Status Quo of the Formative Assessment Enactments in Spoken Language Interpreter Training: A Scoping Review of Research and Practice

Da Yan, Shaidatul Akma Adi Kasuma[1], Mansour Amini
School of Languages, Literacies and Translation, Universiti Sains Malaysia, Penang, Malaysia
Corresponding Author Email: shaidatul@usm.my[1]

Abstract
For the past decades, the enactments of formative assessment in various educational settings have produced fruitful research outcomes. Nonetheless, in interpreter training, the application of formative assessment remains insufficient and unsystematic. The purpose of the review is to provide a comprehensive synopsis of existing knowledge in the enactments of formative assessment for interpreter training settings. Methodologically, the scoping review adhered to the five-step procedure developed by Arksey and O'Malley for the examination of 16 documented on the enactments of formative assessment in spoken language interpreter training which were retrieved from Web of Science and Scopus scholarly databases. The results of the scoping review indicated that: 1) documented research and practices have predominantly utilized a mono-strategic design in preparing assessment tasks; 2) existing studies have displayed highly diversified formative assessment design specifications, and 3) previous research has underlined the effects of formative assessment in enhancing trainee achievement and motivation. Correspondingly, the review of exiting literature has accentuated the knowledge gap of how to design and enact effective and standardized formative assessment practices for interpreter training. The scoping review delivers implications for educators and researchers interested in further incorporating formative assessment in interpreter education.
Keywords: Formative Assessment, Spoken Language Interpreting, Interpreter Training, Scoping Review

Introduction
In the field of educational science, the critical role played by assessment has been increasingly recognized (Murchan, 2017). Formative assessment, taking place during the teaching and learning progress, is used to provide feedback with which educational adjustment were made (Gipps, 2012). However, in the early days, the practice of formative assessment has been glutted with misconceptions, particularly the oversimplification as either "an instrument" or

a process "during instruction" (Bennett, 2011, p. 6). The definitional issue was addressed by Black and Wiliam (1998), who defined formative assessment as a continuous process to elicit students' feedback on learning progression for improvement of teaching and learning. With sustained efforts to theorize and enact formative assessment, it is argued as one of the most frequently adopted and effective types of assessment for learning (Univio & Pérez, 2019).

In recent decades, formative assessment has been applied in various educational settings with remarkable outcomes, e.g., early childhood education (Braund et al., 2021), teacher education (Buchholtz et al., 2018), and technology-enhanced settings (Chu et al., 2019). Copious assessment strategies for formative assessment has been supplied to increase the practicability of assessment tasks (Irons & Elkington, 2021). Additionally, Wiliam and Thompson (2008) has proposed a taxonomy of five key strategies for the classification of all applicable formative assessment strategies: 1) clarifying, sharing, and understanding goals for learning and criteria for success with learners; 2) engineering effective classroom discussions, questions, activities, and tasks that elicit evidence of students' learning; 3) providing feedback that moves learning forward; 4) activating students as owners of their own learning; and 5) activating students as learning resources for one another (p. 16).

A plethora of studies have expanded our knowledge of the effects of formative assessment practices on educational outcomes. To begin with, researchers have labelled formative assessment as one of the most effective means to enhance students' academic achievement. In accordance with the results of the meta-analysis conducted by Black and Wiliam (1998), a typical effect sizes of formative assessment on standardized test varied between 0.4 and 0.7. Similarly, Hattie and Timperley (2007) have reported an even larger effect size (0.79) of feedback on achievement, making it one of the most significant decisive factors for learning outcomes. However, the above effect size was challenged by a group of researchers who have doubted the quality of some of the documented cases (Bennett, 2011; Dunn & Mulvenon, 2009). In a follow-up study, the effect size was amended into between 0.2 and 0.8, indicating a variance in the efficacy of formative assessment per case (Palm et al., 2017). Furthermore, evidence has testified to the role of formative assessment in motivating learners. According to the observation of Leenknecht et al., (2021), formative assessment was impactful on learning motivation since the "feedback loop … [functions]… through students' motivation" (Leenknecht et al., 2021, p. 4). In a medical program, Evans et al. (2014) claimed that formative assessment were "stimulating, interactive and entertaining" among participants who enjoyed their involvement in a "active learning" environment (Evans et al., 2014, p. 296). Similar observation could be found in other studies to use formative assessment in foreign language learning (Dmitrenko et al., 2021), automatic assessment (Barana et al., 2019), and K-12 educational program (Chng & Lund, 2018).

Contextualized in interpreter training, published research and documented practices are few and far between. As early as 2005, Lee (2005) made initial effort to formatively apply self-assessment in interpreting classrooms to promote self-regulation of trainees. Henceforth, formative assessment practices gradually appeared in Asian interpreter training programs (Han, 2018b). It should be noted that beyond the scope of spoken language interpreter training, a few cases existed in incorporating formative assessment in sign language interpreter training (e.g., Stauffer, 2012). Nevertheless, including the studies in sign language interpreting, the total number of studies remain rather limited, especially compared against the popularity in other educational settings.

At current stage, we are facing a paucity of research to summarize the status quo of formative assessment enactments in interpreter training. To the best of the authors' knowledge, the

work by Han (2022), in which formative assessment is briefly reviewed as one of the assessment strategies, is the only research synthetization of assessments implemented in interpreter setting. Consequently, the depth and scope of that review are insufficient to fully display the research progression and knowledge gap for the specific niche. Additionally, Han (2018b) criticized that existing research suffers from flaws in design and blemishes in operation. To facilitate the enactments of formative assessment in interpreter training, the conceptual understanding of formative assessment practices, strategies utilized in existing practices, and its efficacy in interpreter training would be significant. Otherwise, interpreter training would be limited to benefit from formative assessment, which has been evidenced by abundant literature as a key component for the advancement of educational outcomes and learners' affective experiences (Leenknecht et al., 2021).

Against the above backdrops, a scoping review of formative assessment enactments in spoken language interpreter training is conducted. The rationale of choosing a scoping review over a systematic review is that the former tries to identify knowledge gaps and underpinning concepts through a more inclusive search strategy for literature while the latter focuses on seeking answers to specific research questions (Munn et al., 2018). Furthermore, Arksey and O'Malley (2005) has enumerated four major reasons to perform a scoping review: 1) to examine the extent and nature of research activities; 2) to determine the value of undertaking a full systematic review; 3) to summarize and disseminate research findings; and 4) to identify research gap in the existing literature (p. 21). For the research synthetization of formative assessment enactments in interpreter training, a scoping review of the fourth type is applicable as the identification of current knowledge gap is critical but the existing body of literature in the field does not merit a "full systematic review" (Arksey & O'Malley, 2005, p. 22). The current scoping review aims to elucidate the existing knowledge lacuna pertaining to formative assessment implementations within the domain of interpreter training. This endeavor involves a meticulous examination of the assessment strategies employed, the delineation of specifications pertaining to formative assessment designs, and an exploration of the consequential impact of formative assessment enactments on training outcomes, as evidenced in extant scholarly literature.

Methods and Materials
The scoping review was guided by the 5-step methodological framework of Arksey and O'Malley (2005) for the replication of search strategies, transparency in data curation, and reliability in research findings. Specifically, the following procedures were followed: 1) identification of research questions; 2) identifying relevant studies; 3) study selection; 4) data charting; and 5) collocating, summarizing, and reporting the results (Arksey & O'Malley, 2005, p. 22).

Identifying The Research Questions
PICOC strategy was used to formulate the research questions of the scoping review (Petticrew & Roberts, 2005). Specifically, the following details were considered: 1) population: interpreter trainees and interpreter trainers; 2) intervention: the enactments of formative assessment (or other classroom assessment strategies formatively); 3) comparison: the comparison with control groups (interpreter trainees in a learning environment without implementing formative assessment during training); 4) outcomes: learning achievements, evaluated performance, students' motivation and engagement, and learning artifacts of trainees; and 5) context: spoken language interpreter training programs. Thus, the aim of the

scoping review is to identify the knowledge gap of formative assessment enacted in interpreting training through the inquiry into the key aspects of assessment design from existing literature. To attain such research objective, the following research questions would be answered:

RQ1: What assessment strategical choices in preparing formative assessment tasks for interpreter training?

RQ2: How are the enacted formative assessment practices for interpreter training designed?

RQ3: What are the effects of the enacted formative assessment practices on the learning outcomes of interpreter training?

Identifying Relevant Studies

The scoping review followed the recommendation of Arksey and O'Malley (2005) to use wider definitions of key terms for a broad coverage of literature. Consequently, we identified several main terms: "formative assessment", "self-assessment", "peer assessment", "portfolios" and "interpreter/interpreting". The identification and inclusion of the above terms were based on conceptualization and understanding of formative assessment and interpreter education in previous studies (Akpan et al., 2012; Black & Wiliam, 2003; Han, 2018b). Additionally, the term "classroom assessment" was appended to the identified terms as cases of applying classroom assessment strategies formatively were reported in exiting literature (Black & Wiliam, 2004; Earl, 2012). The complete search string provided for literature search is shown in

Table 1.

Table 1.

Search Strings for Web of Science and Scopus.

Search String
("interpreter" OR "interpreting" OR "interpreter training" OR "interpreter education" OR "interpreting training" OR "interpreting education" OR "interpreter trainee" OR "interpreter trainer") AND ("formative assessment" OR "self assessment" OR "peer assessment" OR "portfolio" OR "classroom assessment")

Web of Science (WOS) and Scopus databases were searched to identify relevant publications in July 2023. To facilitate the selection and screening of publications from a large volume of existing literature retrieved from the two major academic databases, the following inclusion criteria (IC) and exclusion criteria (EC) were adhered to:

IC1: Include documents on developing and enacting formative assessment in a spoken language interpreting setting.

IC2: Include documents using formative assessment as a pedagogical intervention.

IC3: Include documents reporting the effects or reception of formative assessment.

EC1: Exclude documents whose assessment strategies are apparently not formatively enacted.

EC2: Exclude documents containing relevant keywords but not focusing on formative assessment in a spoken language interpreter training setting.

EC3: Exclude documents without empirical findings.

EC4: Exclude documents repeating findings testified in earlier works.

Study Selection

After downloading the search results from WOS (n=292) and Scopus (n=429), the deduplication was performed using Zotero. As a result, 460 unique references were retained with 261 duplicates excluded. The researchers excluded irrelevant studies through reading the abstracts and titles of the included publications. When the information provided in the abstract and titles was not sufficient for screening, screening based on full text applied. In accordance with the IC and EC, 65 papers were retained. Finally, full text versions of all retained publications were retrieved and evaluated for their eligibility. Disagreement to include or exclude of a publication was resolved through discussion. Eventually, 16 publications were selected for the scoping review. See **Figure 1** for the PRISMA flow diagram (Moher et al., 2011) displaying the results and process of study selection.
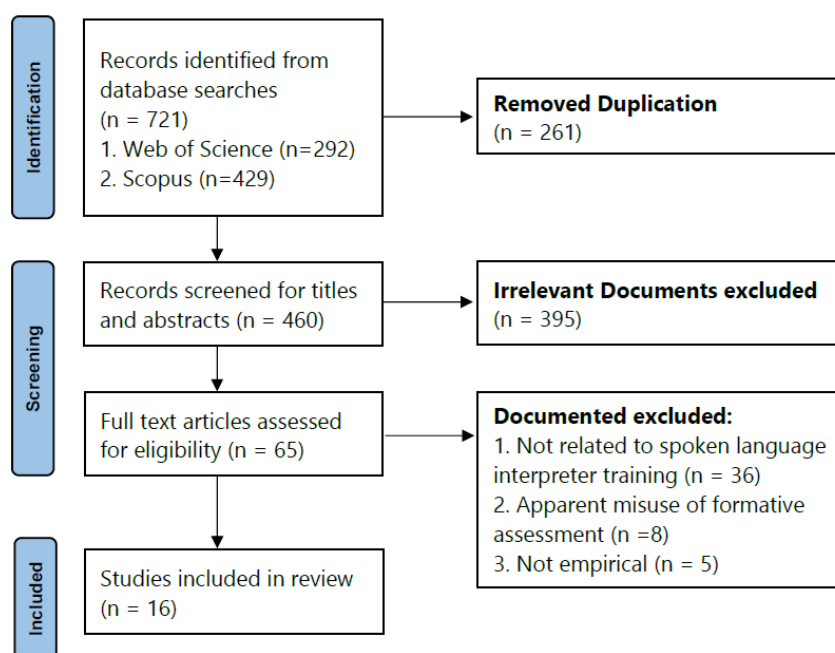
Figure 1. The PRISMA flow diagram of the review.

Data Charting And Collation

The subsequent step is to chart and collate the selected documents. For each selected document, publication details and information related to the research questions, i.e., title, author, publishing year, publication type, research objectives, participants involved in the study, the design of the formative assessment, and the major findings, were compiled. The details of curated data are shown in

Table 2.

Table 2.
Summary of selected documents

Title	Author (Year)	Type	Objectives	Participants	Formative Assessment Design	Findings
(1) Using self-assessment as a formative	Han & Fan (2020)	Journal Article	To investigate the reflective observations of	38 third-year undergraduate students	Students are required to participant	Very complex and even contradictory finding emerge from

assessment tool in an English-Chinese interpreting course: student views and perceptions of its utility			participants on the potential benefits and drawbacks of self-assessment in an interpreter training context.	from an interpreter training program	in three assessment performances over 10 weeks; in each assessment, students need to complete six interpreting tasks; the assessments are graded and used for both formative and summative purposes.	the results of qualitative investigation. Most of the general comments regarding the adopting for self-assessment in a formative assessment manner are favorable. To participants, the positive effects of formative assessment in motivating and informing students of their learning progress are appreciated. However, negative comments are also heard from respondents, especially about the subjectivity and in efficiency of formative assessment. Additionally, students called for the inclusion of synergetic formative assessment strategies to augment the effect of a mono-strategic design of the study.
(2) Conceptualizing and operationalizing a formative assessment model for English-Chinese consecutive interpreting: a case study in an undergraduate interpreting course	Han (2018b)	Book Chapter	To develop a conceptual model for formative assessment in an interpreter training context; to validate the conceptual model by developing and implementing a formative assessment.	41 fourth-year undergraduate students from an interpreter training program; a six-member panel of raters participated in the study to evaluate students' performance and provide their reflections.	During an 18-week semester, students are required to take part in three formative assessments at an interval of four weeks. In the final week, a summative assessment is administered; In each assessment, a Chinese to English and English to Chinese interpreting tasks are provided for students; Students are divided into four groups;	The level of satisfaction of students regarding the design and the implementation of the formative assessment in the interpreter training context is acceptable with an average rating of 6 out of 10. The negative evaluation comes generally from a cohort of student who stubbornly reject any kind of test and assessment during the learning process. Additionally, a preference of teacher assessment over self-assessment or peer assessment could be identified. The reflections of lecturers tend to be affirmative. However, issues worth noting include: the ineffective use of feedback from students, lack of task

					the assessments are rubric-based; peer assessment, self-assessment and teacher assessment are both included with peer assessment randomly assigned to students.	bank or item bank, the extra burden caused by formative assessment to lecturers/rateres.
(3) The accuracy of student self-assessments of English-Chinese bidirectional interpretation: a longitudinal quantitative study	Han & Riazi (2018)	Journal Article	A longitudinal study to determine the accuracy level of student self-assessments of English - Chinese and Chinese - English interpretation and examine its change over time.	38 third-year undergraduate learners majoring in English – Chinese translation; a panel of six rater are recruited as raters	Over 10 weeks, students are asked to take part in three formative assessments at week 4,9 and 10; assessment are both formative and summative; contents of assessment are consecutive interpreter tasks; rubric referenced assessment; peer, self, teacher assessments are included and statistically analyzed.	Self-assessment accuracy for both Chinese-to-English and English-to-Chinese interpreting directions generally increased over time; for the three scoring dimensions of information completeness, fluency of delivery, and target language quality, self-assessment accuracy for English-to-Chinese interpretation was higher than in the other direction; while information completeness was best self-assessed in English-to-Chinese direction, the pattern was reversed in the other direction; a tendency of overrating is observed from students' self-assessment of Chinese-to-English interpreting.
(4) Feedback in conference interpreter education: Perspectives of trainers and trainees	Domínguez Araújo (2019)	Journal Article	To investigate perceptions and practices in three postgraduate conference interpreter training programs	31 trainees and 11 trainers and 14 teachers participated; all students are from a master program with various language background.	Self-assessment and peer-assessment are used; assessments are criteria-referenced; students are trained in advance. The design in flawed in accordance with the key features of	Diversified findings emerge. Respondents argue that feedback should be honest, concise and meaningful/constructive for being used for teaching and learning adjustment. Most importantly, students argued that formative assessment should not be limited to grading and judgement but should be constructive and insightful.

					formative assessment.	However, issues are also mentioned, especially the differences in the attitude towards the question "who should be in charge of the assessment?".
(5) Latent trait modelling of rater accuracy in formative peer assessment of English-Chinese consecutive interpreting	Han (2018c)	Journal Article	To investigate the accuracy of peer assessment ratings on English – Chinese consecutive interpreter performances from a Rasch model perspective.	41 students from an interpreter training program; four lecturers serving as raters.	A direct continuation of (Han, 2018b) with identical formative assessment design; assessments are rubric-referenced.	According to the investigation, rater accuracy tends to be widely varied. Statistically significant differences among peer raters could be identified. Additionally, it was simpler for peer raters to deliver accurate ratings for English-to-Chinese interpretation than the other way. Raters are most accurate in assessing target language quality that other domains of the assessment. Methodologically, the research shows that latent trait modelling could be used to facilitate the analysis and understanding of formative assessment in an interpreter training context.
(6) Scale-referenced, summative peer assessment in undergraduate interpreter training: self-reflection from an action researcher	S.-B. Lee (2019)	Journal Article	A first-person action research to study his own experiences of peer assessment during three semesters for interpreter training	About 30 junior and senior students majoring in consecutive interpreting (change because the study is composed of three semesters)	Designed to be scale-referenced, summative peer assessment, but in the end the design functions like a summative assessment applied formatively.	The research focuses on the validity, benefit and procedural difficulties in implementing a summative peer assessment in an interpreter training context. The most interesting findings from the study is that the assessment, summative as prescribed in its original design, functions as a blending of both summative assessment and formative assessment. Consequently, the author continues to argue the interdependency between the two variants of classroom assessment practices.

(7) Self-assessment as 'assessment as learning' in translator and interpreter education: Validity and washback	Li (2018)	Journal Article	To examine the validity and washback of self-assessment administered for interpreter training.	54 undergraduate students majoring in translation and their instructors.	Self-assessment and teacher assessment are both included; the assessments are criteria-referenced; the design is relatively close to the key features of formative assessment; for four times during the course, students are asked to do self-assessment, in addition to teachers' independent rating.	The students' self-assessment correlates positively with teacher assessment; their self-assessment accuracy improved over time with regular repetition; and self-assessment promotes positive learning attitudes among learners.
(8) Peer Feedback and Reflective Practice in Public Service Interpreter Training	Holewik (2020)	Journal Article	To examine the strength and weakness of peer feedback and reflection.	40 final year postgraduate students from a Polish interpreting program with various language background.	Students joined a three-member group for a simulated role play task in which bi-directional interpreting is conducted. Students are required to provide self-assessment and peer assessment in written format. The assessments are not rubric-referenced or scale-reference.	In the comparison of peer feedback and reflection, a favorable tendency could be among respondents towards the former with 62% of the respondents agreeing with its strength. Specifically, the strength of peer feedback is manifested mainly through three perspectives of interpreting competence: presentation and delivery, accuracy and fidelity, and interpreting skills and strategies. However, the author claimed that the survey was not comprehensive and formal.
(9) A longitudinal quantitative investigation into the concurrent validity of self and peer assessment applied to	Han (2018a)	Journal Article	To explore the concurrent validity of self-assessment and peer assessment in an English – Chinese bidirectional interpreting context.	41 fourth-year undergraduate students from two classes in an interpreter training program	Self and peer assessments are included; the assessments are rubric-based; Formative assessment	Conducted in a nearly identical research context as (Han, 2018c), the research continues to provide insightful findings regarding the correlation between student's assessment and teacher's. The

English-Chinese bi-directional interpretation in an undergraduate interpreting course					occurred at week 5,10 and 15, in the concluding week a summative assessment is administered.	author argues that though the ratings from teachers could not be easily replicated in students' rating, the latter has its own merits, especially for the ability to rank-order their performance and generate useful educational feedback for the adjustment of learning strategies. Additionally, the possibility to gain improvement in self-assessment and peer assessment accuracy over time is verified.
(10) Accuracy of peer ratings on the quality of spoken-language interpreting	Han & Zhao (2021)	Journal Article	To explore the relationship between rating accuracy with factors as interpreting directionality, type of peer raters, rating domain/assessment criterion and quality of interpretation.	39 raters with different level of experiences in interpreter training are recruited as raters.	The study adopted a relatively different design. Participants are required to rate students' performances as audio recordings selected from a large database. The details of the assessment design are not disclosed. Since the rating is after the completion of the teaching and learning process, the design could be recognized as formative assessment. But the internal mechanism is similar, and the results are applicable to formative assessment.	Overall, the English-to-Chinese interpretation was easier for the peer raters to score accurately than the opposite direction, and fidelity was easier to evaluate than fluency and expression. In addition, experienced peer raters were better at scoring the interpretations than their less experienced counterparts, and higher-quality interpretations were simpler to score accurately than lower-quality ones.

(11) University students' experience of 'scale-referenced' peer assessment for a consecutive interpreting examination	S.-B. Lee (2017)	Journal Article	To explore how undergraduate students feel about their scale-referenced, peer assessment activities in/for an interpreting examination.	Three students from those participating in the peer assessment scheme for the qualitative research. Total number of students involved in the assessments is 78.	For a 16-week semester, peer assessments are administered at week 4,5,6,7, followed by a summative assessment at week 8. Week 9 is for peer assessment, and the 10th week for a review session.	Students worry about the proposed peer assessment scheme and their abilities. They feel empowered and responsible Assessment results may be greatly divergent. Students are struggling to apply criteria in practice and experience rating fatigue. Students believe that peer assessment is reliable and helpful. Some students express regrets about feedback for not being able to provided insightful and useful information.
(12) Comparing Student Self-assessment and Teacher Assessment in Korean-English Consecutive Interpreting: Focus on fidelity and language	J. Lee (2022)	Journal Article	To comparing students' self-assessments with teacher assessments in terms of a set of quality categories.	20 first-year graduate students of Korean-English interpretation	Teacher assessment and self-assessment are included in the study. A three-dimensional criterion is used to facilitate assessment. The assessment could be regarded as formative assessment as educational feedback is generated and provided to teachers.	Although the instructor used tougher criteria to assess the integrity of the interpreting, the students tended to concentrate more on the quality of the target language in both directions. Students' self-assessment of their own interpreting performance did not exhibit consistent or reliable mistake detection.
(13) Self-assessment as an Autonomous Learning Tool in an Interpretation Classroom	Y.-H. Lee (2005)	Journal Article	To explore the usefulness of self-assessment in an interpreter training classroom.	23 students of a Korean-English program in a two-year graduate school of translation and interpretation	Students are asked to take the audio-recorded tapes of their interpreting performances home for self-assessment; criteria-referenced assessment	It has been discovered that self-assessment helps students recognize their problems and talents, facilitates focused practice, and enables them to track their own development. Self-evaluation, however, was both time-consuming and emotionally taxing for pupils. However, everyone in the class,

					with a three-dimensional rubric. The design in flawed in accordance with the definition of formative assessment.	including the teacher, concurred that self-assessment helped pupils learn in an interpretation classroom.
(14) Rubric-Based Self-Assessment of Chinese-English Interpreting	Su (2021)	Book Chapter	To examine the effectiveness of the rubric-based self-assessment in an interpreter training context	50 students majoring in interpreting at the postgraduate level from 4 Chinese universities were recruited to participate in the study	Assessment takes place in a six-week workshop. Raters are recruited and trained. Rubric is developed and adopted in the assessment.	Limited in a rather narrow scope, the findings from study indicates that the rubric was beneficial for learners to identify their strength and weakness in completing interpreting tasks. Additionally, the adopting of rubric-referenced assessment contributes to the improvement of their performance in a gradual manner. Finally, students tend to focus more on details in their performance instead of the broader level.
(15) The Interpreting Portfolio as a Self-Regulation Tool in Interpreter Training: Preliminary Findings from a Pilot Study	Mirek (2020)	Journal Article	To examine the effectiveness of portfolio as a self-regulated learning tool for interpreter trainees	Ten second-year master students of English Studies whose curriculum include a one term interpreter training	An iterative loop is formed to use self-evaluation (self-assessment) to understand the learners' progress and issues in learning. The design could be identified as a formative assessment adopting self-assessment in the form of learners' portfolio.	Reflection was obtained from a component in the learning portfolio. Student agreed that the portfolio was helpful to enlarge their horizon regarding their own strength and weaknesses as interpreter learners. Additionally, the possibility to monitor learning progression, reduce stress and tension, and increase their confidence. The value of self-assessment as well as the ability to promote self-planning and discovery of the importance of process-oriented learning. Generally, the assessment strategy is favored by participants who stress the willingness to use it in their own learning process.

(16) Formative assessment in the interpreting classroom: Using the portfolio with students beginning simultaneous interpreting.	Arumí Ribas (2010)	Journal Article	To examine the effectiveness of the formative assessment with portfolio for novice interpreter trainees	3rd and 4th year students in a degree program, in each class about 20 students. The total number of participants not disclosed.	Students work with their portfolios over a period of six to seven months during the whole course. According to the definition of formative assessment, the case is a typical application of using portfolio in learning.	The portfolio-based assessment is favored by participants, with 56 pieces of positive feedback against 9 negative feedback. To students, the assessment provided a continuous assessment with which their metacognition of workload could be improved. Additionally, it is beneficial for continuous learning, knowing their learning progress, and gaining a positive learning attitude.

Reporting The Results
The final stage of the methodological procedures is reporting the results. In the analysis, the researchers identified main aspects of the enactments of formative assessment for interpreter training. The results of the scoping review are presented in the following section.

Results
Out of all the selected documents, 8 were conducted in China (50%), 4 in Korea (25%), 2 in Spain (12.5%), and 2 in Poland (12.5%). Regarding the publication types, two were book chapters (n=2, 12.5%) while fourteen were journal papers (n=14, 87.5%). The reviewed documents were published between 2005 to 2022 with 13 published in recent 5 years (after 2018) (n=13, 81.3%). Among all the researchers of the included documents, Chao Han (from China) has contributed the most publications (n=6, 37.5%). In this section, the results of the scooping review were presented in tandem with the research questions.

Strategical Choices In Assessment Task Preparation
Educators and researchers interested in formative assessment avail themselves of a myriad of assessment strategies for a specific educational purpose or pedagogical setting (e.g., 5 classroom strategies in DiVall et al., 2014; and 75 strategies listed in Keeley, 2008). Reversely, among the selected studies, merely four formative assessment strategies were adopted: 1) self-assessment (#1-#4, #6-#9, #12-#15); 2) peer assessment (#2-#6, #8-#11); 3) teacher rating (#2-#3, 7, 12), and 4) student portfolios (#15-#16). Self-assessment was most favorable (n=12, 75%), followed by peer assessment (n=9, 56.3%), teacher rating (n=4, 25%) and portfolios (n=2, 12.5%). Internally, differences existed in the specification of assessment strategies. For example, in Lee's (2005) study, students were demanded to provide self-assessment in promptly after task completion; in the research of Han (2018a), a post-class self-assessment of audio recording was applied; in Han and Zhao (2021), the raters were required to assess audio recordings retried from a database instead of ones of students' live performances. The varieties in the details of a formative assessment strategy revealed the multifold availabilities of assessment strategies in interpreter training. Additionally, among the four assessment strategies, teacher rating was generally used for comparison. For

example, in Lee's (2022) and Li's (2018) studies, teacher rating were both used to validify the accuracy of students' self-assessment of interpreting performance.

Noticeably, the selected studies principally adopted a mono-strategic design, except merely five studies (n=5, 31.3%) utilizing multiple strategies. For example, in the study of Han and Riazi (2018), three sessions of formative assessment were arranged during a 10-week interpreter training workshop. Per session, students were required to complete simulated interpreter tasks of intermediate difficulty. Upon completion of the interpreting tasks, self-assessment, peer assessment and teacher assessment were conducted. In comparison with mono-strategic case such as the study by Su (2021), the adoption of multiple strategies during a formative assessment would produce detailed results and feedback, in addition to extra possibilities to elicit issues of a certain assessment strategies (Han & Riazi, 2018).

Formative Assessment Design
Regarding the specific formative assessment designs, significant disparities existed among the selected documents. After synthetization, the following three aspects emerged: 1) study context; 2) frequencies and durations of formative assessment enactments; and 3) rating instruments or criteria.

First, the context of most studies was similar. All studies took place in undergraduate (#1-#3, #5-7, #9-#11, and #16; n=10, 62.5%) or postgraduate (#4, #8, and #12-#15; n=6, 37.5%) interpreter training programs. However, the percentage of trainer and trainee involvement was not on a par with each other. Six studies (#2-#5, #7, and #10; n=6, 37.5%) recruited lecturers, trainers, or raters as participants. The mean of lecturer participants was 13.5 and the median 6. Alternatively, in Lee's (2017) study, the researcher recruited three participating students out of a total number of 78 for the phenomenological study. Thirteen studies revealed the instructional context to which the formative assessment belonged (#1-#3, #5-#7, #9-#11, #13-#16; n=13, 81.3%), with the backgrounds undisclosed in three of them (#4, #8, #12; n=3, 18.7%). For example, in the study by Mirek (2020), the total duration for the courses was 15 weeks, in which thirteen weeks were used to provide "intense training" (p. 147). Contrarily, in a recent study of Lee (2022), the details pertaining to the instructional background or the course specification were unknown. Among all the studies with an explicit introduction of its research context, most researchers conducted the formative assessment as an educational intervention or a pedagogical innovation, with the exception of the study of Han and Zhao (2021) which was an evaluation of rating accuracy of concluded peer assessments. Furthermore, pre-study training regarding the specification of rubrics and formative assessment were generally provided to enhance participants' assessment literacy and familiarity with the relevant study.

Second, disparities existed in the frequencies and durations of formative assessment enactments. In five studies, formative assessment was enacted as a one-off attempt (#4, #8, #10, #12, and #13; n=5, 31.3%), while in eleven studies (#1-#3, #5-#7, #9, #11, and #14-#16; n=11, 68.7%), formative assessment was implemented for multiple times. In most cases adopting self- and peer assessment, the duration lasted merely for a few weeks. For instance, in Su's (2021) research to investigate the effectiveness of formative assessment in a 6-week workshop of interpreting, three sessions were arranged for students to complete rubric-based self-assessments of interpreting abilities. Reversely, study using portfolio lasted dramatically longer than those using other strategies. For example, in the study of Arumí Ribas (2010), the whole process for students to incrementally complete self-assessment portfolios lasted for several months. Noticeably, studies conducted in the Chinese interpreter training

context generally adopted repeated enactments of formative assessment. Contrarily, one-off enactment was identified from research out of the Chinese context, i.e., 2 Korean, 1 Polish, and 1 Spanish studies (n=4, 25%).

Finally, several studies have adopted or developed ad hoc rating instruments. For example, Su (2021) adopted a strategy-based rubric encompassing grammar- and meaning-based rating criterion; in Lee's (2005) study, a three dimensional criteria measuring the meaning, language use and delivery of interpreting was adopted; in the efforts to evaluate rater accuracy using a latent trait modelling analysis, Han (2018c) developed an eight-point analytic rating scale to assess students' interpreting competence from the angle of information completeness, fluency of delivery and target language quality.

Effects Of The Formative Assessment Practices

Significant effects of formative assessment on educational outcomes were reported. Participants involved stressed the empowerment and enhancement from the enactments of formative assessment. For instance, the longitudinal study #9 (Han, 2018a) revealed that student's performances in interpreter and assessment would be mutually improved over time. Additionally, the varied efficacies of formative assessment in different micro-settings were observed. For instance, in study #10 (Han & Zhao, 2021), the researchers argued that ratings of L2-to-L1 (i.e., English-to-Chinese) interpreting were more accurate than the L1-to-L2 ones. Similarly, in the case of study #12 (J. Lee, 2022), the behavioral features of student raters were examined, indicating learners were more inclined to assess interpreting performance on the syntactical level instead of the discourse level.

Apart from its effect on training progression, the reviewed publications highlighted the affective enhancement resulted from formative assessment enactments. As in study #15 (Mirek, 2020), students emphasized that formative assessment played a significant role in increasing their confidence and reducing stress. Similarly, the study #16 (Arumí Ribas, 2010) pointed out that formative assessment promoted the self-regulation and metacognitive regulatory skills of novice learners.

However, issues pertaining to the enactments of formative assessment were also reported. The most frequently raised issue was the mental expenditure posed on trainees. For example, in the case of study #13 (Y.-H. Lee, 2005), the participants argued that the formative assessment tasks were time-consuming and mentally taxing. The psychological burnout was not only experienced by students, but also lecturers and raters, as in the case of study #2 (Han, 2018b). Additionally, the credibility and objectivity of formative assessment, especially in self- and peer assessment, were frequently doubted. For instance, in the study of Han and Fan (2020), the trainees complained the strongly subjective grading in self-assessment and the ignorance of "deep-seated problems" in interpreting performance (p. 118).

Discussions

The scoping review investigated the application of formative assessment in interpreter education. In the following sections, the current knowledge and knowledge gap were presented and discussed against existing theoretical and empirical insights from a broader context.

First, we have found a shared preference for a mono-strategic scheme in preparing assessment tasks. Only in a few studies (e.g., #2, #9, and #12), the multi-strategic schemes, which delivered more comprehensive and insightful understanding of formative assessment, were adopted. Conversely, in most of the reviewed studies, the strategic choices in the

formative assessments were monotonous. Incontestably, each individual strategy utilized in the selected documents (i.e., self-assessment, peer assessment, teacher rating and portfolio) could be justified by one of the five key strategies articulated by Wiliam and Thompson (2008). However, the observed strategical preference was contrary to the suggestion from Lyon el al. (2019) that multiple formative assessment strategies, combined with tight integration with instruction and sound assessment design, should be applied for enhanced effects. The invariable strategical choices within the formative assessment enactments could be attributed to the researcher's overemphasis on the authenticity of assessment (Frey et al., 2019). Inarguably, self-assessment, peer-assessment and portfolio, judged by the five-dimension framework proposed by Gulikers et al. (2004), are good practices to attain authenticity of assessment by simulating the working environments and task requirements faced by professional interpreters. However, other critical aspects in developing and enacting formative assessment, e.g., actionability and timeliness of the assessment, were not fully considered (DiVall et al., 2014).

Second, mixed findings emerged from the investigation into the designs of each reported formative assessment practice. Most of the formative assessment designs were in partial agreement with the framework proposed by Andersson and Palm (2017), in which the procedures of formative assessment were in alignment with the pedagogical processes. However, some key features of formative assessment were not reflected in the selected documents, i.e., the iterative nature of formative assessment (Rr et al., 2020), the cyclic process with closing the gap and reidentifying new gaps (Heritage, 2010), and the multiple phases within a formative assessment loop as discussed in the work by Gulikers and Baartman (2017). In the reviewed documents, the noniterative and noncyclic assessment practices has prolonged and curbed the elicitation and utilization of feedback. The shortcomings of the existing designs of formative assessment enactments and the corresponding inefficient feedback processing echoed the criticism of Wiliam (2006) that poorly designed formative assessment practices are those with ineffective the evocation and interpretation of evidence. Furthermore, the drawback to utilize noncyclic and nonsystematic assessment tasks in interpreter training as formative assessment resonates with Vermeulen and van der Kleij's (2012) view that such assessment practices are informal and more suitable for low-stakes assessment. In most of the reviewed documents, in which teacher-, peer- and self-grading were mandatory, the combination of noncyclic assessment designs and high-stakes assessment tasks was highly inappropriate.

Third, we have identified the significant effects of formative assessment enactments on interpreter trainees' achievement and affective experiences. One the one hand, the finding was generally in tandem with evidence accumulated over the years to utilize formative assessment in various educational settings (Arja et al., 2018; Chu et al., 2019; Granberg et al., 2021). On the other hand, formative assessment practices in the included documents have been criticized by participants for increasing mental efforts of trainees. Such finding partially echoed with the claim from existing studies that the quality of formative feedback was subject to time constraints (Palmer & Devitt, 2014). Most importantly, the reviewed studies have reported no significant effects of formative assessment practices on during-enactment learning and teaching. The phenomenon is contradictory to the account of Wiliam and Thompson (2008) that successful formative assessment practices is determined by the during-learning processes, i.e., the formation, processing and utilization of feedback. Finally, the peer and trainee-trainer interactions in the reviewed documents were minimal and mostly took place in the grading of interpreting artifacts. This phenomenon contradicts the findings of

existing research in the field of formative assessment in which classroom interaction, particularly teacher-student interaction, has consistently been highlighted (e.g., Nieminen et al., 2021). Considering the collaborative and interactive nature of interpreter training (Bassnet, 2012), the absence of substantial classroom interaction propels researchers to introspect the strategical choices and assessment designs of the documented enactment cases.

The issues in strategical choices, assessment designs and effects from included studies pointed to the gaps in our current conceptual and practical knowledge of formative assessment in an interpreter training context. To begin with, existing practices of formative assessment were not sufficiently integrated with curriculum or pedagogical activities. Ideally, as explained by the procedural framework of Heritage (2010), the loop of formative assessment commenced with the identification of knowledge gap and the clarification of learning goals. However, in the selected documents, most formative assessment was conducted without clear definition of the specific learning objectives. For example, students were required to complete their interpreting tasks before the chance to self-assess their performance thereof in the case of study #11. Consequently, the respondents reported their lack of confidence in applying the rating or evaluating criteria during the self-assessment. The lack of integration with the curriculum would not only make formative assessment subjective and unreliable (Orsmond et al., 2004; Zheng et al., 2021), but also curb its innate advantage to be iteratively adopted during a relatively longer period of teaching and learning (Xiao & Yang, 2019).

Furthermore, rigorous design and standardized procedures of formative assessment were not upheld in the included studies. With abundant experiences in applying formative assessment in a wide variety of educational domains (Bacquet, 2020; Johnson et al., 2019), the rigor and credibility of formative assessment for interpreting training should not be overshadowed by the domain-specific characteristics of interpreter education. Specifically, developers were encouraged to pay special attention to the key features of formative assessment, e.g., the iterative implementation of formative assessment (Heritage, 2010; Wiliam & Thompson, 2008), the choice of strategies or strategy combinations (DiVall et al., 2014), eliciting students' reflection (Gotwals & Birmingham, 2016), and making educational adjustment based on the feedback during formative assessment (Andersson & Palm, 2017; H. Lee et al., 2020). However, in the reviewed publications, how teachers and learners made use of the outcomes of formative assessment for pedagogical improvements were nearly nonexistent. The situation echoed with the criticism of Andrade (2019) that poorly designed self-assessment failed to offer learners opportunities to "do something about their perceived low competence" (p. 3).

There are a few limitations faced by the scoping review. First, we have only selected and reviewed publications in English from WOS and Scopus databases. The rationale to exclude document written in other languages was to facilitate publication selection and reviewing processes. Judging from the proportion of studies conducted in the Chinese interpreter training context in the present study, follow-up researchers could perform a comparative study between documented enactments of formative assessment in interpreter training from Chinese academic databases and those in other countries. Second, we have applied inductive coding method during the synthetization processes in the scoping review. Consequently, the collation, analysis, and interpretation would produce bias out of the subjectivity and individual experiences of the researchers. To mediate such limitation, we have applied multiple bias-controlling measures: 1) tapping researcher triangulation for internal

consistency during the charting and collation processes; and 2) referring to existing research in the field of formative assessment as theoretical guidance to frame the knowledge gap of enacting formative assessment in interpreter training.

Conclusion

The scoping review examined the status quo of relevant research on the enactments of formative assessment in spoken language interpreter training. The findings revealed that: 1) most existing studies adopted a mono-strategic scheme regarding the selection of assessment tasks; 2) existing studies did not closely follow the recommended operationalizations of formative assessment articulated in previous literature; 3) minimal expansion in knowledge of the effects of formative assessment on educational outcomes could be obtained from the selected studies. Based on the acquired knowledge of formative assessment enactments in interpreter training, the review pointed out to future direction for subsequent research in the field. Being a pioneering scoping review of relevant research from a critical stance, the study was significant for researcher and educators interested in advancing the enactment of formative assessment in interpreter education.

References

Akpan, J. P., Notar, C. E., & Padgett, S. A. (2012). Formative Assessment: A Basic Foundation for Teaching and Learning. *National Teacher Education Journal*, *5*(1).

Andersson, C., & Palm, T. (2017). Characteristics of improved formative assessment practice. *Education Inquiry*, *8*(2), 104–122. https://doi.org/10.1080/20004508.2016.1275185

Andrade, H. L. (2019). A Critical Review of Research on Student Self-Assessment. *Frontiers in Education*, *4*. https://doi.org/10.3389/feduc.2019.00087

Arja, S. B., Acharya, Y., Alezaireg, S., Ilavarasan, V., Ala, S., & Arja, S. B. (2018). Implementation of formative assessment and its effectiveness in undergraduate medical education: An experience at a Caribbean Medical School. *MedEdPublish*, *7*, 131. https://doi.org/10.15694/mep.2018.0000131.1

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Arumí Ribas, M. (2010). Formative Assessment in the Interpreting Classroom: Using the Portfolio with Students Beginning Simultaneous Interpreting. In M. Garant (Ed.), *Current Trends in Translation and Interpreting III* (pp. 97–133). University of Helsinki, Department of Translation Studies.

Bacquet, J. N. (2020). Implications of Summative and Formative Assessment in Japan – A Review of the Current Literature. *International Journal of Education and Literacy Studies*, *8*(2), Article 2. https://doi.org/10.7575/aiac.ijels.v.8n.2p.28

Barana, A., Marchisio, M., & Sacchet, M. (2019). Advantages of Using Automatic Formative Assessment for Learning Mathematics. In S. Draaijer, D. Joosten-ten Brinke, & E. Ras (Eds.), *Technology Enhanced Assessment* (pp. 180–198). Springer International Publishing. https://doi.org/10.1007/978-3-030-25264-9_12

Bassnet, S. (2012). Translation studies at a cross-roads. In E. Brems, R. Meylaerts, & L. van Doorslaer (Eds.), *The Known Unknowns of Translation Studies* (Vol. 24, pp. 17–27). John Benjamins.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25.

https://doi.org/10.1080/0969594X.2010.513678

Black, P., & Wiliam, D. (1998). Inside the Black Box: Raising Standards through Classroom Assessment. *Phi Delta Kappan*, *80*(2), 139–148. https://doi.org/10.1177/003172171009200119

Black, P., & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal*, *29*(5), 623–637. https://doi.org/10.1080/0141192032000133721

Black, P., & Wiliam, D. (2004). Classroom Assessment is Not (Necessarily) Formative Assessment (and Vice-versa). *Teachers College Record*, *106*(14), 183–188. https://doi.org/10.1177/016146810410601408

Braund, H., DeLuca, C., Panadero, E., & Cheng, L. (2021). Exploring Formative Assessment and Co-Regulation in Kindergarten Through Interviews and Direct Observation. *Frontiers in Education*, *6*. https://doi.org/10.3389/feduc.2021.732373

Buchholtz, N. F., Krosanke, N., Orschulik, A. B., & Vorhölter, K. (2018). Combining and integrating formative and summative assessment in mathematics teacher education. *ZDM*, *50*(4), 715–728. https://doi.org/10.1007/s11858-018-0948-y

Chng, L. S., & Lund, J. (2018). Assessment for Learning in Physical Education: The What, Why and How. *Journal of Physical Education, Recreation & Dance*, *89*(8), 29–34. https://doi.org/10.1080/07303084.2018.1503119

Chu, H.-C., Chen, J.-M., Hwang, G.-J., & Chen, T.-W. (2019). Effects of formative assessment in an augmented reality approach to conducting ubiquitous learning activities for architecture courses. *Universal Access in the Information Society*, *18*(2), 221–230. https://doi.org/10.1007/s10209-017-0588-y

DiVall, M. V., Alston, G. L., Bird, E., Buring, S. M., Kelley, K. A., Murphy, N. L., Schlesselman, L. S., Stowe, C. D., & Szilagyi, J. E. (2014). A Faculty Toolkit for Formative Assessment in Pharmacy Education. *American Journal of Pharmaceutical Education*, *78*(9), 160. https://doi.org/10.5688/ajpe789160

Dmitrenko, N., Budas, I., Koliadych, Y., & Poliarush, N. (2021). Impact of Formative Assessment on Students' Motivation in Foreign Language Acquisition. *East European Journal of Psycholinguistics*, *8*(2), Article 2. https://doi.org/10.29038/eejpl.2021.8.2.dmi

Dunn, K., & Mulvenon, S. (2009). A Critical Review of Research on Formative Assessments: The Limited Scientific Evidence of the Impact of Formative Assessments in Education. *Practical Assessment, Research, and Evaluation*, *14*(1).

https://doi.org/10.7275/jg4h-rb87

Earl, L. M. (2012). *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning* (Second edition). Corwin.

Evans, D. J. R., Zeun, P., & Stanier, R. A. (2014). Motivating student learning using a formative assessment journey. *Journal of Anatomy*, *224*(3), 296–303. https://doi.org/10.1111/joa.12117

Frey, B. B., Schmitt, V. L., & Allen, J. P. (2019). Defining Authentic Classroom Assessment. *Practical Assessment, Research, and Evaluation*, *17*(1).

https://doi.org/10.7275/sxbs-0829

Gipps, C. V. (2012). *Beyond testing: Towards a theory of educational assessment* (Classic ed). Routledge.

Gotwals, A. W., & Birmingham, D. (2016). Eliciting, Identifying, Interpreting, and Responding to Students' Ideas: Teacher Candidates' Growth in Formative Assessment

Practices. *Research in Science Education*, *46*(3), 365–388. https://doi.org/10.1007/s11165-015-9461-2

Granberg, C., Palm, T., & Palmberg, B. (2021). A case study of a formative assessment practice and the effects on students' self-regulated learning. *Studies in Educational Evaluation*, *68*, 100955. https://doi.org/10.1016/j.stueduc.2020.100955

Gulikers, J., & Baartman, L. (2017). *Doelgericht professionaliseren. Formatieve toetspraktijken met effect! Wat doet de docent in de klas?* (No. 405-15–722; pp. 1–109). https://edepot.wur.nl/440065

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, *52*(3), 67–86. https://doi.org/10.1007/BF02504676

Han, C. (2018a). A longitudinal quantitative investigation into the concurrent validity of self and peer assessment applied to English-Chinese bi-directional interpretation in an undergraduate interpreting course. *Studies in Educational Evaluation*, *58*, 187–196. https://doi.org/10.1016/j.stueduc.2018.01.001

Han, C. (2018b). Conceptualizing and Operationalizing a Formative Assessment Model for English-Chinese Consecutive Interpreting. In *Quality Assurance and Assessment Practices in Translation and Interpreting* (pp. 89–111). IGI Global. https://doi.org/10.4018/978-1-5225-5225-3.ch004

Han, C. (2018c). Latent trait modelling of rater accuracy in formative peer assessment of English-Chinese consecutive interpreting. *Assessment & Evaluation in Higher Education*, *43*(6), 979–994. https://doi.org/10.1080/02602938.2018.1424799

Han, C. (2022). Interpreting testing and assessment: A state-of-the-art review. *Language Testing*, *39*(1), 30–55. https://doi.org/10.1177/02655322211036100

Han, C., & Fan, Q. (2020). Using self-assessment as a formative assessment tool in an English-Chinese interpreting course: Student views and perceptions of its utility. *Perspectives*, *28*(1), 109–125. https://doi.org/10.1080/0907676X.2019.1615516

Han, C., & Riazi, M. (2018). The accuracy of student self-assessments of English-Chinese bidirectional interpretation: A longitudinal quantitative study. *Assessment & Evaluation in Higher Education*, *43*(3), 386–398. https://doi.org/10.1080/02602938.2017.1353062

Han, C., & Zhao, X. (2021). Accuracy of peer ratings on the quality of spoken-language interpreting. *Assessment & Evaluation in Higher Education*, *46*(8), 1299–1313. https://doi.org/10.1080/02602938.2020.1855624

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Heritage, M. (2010). Assessment With and for Students. In *Formative Assessment: Making it Happen in the Classroom* (pp. 7–20). Corwin Press. https://doi.org/10.4135/9781452219493

Irons, A., & Elkington, S. (2021). *Enhancing Learning through Formative Assessment and Feedback* (2nd ed.). Routledge. https://doi.org/10.4324/9781138610514

Johnson, C. C., Sondergeld, T. A., & Walton, J. B. (2019). A Study of the Implementation of Formative Assessment in Three Large Urban Districts. *American Educational Research Journal*, *56*(6), 2408–2438. https://doi.org/10.3102/0002831219842347

Keeley, P. D. (2008). *Science Formative Assessment: 75 Practical Strategies for Linking Assessment, Instruction, and Learning* (1st edition). Corwin.

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The Effectiveness and Features of Formative Assessment in US K-12 Education: A Systematic Review. *Applied Measurement in Education*, *33*(2), 124–140.
https://doi.org/10.1080/08957347.2020.1732383

Lee, J. (2022). Comparing Student Self-assessment and Teacher Assessment in Korean-English Consecutive Interpreting: Focus on fidelity and language. *INContext: Studies in Translation and Interculturalism*, *2*(3). https://doi.org/10.54754/incontext.v2i3.27

Lee, S.-B. (2017). University students' experience of 'scale-referenced' peer assessment for a consecutive interpreting examination. *Assessment & Evaluation in Higher Education*, *42*(7), 1015–1029. https://doi.org/10.1080/02602938.2016.1223269

Lee, Y.-H. (2005). Self-assessment as an Autonomous Learning Tool in an Interpretation Classroom. *Meta : Journal Des Traducteurs / Meta: Translators' Journal*, *50*(4). https://doi.org/10.7202/019869ar

Leenknecht, M., Wijnia, L., Köhlen, M., Fryer, L., Rikers, R., & Loyens, S. (2021). Formative assessment as practice: The role of students' motivation. *Assessment & Evaluation in Higher Education*, *46*(2), 236–255. https://doi.org/10.1080/02602938.2020.1765228

Li, X. (2018). Self-assessment as 'assessment as learning' in translator and interpreter education: Validity and washback. *The Interpreter and Translator Trainer*, *12*(1), 48–67. https://doi.org/10.1080/1750399X.2017.1418581

Lyon, C. J., Nabors Oláh, L., & Caroline Wylie, E. (2019). Working toward integrated practice: Understanding the interaction among formative assessment strategies. *The Journal of Educational Research*, *112*(3), 301–314. https://doi.org/10.1080/00220671.2018.1514359

Mirek, J. (2020). The Interpreting Portfolio as a Self-Regulation Tool in Interpreter Training: Preliminary Findings from a Pilot Study. *Roczniki Humanistyczne*, *68*(10), 141–153. https://doi.org/10.18290/rh206810-11

Moher, D., Altman, D. G., Liberati, A., & Tetzlaff, J. (2011). PRISMA Statement. *Epidemiology*, *22*(1), 128. https://doi.org/10.1097/EDE.0b013e3181fe7825

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1), 143. https://doi.org/10.1186/s12874-018-0611-x

Murchan, D. (2017). *Understanding and applying assessment in education* (1st edition). SAGE Publications.

Nieminen, P., Hähkiöniemi, M., & Viiri, J. (2021). Forms and functions of on-the-fly formative assessment conversations in physics inquiry lessons. *International Journal of Science Education*, *43*(3), 362–384. https://doi.org/10.1080/09500693.2020.1713417

Orsmond, P., Merry, S., & Callaghan, A. (2004). Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International*, *41*(3), 273–290. https://doi.org/10.1080/14703290410001733294

Palm, T., Andersson, C., Boström, E., & Vingsle, C. (2017). A review of the impact of formative assessment on student achievement in mathematics. *Nordisk Matematikkdidaktikk*, *22*(3), 25–50. http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-133183

Palmer, E., & Devitt, P. (2014). The assessment of a structured online formative assessment program: A randomised controlled trial. *BMC Medical Education*, *14*(1), 8. https://doi.org/10.1186/1472-6920-14-8

Petticrew, M., & Roberts, H. (2005). *Systematic Reviews in the Social Sciences: A Practical Guide* (1st edition). Wiley-Blackwell.

Rr, S., Fox-Turnbull, W., Earl-Rinehart, K., & Calder, N. (2020). Development of formative assessment tool for a primary, technology classroom. *Design and Technology Education: An International Journal*, *25*(2), 101–116.

Stauffer, L. (2012). ASL Students' Ability to Self Assess ASL Competency. *Journal of Interpretation*, *21*(1). https://digitalcommons.unf.edu/joi/vol21/iss1/7

Su, W. (2021). Rubric-Based Self-Assessment of Chinese-English Interpreting. In J. Chen & C. Han (Eds.), *Testing and Assessment of Interpreting: Recent Developments in China* (pp. 67–84). Springer. https://doi.org/10.1007/978-981-15-8554-8_4

Univio, D. J., & Pérez, A. del P. (2019). Ipsative Assessment of Essay Writing to Foster Reflection and Self-Awareness of Progress. In E. White & T. Delaney (Eds.), *Handbook of Research on Assessment Literacy and Teacher-Made Testing in the Language Classroom:* (pp. 157–180). IGI Global. https://doi.org/10.4018/978-1-5225-6986-2

Vermeulen, J. A., & van der Kleij, F. M. (2012). Towards an Integrative Formative Approach of Data-Driven Decision Making, Assessment for Learning, and Diagnostic Testing. *Psychometrics in Practice at RCEC*, 162–185.

Wiliam, D. (2006). Formative Assessment: Getting the Focus Right. *Educational Assessment*, *11*(3–4), 283–289. https://doi.org/10.1080/10627197.2006.9652993

Wiliam, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Lawrence Erlbaum Associates.
https://doi.org/10.4324/9781315086545-3

Xiao, Y., & Yang, M. (2019). Formative assessment and self-regulated learning: How formative assessment supports students' self-regulation in English language learning. *System*, *81*, 39–49. https://doi.org/10.1016/j.system.2019.01.004

Zheng, C., Wang, L., & Chai, C. S. (2021). Self-assessment first or peer-assessment first: Effects of video-based formative practice on learners' English public speaking anxiety and performance. *Computer Assisted Language Learning*, 1–34.
https://doi.org/10.1080/09588221.2021.1946562