beginners, intermediate or final semesters) before evaluators agree on a set of marking criteria.

This indicates that various sources and types of texts are used to assess trainees. Precedence is given to formal documents that are well drafted, or authentic text actually delivered by speakers. According to the respondents, these texts are chosen with a view to harnessing specific aspects of training, to reflect the types of tasks and difficulties involved in real life interpreting and achieve course objectives, while keeping students' needs in mind.

4.4 Defining assessment criteria

After choosing a text, the panel of evaluators generally agrees on a set of marking criteria. Despite respondents using different terminologies to refer to the criteria applied to assess trainees, we identified five basic, common criteria, though called by different terms:

i.   content, message, fidelity in meaning, comprehension and message, accuracy

ii.   fluency, public speaking skills, language, delivery

iii. vocabulary, language register, language, style/diction

iv.   demeanour, attitude, delivery

v.   general knowledge

There seems to be consensus among those who listed assessment criteria that the meaning/content/comprehension item carries no less than 50% of the final mark – sometimes even 60% – followed by fluency (20%), vocabulary (20%), and demeanour (10%). Overall, "Getting the right meaning of the source utterance" and "Rendering the ST [source text] as faithfully as possible" were the most crucial elements for evaluators. They also underscored the necessity to "Avoid making language mistakes in your interpretation" and "Getting the right words in the

TL [target language]". A follow-up question was raised about the weight respondents attributed to each of the items on their list for the final mark. Once again, there seemed to be general agreement that emphasis should be laid on "Getting and conveying the right meaning".

From the answers it is clear that, though no prior objective or subjective assessment guide is known to either students or lecturers, most respondents are of the view that the current assessment system at ASTI is satisfactory. With the above-mentioned assessment criteria, the current interpretation testing at ASTI was described as satisfactory by 75% of respondents, as very satisfactory by 16.7% and unsatisfactory by 8.3%, not only for assessment purposes, but also for preparing trainees for the next step.

42. According to you, how satisfactory is the current interpretation testing?
12 réponses

- a) Very satisfactory
- b) Very unsatisfactory
- c) Satisfactory
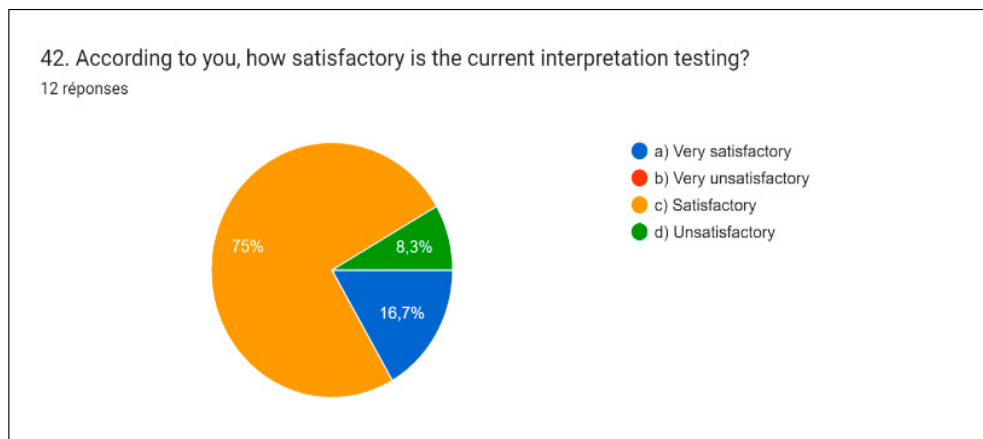- d) Unsatisfactory

75%   8,3%   16,7%

Figure 6:   Level of satisfaction with the current evaluation system

Expressing satisfaction means that the current approach enables evaluators to meet the goals of interpreter assessment with a certain degree of confidence in the process. The selection of materials to be used for assessment and scoring criteria were mainly guided by what the evaluators wished to assess and see in their trainees. Once texts have been selected, there is a session with the trainees before the assessment takes place.

4.5 Preparing the candidates by providing background information

Respondents agreed that, once the panel has been determined, the panel calls in the candidates as a group for a briefing session in preparation for the upcoming assessment exercise. For 66.7% of the respondents, evaluation criteria are "not always" given in advance. Instead, as students are called into the classroom to receive information about how they will be assessed, they are given an opportunity to ask questions and receive guidelines. Background information can also be provided, as can practical advice on how to conduct themselves during exams, though this information is not always provided systematically. Once trainees have been dismissed from the classroom, they are called back one by one to be assessed. After a student's performance, every member of the panel is requested to assess their performance and mark the student. At this level, weaknesses of the obviously subjective assessment approach can be unveiled.

4.6 Weaknesses of the current assessment approach

Respondents were asked to identify some of the major weaknesses of the system currently used at ASTI.

4.6.1 *Lack of consistency in marking criteria and training of trainers*

While being asked whether panel members' evaluations vary (by at least 3 marks), 91.7% of them said yes.
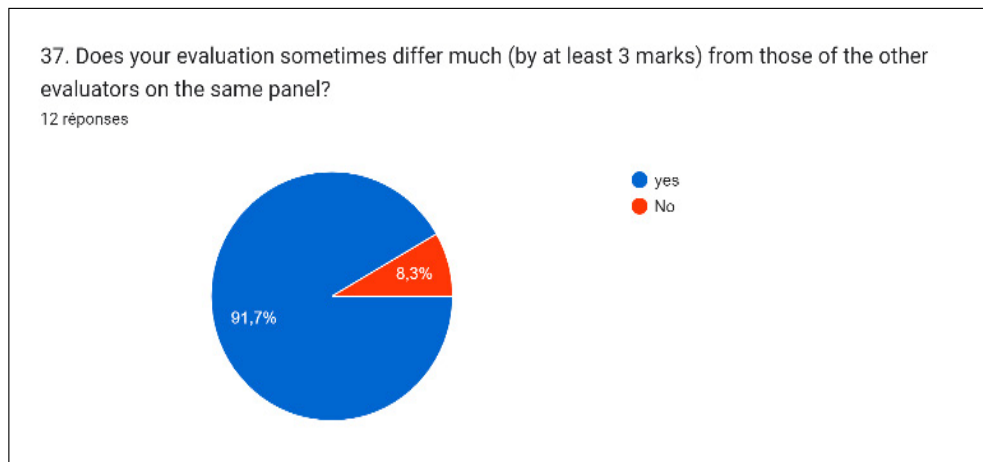

Figure 7:   Gap in marking between evaluators

Though there is always a margin of divergence in the evaluations by panel members, a large number of evaluators who declared that evaluations sometimes differ by more than 3 marks may be illustrative of a discrepancy in the assessment criteria or weight of each criterion as applied by the various evaluators. It may also indicate a lack of consistency in the various approaches adopted by each member of the panel.

Respondents were asked if they had attended an induction seminar before they started teaching on the interpretation programme at ASTI. Although 83.3% of respondents reported attending such a seminar, 16.7% did not; therefore, some lecturers/evaluators started work without receiving prior preparation. The fact that not all lecturers had participated in an induction seminar before starting teaching or assessing indicates that such seminars are not automatically organised for newcomers joining the teaching environment. Lecturers who did not attend induction seminars were also asked if they had observed other lecturers' lessons before starting to teach their own classes. Half of them (50%) did and indicated that it was on their own initiative, and not because school administration had instructed them to do so. The other half did not observe other lecturers teaching before engaging in the activity themselves, giving as reasons that no one asked them to do so (62.5%) and that they did not think it was necessary (25%). In conclusion, there is no requirement for newly recruited lecturers to observe their colleagues teaching before diving into that activity themselves.

When it comes to assessment, the respondents were asked if they had attended seminars on the assessment of interpreting performance after joining interpretation training in ASTI; 66.7% of them replied that they had not.
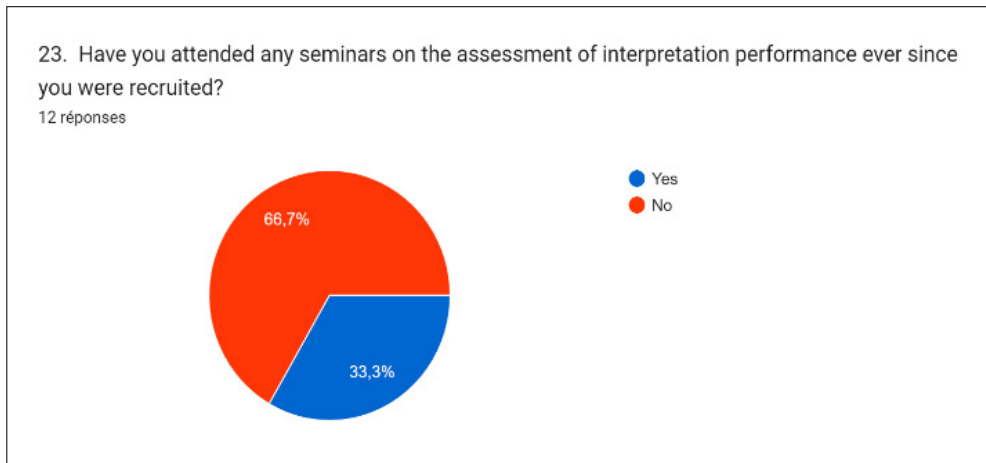
Figure 8:   Participation in training seminars on interpreting assessment

Indeed, most lecturers/evaluators had not attended seminars or training sessions on assessing interpreter trainees' performance. Consequently, there is a lack of grounded knowledge on interpreter performance assessment.

4.6.2 Absence of an assessment guide, prior interpretation

Another aspect that may be a weak link for interpreter assessment at ASTI is related to trainees' knowledge of evaluation criteria. Indeed, the respondents indicated that they do not necessarily provide students with the evaluation criteria in advance. ASTI seemed to lack a marking guide that was known to both students and lecturers (see Figure 9).
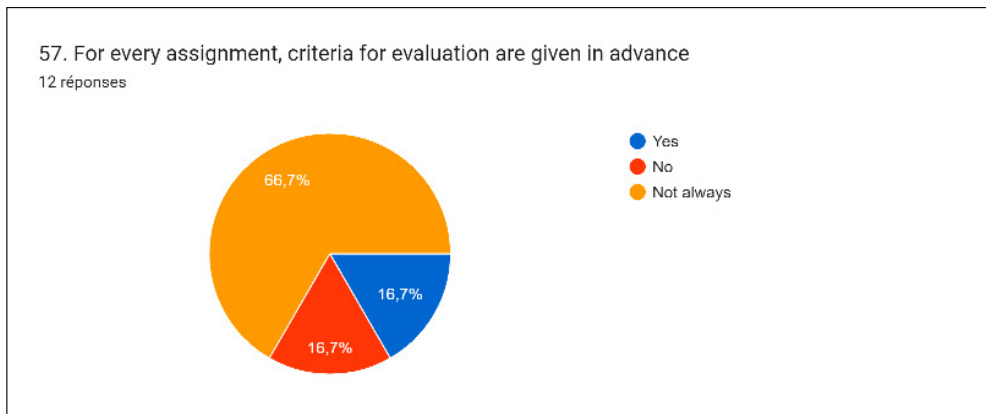

Figure 9:   Prior introduction to evaluation criteria

It was indicated in Section 4.2, where strengths were discussed, that the interpretation programme at ASTI usually brings external evaluators on board for exams. Most of these external evaluators are also professional interpreters or interpretation lecturers; however, their involvement does not seem to be preceded by an induction to help them understand their role.

When asked about the marking criteria they apply to each of the courses they teach, not all respondents provided answers. One might wonder how those who did not have an answer to that question carry out their assessment.

In response to the question whether there was a marking guide known to students and lecturers, few respondents provided answers, some said no and others said that they ignored such a marking guide if it existed. One respondent declared that the marking guide was generally known to lecturers, but that there was a need to explain it to the students too. It could be concluded that there is no common marking guide used by lecturers and, even if such a guide existed, it is only known to lecturers. Hence, students do not know how they are assessed. It is important to develop marking guides for each core course and to share them with students early on in the training, so that students know what is expected of them and how they will be assessed.

In addition, respondents were asked whether they or their colleagues first interpreted the texts or speeches before they used it for student evaluation, and three-quarters of respondents said they did not.

31. Do you or any colleague first interpret the texts or speeches to be proposed to the students for evaluation?
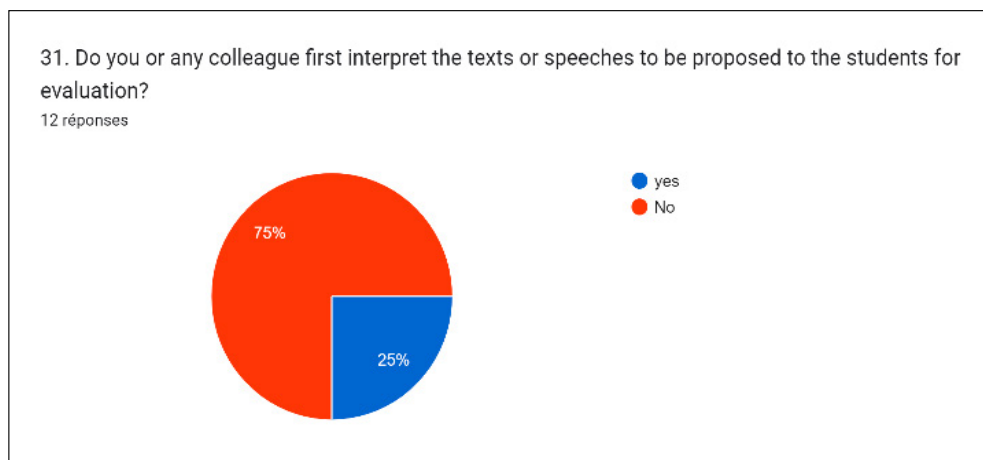
12 réponses

- yes
- No

75%

25%

Figure 10:   Prior interpretation of texts by evaluators

Lecturers/evaluators themselves do not interpret the texts they present to trainees; they only read written texts. Considering the difficulties involved in the transition from a written text to a spoken text, not having lecturers first interpret the text before expecting students to do the exercise is a weakness of the current assessment approach at ASTI. It is worth mentioning that, up to the 2014/2015 academic year, the intake for the ASTI interpretation programme was 10 students. Azambou (2019, 225) observes that "The intake stands at 10, which is reasonable and manageable as enrolment for an interpreter training class". From the 2015/2016 academic year, Azambou reports that "The intake moves to 20 students, without any change in infrastructure or number of staff. A class of 20 students in interpreter training is hardly manageable pedagogically". He considers this mismatch between intake and resource

allocation to pose threat to the programme. Thus, another aspect of the interpretation division at ASTI that can be considered a weakness or as having a significant impact on assessment is the large intake. Indeed, almost 60% of lecturers acknowledged that a large intake has a negative effect on interpreter assessment outcomes.

32. What is the impact of large intake on the interpreter education outcomes in your context?
12 réponses

- a) No impact
- b) Negative
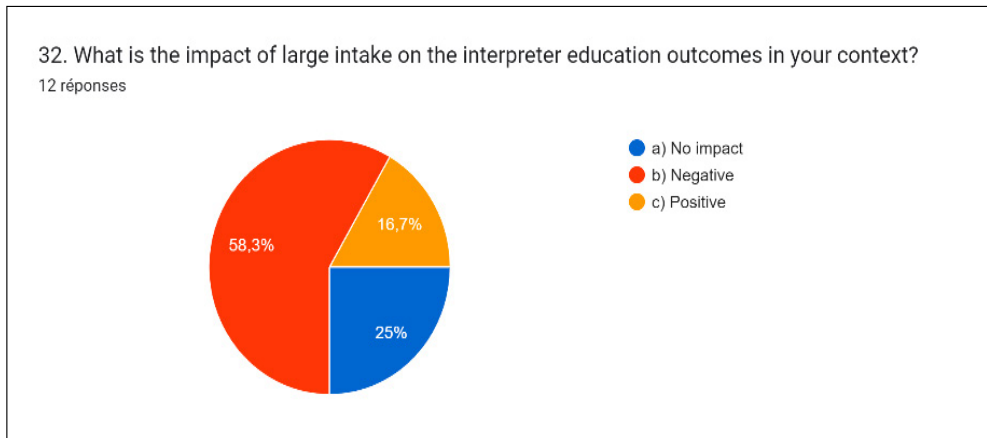- c) Positive

58,3%   16,7%   25%

Figure 11:   Impact of large intake on interpreter training at ASTI

Consequently, many students train in the same lessons and it is difficult to be meticulous with each and every one of them. In the end, the assessment outcomes may not be accurate, as the distribution of assessment time per student is reduced.

4.7 Overcoming the challenges and way forward

Overcoming challenges in assessment requires a systematic development of the capacities of the teaching/assessing team, ensuring that lecturers/evaluators interpret assessment material themselves before the assessment event, and designing a marking list that combines subjective and objective assessment approaches.

4.7.1 Capacity building

Respondents expressed the desire to undergo systematic capacity building for their role as members of the teaching/assessment team of the interpretation division at ASTI. When asked if there was any aspect of their teaching that needed reinforcement or acquisition, most respondents pointed to their pool of teaching exercises (75%), documentation on current trends in interpreting studies and skills development in interpreting performance assessment (50%). They also expressed a need for further training in teaching techniques for specific course components (41%), and class diversity management (33.3%). There is, therefore, a need to identify the range of skills and techniques that will enhance the teaching and evaluation process in the end.

Figure 12 is indicative of a range of skills and techniques that are likely to improve the teaching and evaluation process if training is provided.
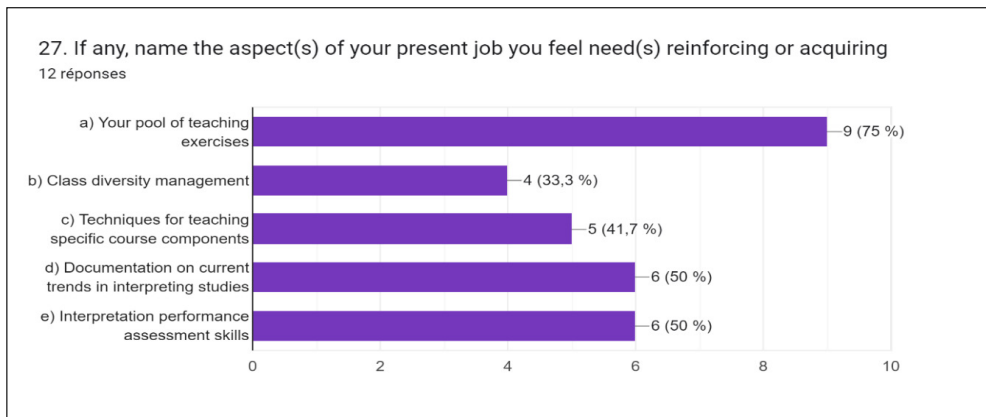
Figure 12:   Capacity building needs

Systematic, continuous training would help to refine marking criteria used collegially, and to clarify the weight of each criterion in the final mark. This would eliminate the apparent lack of consistency in this regard, which is caused by each evaluator deciding the weight to attribute to a given criterion.

One of the major findings of this research is that not all the evaluators work according to predetermined marking criteria and, for those who do, there is a lack of consistency; hence, the need to develop a marking guide.

4.7.2 Designing a marking guide and breaking the closed cycle

Reducing the margin of subjectivity goes hand in hand with the design of a scoring guide, which could, without being prescriptive in nature, serve as guidelines for the whole process of trainee assessment.

When asked for suggestions to enhance the evaluation procedures for student interpreters at ASTI, a respondent said that "it is imperative for teachers to interpret the material themselves before it is proposed to assess students, so that difficulties are better known to them". A scoring guide would help develop common assessment criteria for Division II (Interpretation) and meet the need for predetermined, consensus-based criteria, while also taking into account elements specific to the context of ASTI. As illustrated in Figure 13, respondents agreed (91.7%) that there was a need for such guidelines.
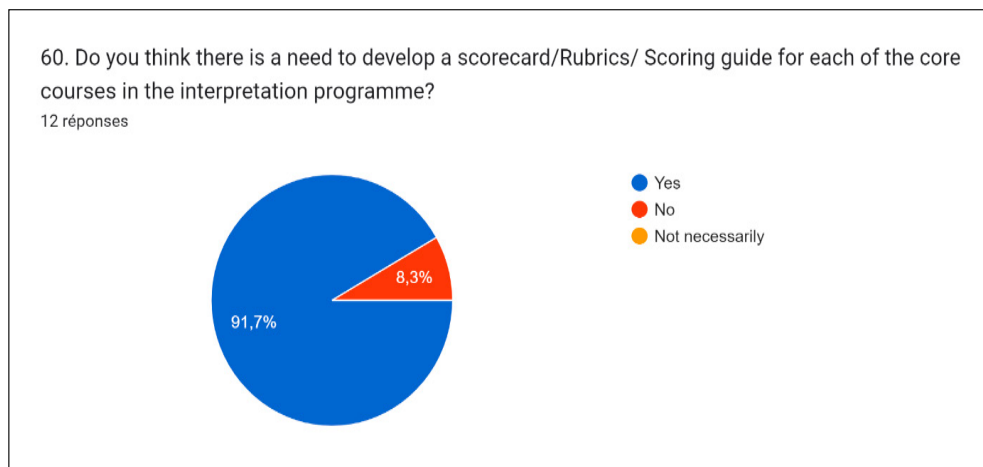
Figure 13:   On developing a scoring guide

Because 75% of respondents reported that neither they nor their colleagues interpreted the texts or speeches proposed for student evaluation, such a guide would systematise the need for lecturers to interpret exams texts themselves before they submit these texts to students.

5.   DISCUSSION OF RESULTS

The enterprise of ITA and its developmental trajectory is shaped by forces originating from the interpreting profession. The fundamental force driving ITA is the societal and political demand for high quality interpreting services in response to the global mobility of goods, services, technologies, ideas and people. The trend of globalisation, which has boosted business, cultural and people-to-people exchanges, has also scaled up the need for interpreter-mediated communication. Furthermore, the provision of interpreting services has been institutionalised in bi/multilateral diplomatic situations and international organisations, notably the United Nations and the European Union.

With reference to developments in China, Chen and Han (2021) explain that testing and assessment practice received a major boost in the 2000s in relation to admission testing, formative assessment (self and peer assessment) and summative assessment at the end of training programmes (Su 2019). Its results play a crucial role in informing relevant stakeholders (interpreting clients, practitioners, certifiers, researchers and educators) about a wide array of contextual decision-making, namely admission into and completion of degree programmes, hiring, certification and so on (Li et al. 2022). Assessment, therefore, entails a judgement on a competence displayed by an interpreter. Though interpreter education includes diagnostic testing at the stage of admission, formative testing during the learning process and summative testing at the end of the programme (Sawyer 2004), the focus here is summative assessment of interpreting competence.

In the final analysis, the following strengths of the interpreter assessment approach of ASTI were identified by this study: collegial selection of topics for examination by a panel of external and internal evaluators, having evaluators representing different backgrounds and generations, collegial choice of a particular text or speech, agreeing on general assessment criteria and providing some background information to students before the start of the examination proper.

The article, however, points out the following as weaknesses of the approach used at ASTI: lack of consistency in marking criteria and training of trainers, the absence of an assessment guide and failure of evaluators themselves to interpret a text or speech prior to the exam. In the absence of a formal guide that all evaluators could reference to harmonise their positions during evaluation, having a variety of assessor backgrounds or levels of expertise or intergenerational knowledge appears to be an advantage rather than a weakness, especially given the variety of elements to consider for evaluation. Gile (2004, 25) argues that

> common issues are fidelity norms (what deviations from the source text/speech are considered 'legitimate'?) and their variability, sensitivity to errors and omissions, as well as the relative importance of various quality parameters (quality of language, fidelity, professional behaviour, etc.).

Faced with such a variety of parameters, the more diverse the assessment panel, the better.

Thus, from "*itemized/atomistic analysis*" (Han 2022, 39–40) through

> paper-based or electronic questionnaires presented in the form of a checklist or an assessment grid… *multi-methods scoring* (incorporating both itemized/atomistic analysis and rating scale-based assessment)… *rubric-referenced, rating scale-based assessment*, also known as *rubric scoring*… *comparative judgment*… to automatic assessment,

none of these evaluation methods seems to have attracted enough interest to be retained as preferred scoring method by ASTI. This may speak in favour of the need of ASTI to design, adopt and formalise a mixed and context-based approach for interpreting testing and assessment.

As Han (2022, 30–31) rightly states, if "one area of oral communication – spoken-language interpreting – seems to have drawn far less attention from language testers than it rightfully deserves, given that interpreting, in and of itself, is language-mediated communication", it is more than clear that very little has been heard from the Global South, as the literature review has shown.

While the interpreter training programme of ASTI was launched in 1987, its first ever permanent lecturer was recruited in 2010, some 23 years later, while research on interpreting studies started to appear in the late 1950s. To date (2024), ASTI has four permanent lecturers for its programme. This has opened prospects for academic research, including action re-search. As indicated earlier, "one-fourth of the respondents were PhD holders […] and there were also PhD candidates". Up to 2019, when the first two PhD topics on interpretation were defended at ASTI, no panel had ever included PhD or PhD candidates as members. Thus, capacity building initiatives would help to consolidate, root and possibly expand this initial nucleus of professional interpreters-cum-researchers. Also, as potential visiting lecturers in the programme are known at the beginning of each academic year, training seminars can be programmed and organised with their participation to discuss assessment issues and agree on consistent general principles. They could contribute to making assessment panels less heterogeneous, thus reducing inconsistencies among the personal approaches adopted by each member of the panel. Building a strong permanent team of lecturers is, therefore, key to the harmonisation and improvement of the general assessment approach of a panel. As long as panels are made up of a majority of external evaluators who represent different organisations and who have different assessment philosophies, it will be difficult to substantially reduce assessment subjectivity.

In addition to expanding and strengthening the permanent pool of lecturers, developing a marking guide known to students and lecturers (both permanent and visiting) would also be a major contribution to improving performance assessment in the programme. Such a guide would be a living document that is likely to be amended regularly to reflect new realities, developments and changes, both in the immediate environment of the programme and in the global village at large. It is, therefore, hoped that this action research would trigger concrete steps to achieve greater fairness and objectivity in the interpreter assessment endeavour of ASTI.

In the ever-evolving landscape of language assessment and testing, the quest for innovation and accuracy remains paramount. As we navigate the complexities of linguistic evaluation, it becomes increasingly evident that a multifaceted approach is necessary to capture the nuances and intricacies of language proficiency.

One of the fundamental challenges in language assessment is the need for instruments that are not only reliable and valid, but also sensitive to the diverse contexts in which languages are used. Contextual factors, such as cultural background, educational environment and communicative purpose, exert profound influences on language performance, necessitating a nuanced approach to assessment design.

6. CONCLUSION

Assessing interpreter trainees is an exercise that can be constantly refined, taking into account the developments and needs arising within a specific context. This article aimed to take stock of the evaluation approach used at the Advanced School of Translators and Interpreters (ASTI) of the University of Buea in Cameroon. Assessment methods used to gauge the proficiency of trainees must constantly be evaluated to help overcome challenges in the assessment process that are related to fairness and objectivity. Systematic training of trainers, designing a scoring guide and breaking the closed circle could help finetune the process while upholding ethics considerations related to the evaluation process. Possibilities to combine objective and subjective assessment criteria, as suggested in literature, could be researched further.

REFERENCES

Albl-Mikasa M 2013, 'Developing and cultivating expert interpreter competence,' *The Interpreters' Newsletter*, vol. 18, pp. 17–34. https://core.ac.uk/download/pdf/149227171.pdf

Amato M, & Mack, G 2022, 'Interpreter education and training,' in F Zanettin & C Rundle (Eds), *The Routledge handbook of translation and methodology*, Routledge, London, pp. 457–475. https://doi.org/10.4324/9781315158945-32

Angelelli CV 2005, *'Longitudinal studies and the development of assessment for advanced competencies,'* paper presented at Invited Symposium: The longitudinal study of advanced foreign language capacities: theoretical and methodological issues, Georgetown, District of Columbia, United States.

Azambou NLR 2019, *Evaluation of the interpretation training programme of the Advanced School of Translators and Interpreters (ASTI) of the University of Buea Cameroon*, PhD thesis, University of Buea.

Bartłomiejczyk M, 2006, 'Strategies of simultaneous interpreting and directionality,' *Interpreting*, vol. 8, no. 2, pp. 149–174. https://doi.org/10.1075/intp.8.2.03bar

Cai R-D, Dong, Y-P, Zhao N, & Lin J-X 2015, 'Factors contributing to individual differences in the development of consecutive interpreting competence for beginner student interpreters,' *The Interpreter and Translator Trainer*, vol. 9, no. 1, pp. 104–120. https://doi.org/10.1080/1750399X.2015.1016279

Chen S 2020, 'The impact of directionality on the process and product in consecutive interpreting between Chinese and English: Evidence from pen recording and eye tracking,' *Journal of Specialised Translation*, vol. 34, pp. 100–117. https://jostrans.org/issue34/art_chen.pdf

Chen, S & Han C (Eds.) 2021, 'Testing and assessment of interpreting', in D Li (ed.), *New Frontiers in Translation Studies*, Springer. https://doi.org/10.1007/978-981-15-8554-8_1

Ericsson KA 2000, 'Expertise in interpreting: An expert-performance perspective,' *Interpreting*, vol. 5, no. 2, pp. 187–220. https://doi.org/10.1075/intp.5.2.08eri

Gerver D 1975, 'Psychological approach to simultaneous interpretation,' *Meta*, vol. 20, no. 2, pp. 119–128. https://doi.org/10.7202/002885ar

Gile D 2009, *Basic concepts and models for interpreter and translator training*, John Benjamins, Amsterdam. https://doi.org/10.1075/btl.8

Gile D 2004, 'Translation research and interpreting research: kinship, differences and prospects for partnership,' in C Schäffner (ed), *Translation research and interpreting research. Traditions, gaps and synergies*, Multilingual Matters, pp. 10-34. https://doi.org/10.21832/9781853597350-003

Gile D 1999, 'Variability in the perception of fidelity in simultaneous interpretation,' *Hermes: Journal of Linguistics*, vol. 22, no. 22, pp. 51–79. https://doi.org/10.7146/hjlcb.v12i22.25493

González RD, Gatto P, & Bichsel J 2010, 'Constructing a valid and reliable trilingual interpreting testing instrument,' in RL McKee & JE Davis (eds), *Interpreting in multilingual, multicultural contexts,* Gallaudet University Press, pp. 55–82. https://muse.jhu.edu/ chapter/272584/pdf. https://doi.org/10.2307/j.ctv2rcnft1.6

Han C 2022, 'Interpreting testing and assessment: A state-of- the-art review,' *Language Testing*, vol. 39, no. 1, pp. 30–55. https://doi.org/10.1177/02655322211036100

Han C 2021, 'Analytic rubric scoring versus comparative judgment: A comparison of two approaches to assessing spoken-language interpreting', *Meta*, vol. 66, no. 2), pp. 337−361. https://doi.org/10.7202/1083182ar

Han C 2015, Building the validity foundation for interpreter certification performance testing [Unpublished doctoral dissertation], Macquarie University, Australia. http://minerva.mq.edu.au:8080/vital/access/services/Download/mq:44348/SOURCE1

Han C, Hu J, & Deng Y 2023, 'Effects of language background and directionality on raters' assessments of spoken-language interpreting,' *Spanish Journal of Applied Linguistics*, vol. 36, no. 2, pp. 556–584. https://doi.org/10.1075/resla.21009.han

Han C, & Lu X 2021 'Interpreting quality assessment re-imagined: The synergy between human and machine scoring,' *Interpreting and Society: An Interdisciplinary Journal*, vol. 1, no. 1, pp. 70–90. https://doi.org/10.1177/27523810211033670

Han C, Zheng B, Xie M, & Chen S 2024, 'Raters' scoring process in assessment of interpreting: An empirical study based on eye tracking and retrospective verbalisation,' *The Interpreter and Translator Trainer*, pp. 1–23. https://doi.org/10.1080/1750399X.2024.2326400

Lee J 2019, 'Subjective assessment of interpreting: A review of methods and issues,' *Interpreting Studies Quarterly*, vol. 8, no. 2, pp. 143–162.

Lee S 2015, 'Developing an analytic scale for assessing undergraduate students' consecutive interpreting performances,' *Interpreting*, vol. 17, pp. 226–254. https://doi.org/10.1075/intp.17.2.04lee

Li X 2013, 'Are interpreting strategies teachable? Correlating trainees' strategy use with trainers' training in the consecutive interpreting classroom,' *The Interpreters' Newsletter*, vol. 18, pp. 105–128.

Li X, Li X, Chen S, Ma S, & Xie, F 2022, 'Neural-based automatic scoring model for Chinese–English interpretation with a multi-indicator assessment,' *Connection Science*, vol. 34, no. 1, pp. 1638–1653. https://doi.org/10.1080/09540091.2022.2078279

Moser-Mercer B 2008, 'Skill acquisition in interpreting: A human performance perspective,' *Interpreter and Translator Trainer*, vol. 2, no. 1, pp. 1−28 https://doi.org/10.1080/1750399X.2008.10798764

Moser B 1978, 'Simultaneous interpretation: A hypothetical model and its practical application,' in D Gerver & HW Sinaiko (eds), *Language interpretation and communication,* Plenum Press, New York, pp. 353–368. https://doi.org/10.1007/978-1-4615-9077-4_31

O'Brien S, Li J, & Tawil I 2019, 'Interpreting proficiency testing,' in S Laviosa & M González-Davies (eds), *Routledge handbook of translation and education,* Routledge, London, pp. 504–520.

Pöchhacker F 2007, *Developing interpreting expertise: Cognitive processes and training implications,* John Benjamins.

Pöchhacker F 2000, *Dolmetschen: Konzeptuelle Grundlagen und Deskriptive Untersuchungen [Interpreting: Conceptual foundations and descriptive studies]*, Stauffenburg-Verlag.

Schaeffer M, Huepe D, Hansen-Schirra S, Hofmann S, Muñoz E, Kogan B, Herrera E, Ibáñez A, & García AM 2020, 'The Translation and Interpreting Competence Questionnaire: An online tool for research on translators and interpreters,' *Perspectives*, vol. 28, no. 1, pp. 90–108. https://doi.org/10.1080/0907676X.2019.1629468

Setton R, & Dawrant A (eds) 2016, *Conference interpreting: A complete course*. John Benjamins, Amsterdam. https://doi.org/10.1075/btl.120

Setton R 1999, *Simultaneous interpretation: A cognitive-pragmatic analysis*, John Benjamins. https://doi.org/10.1075/btl.28

Sawyer DB 2004, *Fundamental aspects of interpreter education: Curriculum and assessment*, John Benjamins. https://doi.org/10.1075/btl.47

Su W 2019, 'Interpreting quality as evaluated by peer students,' *The Interpreter and Translator Trainer*, vol. 13, no. 2, pp. 177–189. https://doi.org/10.1080/1750399X.2018.1564192

Tiselius E & Sneed K 2020, 'Gaze and eye movement in dialogue interpreting: An eye-tracking study,' *Bilingualism: Language and Cognition*, vol. 23, no. 4, pp. 780–787. https://doi.org/10.1017/S1366728920000309

Wang W-W, Xu Y, Wang B-H, & Mu L 2020, 'Developing interpreting competence scales in China,' *Frontiers in Psychology*, vol. 11, Article 481. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.0048. https://doi.org/10.3389/fpsyg.2020.00481

Wang BH 2007 'Assessment of interpreting competence and interpreter competence. Towards an objective model,' *Foreign Language World*, vol. 120, pp. 44–50.

Wen H & Dong Y-P 2019, 'How does interpreting experience enhance working memory and short-term memory: A meta-analysis,' *Journal of Cognitive Psychology*, vol. 31, no. 8, pp. 769–784. https://doi.org/10.1080/20445911.2019.1674857