OmniaScience

Journal of Technology and Science Education

JOTSE, 2025 – 15(1): 186-203 – Online ISSN: 2013-6374 – Print ISSN: 2014-5349

https://doi.org/10.3926/jotse.2933

DESIGN AND VALIDATION OF A SELF-ASSESSMENT TOOL FOR STE(A)M TEACHERS IN CLIL CONTEXTS

María José Luelmo del Castillo , Elvira Izquierdo-Sánchez-Migallón ,
Virginia Vinuesa-Benítez , Nuria García-Manzanares[*]

Innovation and Educational Improvement (IMEI) Research Group. Rey Juan Carlos University (Spain)

mariajose.luelmo@urjc.es, elvira.izquierdo@urjc.es, virginia.vinuesa@urjc.es
[*]*Corresponding author: nuria.garcia@urjc.es*

Received May 2024
Accepted March 2025

Abstract

In bilingual education, the integration of STE(A)M (Science, Technology, Engineering, Art, and Mathematics) with CLIL (Content and Language Integrated Learning) establishes a dynamic learning environment wherein students concurrently develop scientific, technical, linguistic, and creative competencies. These methodologies cultivate essential 21st-century skills, including collaboration, communication, and critical thinking. This investigation aims to accomplish two primary objectives: to develop a self-assessment rubric for integrating STE(A)M and CLIL methodologies and to validate this rubric utilizing the Delphi method. This research employs a mixed-methods approach, combining qualitative and quantitative techniques to develop and validate a self-assessment rubric for integrating STE(A)M and CLIL methodologies in bilingual education. The study was conducted in three phases: initially, the research team developed and refined the rubric through a comprehensive literature review to ensure it was grounded in theoretical frameworks and pedagogical practices. Subsequently, a panel of specialists was convened using the Nominal Group Technique to collaboratively design the initial rubric. Finally, the Delphi method was employed to validate and refine the rubric through two rounds of expert consultation, involving twelve academic professionals. The final rubric comprises 23 items, structured into two key dimensions: (1) the integration of STE(A)M and CLIL and (2) the design and planning of the teaching and learning process. The results indicate that the CLIL-STE(A)M-SAT rubric is a reliable instrument for assessing the integration of these methodologies. It demonstrates potential for future research and practical application by primary school educators in bilingual STE(A)M-focused contexts.

Keywords – CLIL, STE(A)M, Bilingual education, Validation rubric, Self-assessment tool.

To cite this article:

Luelmo del Castillo, M.J., Izquierdo-Sánchez-Migallón, E., Vinuesa-Benítez, V., & García-Manzanares, N. (2025). Design and validation of a self-assessment tool for STE(A)M teachers in CLIL contexts. *Journal of Technology and Science Education,* 15(1), 186-203. https://doi.org/10.3926/jotse.2933

----

1. Introduction

The integration of education in STE(A)M disciplines, which encompass Science, Technology, Engineering, Art, and Mathematics, has become a critical approach in the 21st century and is increasingly evident at various educational levels. This trend is reflected in the proliferation of initiatives stressing the importance of this knowledge and its practical application in academic settings. In a similar vein, Content and Language Integrated Learning (CLIL), in which the content of subjects like science or technology is taught through an additional language as a medium of instruction other than the mother tongue, is recognized as a key methodology in the field of education. This is particularly evident in the expansion of bilingual programs over the past two decades in Europe and, particularly, in Spain, where STE(A)M disciplines are part of the content studied through the target language. Both the significance of STE(A)M education for a country's economic and social development and the potential of the CLIL methodology as a tool for language teaching and skill acquisition have been acknowledged by organizations such as the Organisation for Economic Cooperation and Development (OECD, 2023).

In the context of globalization and the so-called "fourth industrial revolution" (Schwab, 2016; World Economic Forum, 2016; OECD, 2023), language and STE(A)M disciplines are essential for accessing scientific and technological knowledge worldwide, fostering innovation driven by cultural diversity, and maintaining competitiveness in a globalized job market. This combination of skills expands career prospects and promotes advancement in a globally competitive and constantly evolving environment. In a similar vein, the European Commission (2014), through its Erasmus+ program, has made it clear that it aims to promote education in the areas of CLIL and STE(A)M by funding various projects that approach these subjects from different perspectives.

This paper is derived from the authors' participation in one of these initiatives: Erasmus+ *MiniOpenLab: Open Community and Hands-On Approach to Sustainable Development and STEM Education – An Innovative Approach*. This project aimed to motivate students not only to learn Mathematics and Science but also to envision themselves working in these fields and to develop attitudes and behaviours that align with the United Nations Sustainable Development Goals. It offered a unique approach that prioritized experiential learning and cooperation between science and technology organizations, businesses, and civil society. The primary goal was to establish and test an open community and hands-on approach to Sustainable Development and STE(A)M Education for students aged 6 to 12, involving the design and implementation of small STE(A)M labs open to the community (MiniOpenLabs), teacher training workshops, and the publication of a book containing the activities that could be conducted in the labs.

The plan brought together various educational institutions and organizations from three countries, including Spain, Portugal, and Greece. These institutions included higher education institutions, an engineering and product development centre, and schools. The project's objectives required the adaptation of activities to CLIL environments, which highlighted the need to modify the activities created for the MiniOpenLabs.

The methodological framework that was used required a major shift in how we teach and learn as it tried to integrate both approaches. As a result, it became clear that creating a tool to help educators assess their strengths and shortcomings when teaching scientific disciplines using an additional language was essential. Therefore, we set out two main objectives: to develop a self-assessment rubric for integrating STE(A)M and CLIL methodologies to support educators in evaluating and enhancing their teaching practices and to validate this rubric utilizing the Delphi method.

To achieve these aims, the study was conducted in three phases: Phase I consisted of developing and refining the initial draft of the rubric by the researchers' group before it was sent to a panel of experts for validation, conducting a comprehensive literature review to identify the key criteria and dimensions relevant to integrating STE(A)M and CLIL methodologies and select potential items, to ensure the rubric is firmly grounded in both theoretical frameworks and current pedagogical practices; Phase II focused on assembling a panel of specialists using the nominal group technique to collaboratively design the initial rubric and Phase III employed the Delphi method to validate and refine the rubric, ensuring its reliability and applicability for educational use.

2. Theoretical Framework

2.1. Content and Language Integrated Learning (CLIL)

Content and Language Integrated Learning (CLIL) is a dual educational approach that enables students to learn subjects, such as Science, History, or Technology, through an additional language, thereby not only enhancing students' language proficiency but also fostering deep learning in content areas (Coyle, Hood & Marsh, 2010; Meyer, 2010). A range of traits and tenets from learning theories, methodologies, and approaches, including new socio-cultural and constructivist perspectives, are compiled into its methodological foundation. However, what sets it apart is the integration of not only language and content but also cognitive and cultural aspects (Vinuesa, 2017).

The fundamental components of CLIL are encapsulated in the "4Cs-Framework," which comprises four dimensions (Coyle, 1999, 2006). As an integrated approach, it emphasizes the interrelationship between the four building blocks to optimize the benefits of combining learning (Content and Cognition) with language acquisition (Communication and Cultural understanding). By integrating Content, Cognition, Communication, and Culture, the 4Cs framework offers a comprehensive approach to maximize the benefits of CLIL.

The essence of content, or subject matter, encompasses more than the mere acquisition of knowledge and skills; learners actively construct their understanding of the discipline and their knowledge while acquiring specialized skills tailored to meet their individual needs (Lantolf, 2000; Vygotsky, 1978). Furthermore, this process is intricately intertwined with communication, which refers to learning through interaction, reconstructing content, engaging in cognitive processes, utilizing language within a learning environment, and acquiring language skills (Coyle et al., 2010). Effective communication in a CLIL environment involves not only the exchange of information but also the creation of opportunities for knowledge construction and meaning making (Meyer, 2016), which enhances language proficiency and deepens comprehension of the subject matter. Additionally, cognition, the process of learning and thinking, entails not only the acquisition and application of knowledge related to the subject matter being taught but also the cognitive abilities employed in this process.

The cognitive processes involved in CLIL extend beyond basic recall and comprehension. It emphasizes how students enhance their language proficiency in the target language while applying and analysing material connected to the curriculum topic they are studying (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich et al., 2001; Bloom, 1984; McGuinness, 1999; Dalton-Puffer, 2007, 2013). By engaging in these deeper cognitive activities, learners construct meaningful knowledge and develop a more profound understanding of the content. Moreover, culture plays a crucial role in building interculturality, involving the consideration of cultural contexts and values, the development of intercultural competence, and the enhancement of knowledge regarding global issues (Byram, Nichols & Stevens, 2001; Crozet & Liddicoat, 2000). Understanding and appreciating cultural contexts enriches the learning experience by providing diverse perspectives, enabling students to establish connections between their own experiences and those of others, and fostering empathy and intercultural competence. By integrating these elements–content, communication, cognition, and culture—educators can create a holistic learning experience that not only imparts knowledge but also fosters critical thinking, enhances effective communication, and promotes cultural awareness.

2.2. Integrated Science, Technology, Engineering, (Arts) and Mathematics Learning

The STE(A)M educational approach advocates for an integrated learning experience that encompasses Science, Technology, Engineering, and Mathematics. The primary objective of this approach is to cultivate critical thinking, creativity, collaboration, and innovation among students, empowering them to confront the complexities of today's and tomorrow's world. By immersing learners in these disciplines, STEM not only fosters a profound comprehension of their theoretical underpinnings but also encourages their practical application across diverse contexts, from everyday scenarios to cutting-edge scientific endeavours and technological advancements. Unlike traditional STEM frameworks, STE(A)M broadens the scope by incorporating the arts, acknowledging the pivotal role of creativity, artistic expression, and interdisciplinary exploration in addressing contemporary challenges. Through this holistic approach, STE(A)M equips students with the multifaceted skills and perspectives essential for success across various fields and professions.

In the realm of education, incorporating STE(A)M activities into the classroom setting often relies on the 5E's instructional model, originally conceived by Dr. Rodger W. Bybee in 1987, and based on the fundamental principles of constructivist learning, as outlined by Atkin and Karplus (1962), Bybee (2014; 2019) and DeBoer (1991). This model is rooted in the idea that students utilize their pre-existing ideas, knowledge, and prior experiences to construct new concepts and acquire fresh insights.

A substantial body of scholarly research has supported the efficacy of the 5E's model in STE(A)M education, underscoring its essential role as a structured framework that guides students through distinct stages of the learning process (Tezer, 2019; Conradty & Bogner, 2018, Conradty, Sotiriou, & Bogne, 2020; Anggraeni, 2021). The process commences with the Engage phase, in which the instructor presents a thought-provoking question, scenario, or challenge to stimulate curiosity and prompt students to draw upon their prior knowledge and experiential learning. This is followed by the Explore phase, where students actively engage in formulating hypotheses, devising research strategies, and collecting data to enhance their understanding through hands-on experience. In the subsequent Explain phase, learners demonstrate their comprehension by undertaking research projects and experiments, while the instructor provides scaffolded guidance to navigate the complexities of STE(A)M content. The Elaborate phase allows students to articulate their findings and consolidate their understanding of scientific concepts, and finally, the Evaluate phase facilitates reflective assessment, enabling both students and educators to identify strengths, address weaknesses, and tailor future instruction. This cyclical process of engagement, exploration, explanation, elaboration, and evaluation fosters a deeper, more meaningful understanding of STE(A)M concepts, ultimately preparing students for real-world problem-solving.

2.3. Integrating STE(A)M and CLIL

The topic of STEM and foreign language instruction represents a relatively nascent area of research. Extant literature indicates the benefits of integrating an additional language into the pedagogical framework of diverse disciplines. As postulated by Han (2015), Schoettler (2015), and Banergee (2016), foreign languages function as vehicles for expanding and enriching students' comprehension across various knowledge domains. The integration of STE(A)M (Science, Technology, Engineering, (Arts) and Mathematics) and CLIL (Content and Language Integrated Learning) can provide numerous significant advantages to the teaching and learning process. Both approaches adhere to an interdisciplinary, experiential educational methodology that emphasizes the development of skills such as critical thinking, problem-solving, creativity, collaboration, and effective communication. Figure 1 presents a conceptual framework for STE(A)M and CLIL integration, elucidating the core elements of both approaches.

Moreover, both models are centred on the practical application and contextualization of knowledge in authentic and meaningful contexts, facilitating students' comprehension of the practical relevance and utility of the concepts they are learning through projects and activities that connect new ideas to real-world situations (Han, 2015; Banergee, 2016; Nga, Lan & Nguyen, 2018; Martín-Cudero, Cid-Cid & Guede-Cid, 2024). Communication and interaction are also crucial elements of these educational approaches, as students are required to contribute their thoughts, results, and reflections, while teachers introduce situations or content designed to stimulate interest, generate curiosity, and activate the students' prior knowledge.

Consequently, learning becomes a continuous process of development, wherein students acquire new information, enhance their skills, and gain a deeper understanding of one or more disciplines (Dalton-Puffer, 2008; Gabillon & Ailincai, 2013; Haas, Grapin & Lee, 2018). Furthermore, both STE(A)M and CLIL heavily depend on the development of cognitive abilities, such as problem-solving and critical thinking, emphasizing the capacity to link, evaluate, synthesize, and apply knowledge in diverse contexts, rather than merely memorizing it. Ultimately, culture is a significant factor in either approach when addressing the global issues of the twenty-first century. Students must be able to function respectfully in diverse cultural contexts. Making decisions, participating in civic life, and accepting and valuing cultural diversity are all integral components of understanding culture. These elements are essential as they support the development and reinforcement of students' intercultural competencies and global consciousness.

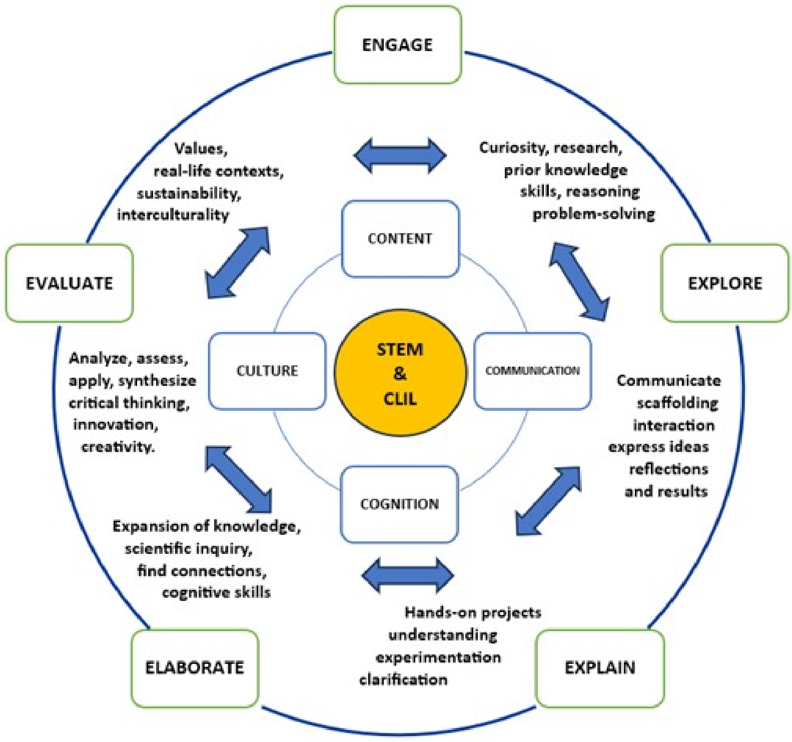

Figure 1. STE(A)M+CLIL integration framework

3. Research Methodology

The research followed a mixed-methods approach, combining qualitative and quantitative techniques to develop and validate the self-assessment rubric for integrating STE(A)M and CLIL methodologies. In the qualitative phase, the research team began by conducting a comprehensive literature review and using the nominal group technique (McMillan, King & Tully, 2016; Manera, Hanson, Gutman & Tong, 2019; Rodríguez-Pavón, Morales Salas, Infante-Moro & Infante-Moro, 2024) to develop the initial rubric, identifying key dimensions and criteria based on expert input and pedagogical theory (Creswell & Plano-Clark, 2017; Donnelly, 2010; Ponce & Pagán-Maldonado, 2015). This process provided in-depth insights into effective integration of STE(A)M and CLIL. In the quantitative phase, the Delphi method, a rigorously validated research approach that has been extensively documented in methodological literature over the past five decades (Landeta, 1999; Hasson, Keeney & McKrenna, 2000; Varela-Ruiz, Díaz-Bravo & García-Durán, 2012) was applied, involving multiple rounds of expert consultation to validate the rubric's items. This phase measured expert consensus on each rubric item, ensuring both reliability and content validity (Linstone & Turoff, 2002; Escobar-Pérez & Cuervo-Martínez, 2008). The mixed-methods approach enabled the research to leverage both expert insights (qualitative) and statistical measures (quantitative), ensuring that the final rubric was both conceptually sound and practically applicable in bilingual STE(A)M classrooms (Creswell & Plano-Clark, 2017).

3.1. Methodology in the Design of the Instrument

The literature review conducted on CLIL-STE(A)M teacher competencies led to the configuration of the first structure of the rubric around two fundamental dimensions that cover the spectrum of analysis of teaching practice: the integration of the basic elements of both approaches (Engage, Explore, Explain, Elaborate, Evaluate and Content, Communication, Cognition, and Culture) and the design and planning of the teaching-learning process (Bertaux, Coonan, Frigols-Martin & Mehisto, 2010; Bybee, 2009; Coyle 2007; Duran & Duran, 2004; Gresnigt, Taconis, van Keulen, Gravemeijer, & Baartman, 2014; Meyer, 2010; Trevallion & Trevallion, 2020).

The coordinating team selected a group of specialists who would elaborate the items associated with each of these dimensions to construct the first draft of the CLIL-STE(A)M SAT rubric. The group comprised English language teachers with CLIL training and experience in CLIL-STE(A)M projects. The first draft was constructed following a process based on the design-based methodology (DBR methodology) described by Brown (1992). The process was structured into three phases: in the first phase, a literature review was carried out; in the second phase, the dimensions were defined and the items that would make up the rubric were elaborated; and in the third phase, the instrument was validated through the Delphi method with a panel of experts. This resulted in a final rubric that retained the structures of the two initial dimensions. The first with 14 indicators and the second with 13 were assigned a numerical score from 1 to 5, with 1 being not relevant or unclear and 5 being very relevant or very clear.

The indicators grouped in dimension 1 focus on the effective integration of STE(A)M and CLIL, from promoting active participation in STEM subjects and second language use to adapting pedagogy, with the overall goal of facilitating equitable, communicative, and contextualized learning in STE(A)M disciplines (Aguilera, García-Yeguas, Perales-Palacios & Vílchez-González, 2022; Borg & Edmett, 2019; Evnitskaya & Dalton-Puffer, 2023).

Dimension 2 includes indicators of the design and planning of the teaching-learning process. It aims to help the teacher check if he/she can provide adequate linguistic support, adapt educational materials, create a good collaborative learning environment, encourage critical reflection, connect STE(A)M content with previous experiences, and offer multiple learning modalities. In addition, this dimension highlights the integration of communication strategies (concept maps, substitution tables, glossaries, etc.) that help students use a second language to understand and express knowledge in the context of STE(A)M disciplines. Considering these aspects ensures that students develop language skills and STE(A)M competencies simultaneously and effectively (Coyle, 2005, 2008; Mehisto, 2010; Morrison, 2006; Roth & Bogner, 2024).

3.2. Validation of the Instrument: The Delphi Method

The second phase was aimed at validating the content of the draft rubric generated previously. The Delphi method, usually used in Educational Sciences, as it provides clarity around a problem through "a communicative process of various experts organized in a panel group" (López-Gómez, 2018, p. 21).

The instruments used to collect information were questionnaires for assessing dimensions and items, including opinion comments. The process of collecting the information was planned with flexible, albeit limited, time to allow the participation of a group located in different places and at different times.

3.3. Selection and Composition of the Panel of Experts

To ensure the rigor and validity of the study's findings and given the initial dimensions, the research team initiated a systematic process to assemble a diverse panel of experts whose combined experiences in STE(A)M education and CLIL pedagogy would provide balanced and insightful perspectives. The expert panel was constituted through a systematic process that prioritized representativeness over quantity, drawing on clearly defined inclusion criteria (López-Gómez, 2018). To ensure a balanced combination of current knowledge and diverse perceptions regarding the teaching of STE(A)M subjects through a second language using the CLIL approach, the selection criteria mandated the inclusion of both primary school teachers experienced in STE(A)M instruction within CLIL environments and teacher trainers. Although the Delphi method does not require a statistically representative sample nor a fixed number of participants (Steurer, 2011), the optimal panel size typically depends on the research problem and available resources (Powell, 2003). Early studies suggested that larger panels might enhance reliability (Dalkey, Rourke, Lewis, & Snyder, 1972); however, subsequent research has indicated that homogeneous panels, such as the one in this study, generally require between 10 and 15 experts (Skumolski, Hartman & Krahn, 2007; Delbecq, van de Ven & Gustafson, 1975). In this context, the selection of 12 experts is justified by a flexible, context-dependent approach (Landeta, 1999) that aligns with the research objectives while ensuring a comprehensive range of practical and pedagogical perspectives.

Twelve experts, who met the criteria mentioned above, were invited. The invitation included questions about their professional experience, a description of their teaching and research activities, and a self-evaluation based on the objective of the project. The 12 invitations were accepted and, therefore, the group of experts was formed. It encompassed seven teacher trainers in Spanish universities, three teachers of primary and secondary education in bilingual schools in the community of Madrid who teach STE(A)M subjects, and two who belong to or have belonged to both bodies. The group comprised three men and nine women, of whom eight were doctors, three were graduates, and one was a doctoral candidate. All of them had more than five years of teaching experience.

Thirty-seven percent of experts had university teaching experience, 19% had teaching experience in primary education, 26% in secondary education, and 19% in early childhood education. 58% of the experts teach or have taught in a bilingual school in the Community of Madrid.

3.4. Procedure for Applying the Delphi Method

The research team oversaw collecting the information provided by the experts through an iterative process in which e-mail was used as a means of communication. This process was structured into two rounds, a number considered adequate to ensure convergence (Linstone & Turoff, 1975). The experts were informed of this from the outset so that the panel was aware of the dimensions of the study and the degree of commitment expected on their part (López-Gómez, 2018).

The first round began with both quantitative and qualitative assessments of the items comprising the instrument. Regarding the quantitative assessment, experts were asked to rate the relevance and clarity of each item on a 5-option Likert scale, with 1 being not relevant or unclear and 5 being very relevant or clear.

Since open questions of a qualitative nature generate valuable information at this point (Hung, Altschuld, & Lee, 2008), they were included in this first round. The comments that the experts would provide would be used later to formulate some of the statements of the next round, as suggested by López-Gómez (2018).

The evaluation form for the second round of consultation was prepared based on an analysis of the results of the first round. The objective was to analyse the clarity of the reformulated items using the comments made by the experts on the open-ended questions. The experts evaluated the items on a 5-choice Likert scale. The results of this round established the consensus that would lead to the final instrument.

The statistical package IBM SPSS (Statistical Package for Social Sciences), version 28.0.1.0 for Windows, was used for data processing. As the results were obtained, decisions were made based on the degree of consensus reached (Martínez-Piñeiro, 2003). To establish this consensus, two analyses were conveyed: Kendall's coefficient to determine the level of agreement of the answers given by the experts and a descriptive analysis that provided the mean, standard deviation, and percentiles of the answers.

4. Results

4.1. Results of Concordance Analysis

Kendall's coefficient of concordance (W) was used to perform concordance tests (Table 1). It revealed that although the agreement between rankings was relatively low, there was a significant association between them. This suggests that there is statistically significant agreement among the experts regarding the relevance and clarity of the items, which is a valuable finding as it indicates that there is consensus among the experts despite individual differences in interpretation and judgment.

	Construct validity (relevance)	Content validity (clarity)
Kendall's W	0,27	0,24
Chi-square	84,44	74,53
Degree of freedom (df)	26	26
Asymp. Sig (p-value)	<,001	<,001

Table 1. Agreement between construct and content validity

4.2. Evaluation of Construct and Content Validity

A descriptive analysis is used to characterize Construct and Content validity, and a criterion based on percentiles is implemented to determine the permanence of the items. Additionally, a conditional relationship between relevance and clarity is established, where only the clarity of the items considered relevant is evaluated. Those reaching, at least, 80% relevance, but lower clarity percentage, have not been removed, but reformulated.

The criterion proposed by Pozo-Llorente, Gutiérrez-Pérez and Rodríguez-Sabiote (2007) is employed, but instead of applying it to the mean, the 80% percentile is used as the cut-off point. Items that did not reach this percentile in relevance or clarity were eliminated. Furthermore, it is established that the assessment of clarity only proceeds for those items that have been considered relevant by the group of experts. This implies that non-relevant items would not be analysed in terms of their clarity.

This study employs a rigorous approach to assessing the relevance and clarity of items using statistical methods and based on the ratings of the expert group. The percentile criterion and the conditionality between relevance and clarity are novel elements that add robustness to the analysis and allow for the precise identification of items that do not meet the established standards.

4.2.1. Construct Validity

To evaluate construct validity, experts were consulted regarding the relevance of the items. As indicated in Table 2, two items needed to be removed from the original questionnaire: item 5 (*To ensure equitable access of STE(A)M subject matter, in my STE(A)M-CLIL lessons there is evidence of explicit teaching of academic language (CALP))*, which only 75% of the experts deemed highly relevant, and item 12 (*I can adapt social and academic styles of communication (BICS / CALP) based on the requirements of a specific subject content)* which achieved only 58% acceptance.

4.2.2. Content Validity

The experts' assessments of item clarity were utilized to evaluate content validity. The results of the first round, outlined in Table 2, indicated that 6 items required reformulation. Item 20 had low acceptability, as fewer than 42% of the experts considered it to be clearly formulated. Similarly, items 13, 22, 26, and 27 also had low acceptability, with only 50% of the experts deeming them clear. Item 24 was deemed moderately acceptable by 67% of the experts, who assessed it as being as clear as possible. The reformulation of the items was performed considering the experts' comments and to justify the score obtained. This was done by pinpointing areas for improvement based on the opinions collected during the first round.

Items	Validity	Mean	Standard deviation	Percentage	Permanence
1. I try to get the students engaged in STE(A)M subjects and in the habit of using a second language.	Construct	4,75	0,452	100%	Yes
	Content	4,08	0,9	83%	Yes
2. I try to encourage students to work together to help them think freely, try alternatives, and discuss them with others keeping instruction to a minimum.	Construct	4,5	0,674	92%	Yes
	Content	3,83	1,03	92%	Yes
3. I incorporate a variety of communication strategies that support students with the language needed to access the content.	Construct	4,92	0,289	100%	Yes
	Content	4,42	0,9	92%	Yes
4. I carry out activities that develop subject communication skills (language for predicting, hypothesizing, cause/effect, describing a process, etc.).	Construct	4,5	0,522	100%	Yes
	Content	4,5	0,522	100%	Yes
5. To ensure equitable access of STE (A)M subject matter, in my STE (A)M-CLIL lessons there is evidence of explicit teaching of academic language (CALP).	Construct	4	0,953	75%	No
	Content				
6. To ensure equitable access of STE (A)M subject matter, in my STE (A)M-CLIL lessons there is evidence of explicit teaching of text structure.	Construct	4,5	0,674	92%	Yes
	Content	4,42	0,793	83%	Yes
7. To ensure equitable access of STE(A)M subject matter, in my STE(A)M-CLIL lessons there is evidence of explicit teaching of content specific vocabulary and/or language features.	Construct	5	0	100%	Yes
	Content	4,83	0,389	100%	Yes
8. I understand the significance and meaning of the STE(A)M educational approach: Engage, Explore, Explain, Elaborate and Evaluate.	Construct	4,5	0,674	92%	Yes
	Content	4,08	0,996	92%	Yes
9. I understand the significance and meaning of the CLIL educational approach: Content, Communication, Cognition and Culture.	Construct	4,42	0,669	92%	Yes
	Content	4	0,953	92%	Yes
10. I consider the cultural implications of the STE(A)M-CLIL content.	Construct	4,42	0,669	92%	Yes
	Content	4,33	0,651	92%	Yes
11. I can use the target language for: providing explanations, presenting information, issuing instructions, elucidating, and confirming comprehension, in order to adapt it to the students' level.	Construct	4,83	0,389	100%	Yes
	Content	4,67	0,651	92%	Yes
12. I can adapt social and academic styles of communication (BICS / CALP) based on the requirements of a specific subject content.	Construct	3,58	1,165	58%	No
	Content				
13. I consistently employ culturally responsive pedagogy, as demonstrated by inclusivity towards diverse cultural and linguistic backgrounds, and consideration for multiple points of view.	Construct	4,5	0,905	92%	Yes
	Content	3,75	1,215	50%	Reformulate
14. I am aware of the scientific topics as relevant for the culture / community.	Construct	4,5	0,522	100%	Yes
	Content	4,17	0,835	92%	Yes
15. I incorporate a variety of communication strategies that support students with the language needed to access the content.	Construct	4,83	0,389	100%	Yes
	Content	4,83	0,389	100%	Yes
16. I plan in advance the language needed for the STE(A)M subject content and provide students with useful examples to communicate subject knowledge.	Construct	4,75	0,452	100%	Yes
	Content	4,67	0,492	100%	Yes

Items	Validity	Mean	Standard deviation	Percentage	Permanence
17. I provide students with multiple opportunities to reflect and discuss on higher order questions (connect ideas, analyse, organize, and interpret information, compare, and contrast information, analyse, evaluate, etc.). Ex. What would happen if…? What would happen if …? What's a different way to carry out the experiment? How would you have solved the problem? What plan would you carry out if this happen to you?	Construct	4,5	0,674	92%	Yes
	Content	4,5	0,674	92%	Yes
18. I use a variety of instructional strategies, materials, and input that clearly enhance all students' participation in the STE(A)M-CLIL lessons.	Construct	5	0	100%	Yes
	Content	4,83	0,389	100%	Yes
19. I create opportunities for students to connect STE(A)M content to previous lessons to deepen students' understanding of STE(A)M concepts.	Construct	4,67	0,492	100%	Yes
	Content	4,75	0,452	100%	Yes
20. I provide students multiple modalities through which to practice and represent the STE(A)M content and these modalities benefit all students. All students' needs and abilities are considered.	Construct	4,58	0,669	92%	Yes
	Content	3,67	1,231	42%	Reformulate
21. In my STEM-CLIL lessons, there is evidence of explicit teaching of academic language (CALP), text structure, content specific vocabulary and/or language features to ensure equitable access of STE(A)M subject matter.	Construct	4,67	0,492	100%	Yes
	Content	4,92	0,289	100%	Yes
22. In my STE(A)M-CLIL lessons, there is evidence of content and language scaffolds (suitable tasks, students' backgrounds, and prior knowledge, variety of supports- prompts, hints, visual organizers, substitution tables, glossaries, word banks, sentence frames/stems) to help students access the subject content.	Construct	4,5	0,674	92%	Yes
	Content	3,83	0,937	50%	Reformulate
23. In my STE(A)M-CLIL lessons, I use innovative equipment and technology.	Construct	4,5	0,674	92%	Yes
	Content	4,33	0,651	92%	Yes
24. I choose educational materials and adapt them according to the students' linguistic competence, whether by structuring or adjusting them as necessary.	Construct	4,83	0,389	100%	Yes
	Content	4,08	1,084	67%	Reformulate
25. I choose educational materials and adapt them according to the students' cognitive skills, whether by structuring or adjusting them as necessary.	Construct	4,83	0,389	100%	Yes
	Content	4,5	0,674	92%	Yes
26. I design balanced formative and summative assessment tools measuring uptake in both language and content.	Construct	4,92	0,289	100%	Yes
	Content	3,92	0,996	50%	Reformulate
27. In my STE(A)M-CLIL lessons I make sure that the classroom atmosphere fosters a collaborative learning environment.	Construct	5	0	100%	Yes
	Content	3,92	1,165	50%	Reformulate

Table 2. Results on construct and content validity

Again, Kendall's coefficient of concordance was used to assess the agreement among raters in the second round (Table 3). The results indicated a weak ($W = 0.206$) but statistically significant ($p = 0.030$) level of agreement, suggesting that the observed concordance was not due to random chance. The chi-square statistic (12.338) with 5 degrees of freedom was used to test whether the observed concordance differed significantly from what would be expected by chance. With a p-value of 0.030, which is below the conventional threshold of 0.05, the agreement among raters was found to be statistically significant. This suggests that the observed concordance is unlikely to be due to random chance, providing evidence that evaluators share a common assessment pattern to some extent.

	Content validity (clarity)
Kendall's W	0.206
Chi-square	12.338
Degree of freedom (df)	5
Asymp. Sig (p-value)	0.030

Table 3. Agreement among the experts on content validity

Table 4 shows the results of the second round of content validity assessment. The group of experts was asked to evaluate the clarity of the six items that were reformulated following the initial round. The revalidation was conducted using the same 5-point Likert scale used in the first round. The results of the content revalidation indicate that only two of the six reformulated items (items 13 and 27) meet the clarity criteria. Only 66.66% of the experts considered these items to be formulated 'very clearly', which is below the minimum threshold for acceptance. Given these findings, it is recommended that these items be removed from the measuring instrument. Despite being reformulated, these items have failed to achieve a satisfactory level of clarity, which could impact the validity and reliability of the instrument.

Items	Reformulated items	Mean	Standard Deviation	Percentage	Permanence
13. I consistently employ culturally responsive pedagogy, as demonstrated by inclusivity towards diverse cultural and linguistic backgrounds, and consideration for multiple points of view.	13. I use culturally inclusive teaching strategies that take into account diverse cultural and linguistic backgrounds and points of view.	4,17	1,115	66,66	No
20. I provide students multiple modalities through which to practice and represent the STE(A)M content and these modalities benefit all students. All students' needs and abilities are considered.	20. I use STE(A)M-related multimodal materials (texts, graphs, satellite images, digital materials, etc) so that all students' needs and abilities are considered.	4,83	0,389	100	Yes
22. In my STE(A)M-CLIL lessons, there is evidence of content and language scaffolds (suitable tasks, students' backgrounds, and prior knowledge, variety of supports-prompts, hints, visual organizers, substitution tables, glossaries, word banks, sentence frames/stems) to help students access the subject content.	22. In my STE(A)M-CLIL lessons, there is evidence of content and language scaffolds (visual organizers, substitution tables, glossaries, word banks, sentence frames/stems) to help students access the subject content.	4,58	0,996	83,33	Yes
24. I choose educational materials and adapt them according to the students' linguistic competence, whether by structuring or adjusting them as necessary.	24. I select, adapt, and adjust educational materials considering the students' language competence.	4,92	0,289	100	Yes

Items	Reformulated items	Mean	Standard Deviation	Percentage	Permanence
26. I design balanced formative and summative assessment tools measuring uptake in both language and content.	26. I design formative and summative assessment tools that measure language and content learning in a balanced way.	4,58	0,669	91,66	Yes
27. In my STE(A)M-CLIL lessons I make sure that the classroom atmosphere fosters a collaborative learning environment.	27. In my STE(A)M-CLIL lessons I create a collaborative learning environment encouraging group and teamwork and meaningful group interactions in which students give their opinions and exchange ideas and critical feedback from their peers.	4,17	1,115	66,66	No

Table 4. Results of the content revalidation of the reformulated items

4.2.3. Summary of Item Validation Process

A total of 27 initial items were assessed for relevance and clarity in the item validation process. The final rubric, comprised of 23 validated items, is a robust and dependable tool for the self-assessment of teachers who teach STE(A)M subjects through CLIL in bilingual education. The final version of the rubric is organized as follows: 1) purpose of the questionnaire and instructions; 2) respondent identification and information on experience as a CLIL/STE(A)M teacher; 3) 23 items, divided into two sections or dimensions, designed for self-assessment when planning teaching in the CLIL/STE(A)M environment. The final version of the questionnaire, as well as the link to access it, are presented below: https://forms.office.com/e/3NUV0f50SY.

TEAM-CLIL SELF-ASSESSMENT TOOL (STEAM-CLIL SAT)

This questionnaire has been developed by considering the quality criteria of CLIL and/or STEAM experiences from various experts.

The intention of this questionnaire is to assist teachers in planning their CLIL/STE(A)M teaching. To achieve this, they must evaluate whether each item is never, almost never, sometimes, almost always, or always fulfilled. The estimated completion time for this task is 10 minutes.

DATA OF THE PERSON COMPLETING THE QUESTIONNAIRE

1. Education
2. Institution or organization.
3. Country of institution or organization.
4. Years of professional experience:
   - Between 1 and 5
   - Between 6 and 10
   - More than 10
   - None
5. Educational levels at which you have taught or are teaching:
   - Early Childhood Education
   - Primary Education
   - Secondary Education and/or Baccalaureate
   - Higher Education
6. Do you teach, or have you taught in a bilingual school?
   - Yes
   - No
7. What is your level of knowledge of CLIL (where 1 is none and 5 is very much)?
8. What is your level of knowledge of STE(A)M subjects (where 1 is none and 5 is a lot)?
9. What is your experience with CLIL (where 1 is none and 5 is a lot)?

FIRST DIMENSION: STE(A)M AND CLIL INTEGRATION

Please, self-assess each item (1 = Never; 2=Rarely; 3=Sometimes; 4=Often; 5=Always)

1. I understand the significance and meaning of the CLIL educational approach: Content, Communication, Cognition and Culture.
2. I understand the significance and meaning of the STE(A)M educational approach: Engage, Explore, Explain, Elaborate and Evaluate.
3. I try to get the students engaged in STE(A)M subjects and in the habit of using a second language.
4. I try to encourage students to work together to help them think freely, try alternatives, and discuss them with others, keeping instruction to a minimum.
5. I incorporate a variety of communication strategies that support students with the language needed to access the content.
6. I carry out activities that develop subject communication skills (language for predicting, hypothesizing, cause/effect, describing a process, etc.).
7. To ensure equitable access of STE(A)M subject matter, in my STE(A)M-CLIL lessons there is evidence of explicit teaching of text structure.
8. To ensure equitable access of STE(A)M subject matter, in my STE(A)M-CLIL lessons there is evidence of explicit teaching of content specific vocabulary and/or language features.
9. I consider the cultural implications of the STE(A)M-CLIL content.
10. I can use the target language for: providing explanations, presenting information, issuing instructions, elucidating, and confirming comprehension, in order to adapt it to the students' level.
11. I am aware of the scientific topics as relevant for the culture / community.

SECOND DIMENSION: DESIGN AND PLANNING OF THE TEACHING AND LEARNING PROCESS

Please, self-assess each item (1 = Never; 2=Rarely; 3=Sometimes; 4=Often; 5=Always)

12. I incorporate a variety of communication strategies that support students with the language needed to access the content.
13. I plan in advance the language needed for the STE(A)M subject content and provide students with useful examples to communicate subject knowledge.
14. I provide students with multiple opportunities to reflect and discuss on higher order questions (connect ideas, analyse, organize, and interpret information, compare, and contrast information, analyse, evaluate, etc.). Ex. What would happen if…? What is a different way to carry out the experiment? How would you have solved the problem? What plan would you carry out if this happened to you?
15. I use a variety of instructional strategies, materials, and input that clearly enhance all students' participation in the STE(A)M-CLIL lessons.
16. I create opportunities for students to connect STE(A)M content to previous lessons to deepen students' understanding of STE(A)M concepts.
17. I use STEAM-related multimodal materials (texts, graphs, satellite images, digital materials, etc) so that all students' needs and abilities are considered.
18. In my STEM-CLIL lessons, there is evidence of explicit teaching of academic language (CALP), text structure, content specific vocabulary and/or language features to ensure equitable access of STE(A)M subject matter.
19. In my STE(A)M-CLIL lessons, there is evidence of content and language scaffolds (visual organizers, substitution tables, glossaries, word banks, sentence frames/stems) to help students access the subject content.
20. In my STE(A)M-CLIL lessons, I use innovative equipment and technology.
21. I select, adapt, and adjust educational materials considering the students' language competence.
22. I choose educational materials and adapt them according to the students' cognitive skills, whether by structuring or adjusting them as necessary.
23. I design formative and summative assessment tools that measure language and content learning in a balanced way.

5. Conclusions

Our work began by verifying existing guidelines when designing STE(A)M activities in bilingual environments that require CLIL methodology and the challenges this poses for teachers in these contexts, which are becoming increasingly common in our classrooms. We consider that a self-assessment tool that helps teachers become aware of the essential aspects to bear in mind when designing these activities could be beneficial as a tool for reflection leading to an improvement in their teaching practice. Systematic self-evaluation plays a key role in ensuring that teaching practices are continuously improved (Manea, 2021). Thus, this paper was focused on creating a self-assessment rubric for teachers by forming a nominal group and to improve and validate it using the Delphi method. After reviewing the relevant literature, the nominal group developed the first draft of the STE(A)M-CLIL SAT rubric containing 2 dimensions and 27 items. The two validation rounds gave rise to the elimination of four items, the rewording of 6, and the reordering of two. As a result, the final instrument consists of two dimensions and 23 items.

A limitation of this work is the number of experts who participated in the validation process. Although we adhere to the recommendations of Skumolski et al. (2007), who suggest that a homogeneous group of experts typically requires a sample size between 10 and 15, and López-Gómez (2018), who recommends that the sample should always exceed 7 participants, a larger number of experts would have enriched the process. Therefore, it is proposed that future research should increase the number of experts.

The final purpose of the present investigation is to introduce the STE(A)M-CLIL SAT tool, which is designed for use by primary educators in bilingual institutions that offer STE(A)M subjects in English. This tool aims to facilitate self-reflection and identify areas for professional development among teachers. It is essential to recognize that teacher self-assessment plays a crucial role in enhancing individual teaching techniques and benefiting the educational community.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest for the research, authorship, and/or publication of this article.

Funding

This research forms part of the Erasmus+ project ref. 2020-1-ES01-KA201-082706, financed by the European Commission.

References

Aguilera, D., García-Yeguas, A., Perales-Palacios, F.J., & Vílchez-González, J.M. (2022). Diseño y validación de una rúbrica para la evaluación de propuestas didácticas STEM (RubeSTEM). *Revista interuniversitaria de formación del profesorado: continuación de la antigua Revista de Escuelas Normales,* 36(97), 1. https://doi.org/10.47553/rifop.v97i36.1.92409

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R. et al. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives.* New York: Longman.

Anggraeni, R.E. (2021). The analysis of the development of the 5E-STEAM learning model to improve critical thinking skills in natural science lesson. *Journal of Physics: Conference Series,* 1832(1), 012050). https://doi.org/10.1088/1742-6596/1832/1/012050

Atkin, J.M., & Karplus, R. (1962). Discovery or invention? *The Science Teacher,* 29(5), 45-51.

Banergee, S. (2016). STEM-A tool for teaching and learning a foreign language. In *Proceedings of IEEE Integrated STEM Education Conference (ISEC)* (1618). Piscataway: IEEE. https://doi.org/10.1109/ISECon.2016.7457523

Bertaux, P., Coonan, C.M., Frigols-Martin, M.J., & Mehisto, P. (2010). *CLIL Competence Grid.* CLIL Cascade Network.

Bloom, B.S. (1984). *Taxonomy of educational objectives.* Boston: Allyn and Bacon.

Borg, S., & Edmett, A. (2019). Developing a self-assessment tool for English language teachers. *Language Teaching Research,* 23(5), 655-679. https://doi.org/10.1177/1362168817752543

Brown, A.L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences,* 2(2), 141-178. https://doi.org/10.1207/s15327809jls0202_2

Bybee, R.W. (2009). *The BSCS 5E instructional model and 21st century skills.* Colorado Springs, CO: BSCS (Biological Sciences Curriculum Study).

Bybee, R.W. (2014). The BSCS 5E instructional model: Personal reflections and contemporary implications. *Science and Children,* 51(8), 10-13. https://doi.org/10.2505/4/sc14_051_08_10

Bybee, R.W. (2019). Using the BSCS 5E instructional model to introduce STEM disciplines. *Science and Children,* 56(6), 8-12. https://doi.org/10.2505/4/sc19_056_06_8

Byram, M., Nichols, A., & Stevens, D. (Eds.) (2001). *Developing Intercultural Competence in Practice.* Toronto: Multilingual Matters. https://doi.org/10.21832/9781853595356

Conradty, C., & Bogner, F.X. (2018). From STEM to STEAM: How to Monitor Creativity. *Creativity Research Journal,* 30(3), 233-240. https://doi.org/10.1080/10400419.2018

Conradty, C., Sotiriou, S.A., & Bogner, F.X. (2020). How creativity in STEAM modules intervenes with self-efficacy and motivation. *Education Sciences,* 10(3), 70. https://doi.org/10.3390/educsci10030070

Coyle, D. (1999). Supporting students in content and language integrated contexts: planning for effective classrooms. In Masih, J. (Ed.), *Learning through a foreign language – models, methods, and outcomes* (46-62). London: Centre for Information on Language Teaching and Research (CILT).

Coyle, D. (2005). *CLIL planning tools for teachers. 4Cs curriculum guidance. 3As lesson planning tool. Matrix audit tool for tasks & materials.* Nottingham: University of Nottingham.

Coyle, D. (2006): Content and Language Integrated Learning – Motivating Learners and Teachers. *The Scottish Language Review,* 13, 1-18.

Coyle, D. (2007). Content and Language Integrated Learning: Towards a Connected Research Agenda for CLIL Pedagogies. *International Journal of Bilingual Education and Bilingualism,* 10(5), 543-562. https://doi.org/10.2167/beb459.0

Coyle, D. (2008). CLIL–A Pedagogical Approach from the European Perspective. In Hornberger, N.H. (Eds.), *Encyclopedia of Language and Education* (1200-1214). Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30424-3_92

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and Language Integrated Learning.* Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009024549

Creswell, J.W., & Plano-Clark, V.L. (2017). *Designing and conducting mixed methods research* (3rd ed.). Sage Publications.

Crozet, C., & Liddicoat, A.J. (2000). Teaching culture as an integrated part of language: Implications for the aims, approaches, and pedagogies of language. *Australian Review of Applied Linguistics. Supplement Series,* 14(1), 1-22. https://doi.org/10.1075/aralss.14.01cro

Dalkey, N., Rourke, D.L., Lewis, R., & Snyder, D. (Eds.) (1972). *Studies in the quality of life: Delphi and decision-making*. Lexington, MA: Lexington Books.

Dalton-Puffer, C. (2007). *Discourse in Content-and-Language-Integrated Learning (CLIL) classrooms.* Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.20

Dalton-Puffer, C. (2008). Outcomes and processes in Content and Language Integrated Learning (CLIL): current research from Europe. In Delanoy, W., & Volkmann, L. (Eds.), *Future Perspectives for English Language Teaching* (139-157). Heidelberg: Carl Winter.

Dalton-Puffer, C. (2013). A construct of cognitive discourse functions for conceptualizing content-language integration in CLIL and multilingual education. *European Journal of Applied Linguistics,* 1(2), 216-253. https://doi.org/10.1515/eujal-2013-0011

DeBoer, G.E. (1991). *A History of Ideas in Science Education: Implications for Practice. Teachers College.* Columbia University: New York, NY, USA. https://doi.org/10.47553/rifop.v97i36.1.92409

Delbecq, A.L., van de Ven, A.H., & Gustafson D. (1975). *Group Techniques for Program Planning; a guide to nominal group and Delphi processes.* Glenview IL: Scott Foresman and Company.

Donnelly, R. (2010). The role of rubrics in student assessment and the development of rubrics in an educational setting. *Assessment & Evaluation in Higher Education,* 35(2), 217-227.

Duran, L.B., & Duran, E. (2004). The 5E Instructional Model: A Learning Cycle Approach for Inquiry-Based Science Teaching. *Science Education Review,* 3(2), 49-58.

Escobar-Pérez, J., & Cuervo-Martínez, A. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición,* 6, 27-36.

European Commission (2014). Improving the effectiveness of language. learning: CLIL and computer assisted language learning. *Education and Training.* Available at: https://ec.europa.eu/assets/eac/languages/library/studies/clil-call_en.pdf

Evnitskaya, N., & Dalton-Puffer, C. (2023). Cognitive discourse functions in CLIL classrooms: Eliciting and analysing students' oral categorizations in science and history. *International Journal of Bilingual Education and Bilingualism,* 26(3), 311-330. https://doi.org/10.1080/13670050.2020.1804824

Gabillon, Z., & Ailincai, R. (2013). CLIL: A Science lesson with breakthrough level. *Education,* 3(3), 168-177. https://doi.org/10.5923/j.edu.20130303.05

Gresnigt, R., Taconis, R., van Keulen, H., Gravemeijer, K., & Baartman, L. (2014). Promoting science and technology in primary education: a review of integrated curricula. *Studies in Science Education,* 50(1), 47-84. https://doi.org/10.1080/03057267.2013.877694

Haas, A., Grapin, S., & Lee, O. (2018). How the NGSS Science Instructional Shifts and Language Instructional Shifts Support Each Other for English Learners: Talk in the Science Classroom. In Bailey, A.L., & Maher, C.A. (Eds.), *Wilkinson Language, Literacy, and Learning in the STEM Disciplines* (35-52). Routledge. https://doi.org/10.4324/9781315269610-3

Han, T. (2015). Foreign Language Learning: Strategies in the Context of STEM Education. *GIST Education and Learning Research Journal,* 11, 79-102. https://doi.org/10.26817/16925777.287

Hasson, F., Keeney, S., & McKrenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advance Nursing,* 32(4), 1008-1015. https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x

Hung, H.L., Altschuld, J.W., & Lee, Y.F. (2008). Methodological and conceptual issues confronting a cross-country Delphi study of educational program evaluation. *Evaluation and Program Planning,* 31(2), 191-198. https://doi.org/10.1016/j.evalprogplan.2008.02.005

Landeta, J. (1999). *El método Delphi. Una técnica de previsión para la incertidumbre.* Barcelona: Editorial Ariel.

Lantolf, J.P. (2000) Second language learning as a mediated process. *Language Teaching,* 33(2):79-96. https://doi.org/10.1017/S0261444800015329

Linstone, H.A., & Turoff, M. (2002). *The Delphi method: Techniques and applications.* Addison-Wesley.

Linstone, H.A., & Turoff, M. (Eds.) (1975). *The Delphi method.* Reading, MA: Addison-Wesley. https://doi.org/10.2307/3150755

López-Gómez, E. (2018). El método Delphi en la investigación actual en educación: una revisión teórica y metodológica. *Educación XXI,* 21(1), 17-40. https://doi.org/10.5944/educxx1.20169

Manea, A.D. (2021). Teacher Self Evaluation - Indicator of Quality Management in Education. *Astra Salvensis,* 9(18), 37-46.

Manera, K., Hanson, C.S., Gutman, T., & Tong, A. (2019). Consensus Methods: Nominal Group Technique. In Liamputtong, P. (Ed.), *Handbook of Research Methods in Health Social Sciences* (737-750). Springer, Singapore. https://doi.org/10.1007/978-981-10-5251-4_100

Martín-Cudero, D., Cid-Cid, A.I., & Guede-Cid, R. (2024). Analysis of mathematics education from a STEAM approach at secondary and pre-universitary educational levels: A systematic review. *Journal of Technology and Science Education,* 14(2), 507-528. https://doi.org/10.3926/jotse.2349

Martínez-Piñeiro, E. (2003). La técnica Delphi como estrategia de consulta a los implicados en la evaluación de programas. *Revista de Investigación Educativa,* 21(2), 449-463.

McGuinness, C. (1999). *From Thinking Skills to Thinking Classrooms: A review and evaluation of approaches for developing pupil's thinking. Research report, 115.* School of Psychology, Queen's University, Belfast.

McMillan, S.S., King, M., & Tully, M.P. (2016). How to use the nominal group and Delphi techniques. *International Journal of Clinical Pharmacy,* 38(3), 655-662. https://doi.org/10.1007/s11096-016-0257-x

Mehisto, P. (2010). Criteria for producing CLIL learning material. *Encuentro,* 21, 15-33.

Meyer, O. (2010). Towards quality-CLIL: Successful Planning and Teaching Strategies. *Pulso,* 33, 11-29. https://doi.org/10.58265/pulso.5002

Meyer, O. (2016). Putting a pluriliteracies approach into practice. *European Journal of Language Policy,* 8(2), 235-242. Available at: https://muse.jhu.edu/article/645982

Morrison, J. (2006). *Attributes to STEM education: The student, the school, the classroom.* Teaching Institute for Excellence in STEM (TIES).

Nga, H.T., Lan, T.T., & Nguyen, N. (2018). Teaching English for 5th grade student in primary school via science topics on approach of STEM education: a case study of learning outside school activities. *Hnue Journal of Science. Educational Sciences,* 63(9), 61-69.

OECD (2023), *Education at a Glance 2023: OECD Indicators.* Paris: OECD Publishing. https://doi.org/10.1787/e13bef63-en

Ponce, O.A., & Pagán-Maldonado, N. (2015). Mixed methods research in education: Capturing the complexity of the profession. *International Journal of Educational Excellence,* 1(1), 111-135. https://doi.org/10.18562/IJEE.2015.0005

Pozo-Llorente, M.T., Gutiérrez-Pérez, J., & Rodríguez-Sabiote, C. (2007). El uso del método Delphi en la definición de los criterios para una formación de calidad en Animación sociocultural y tiempo libre. *Revista de Investigación Educativa*, 25(2), 351-366. Available at: https://revistas.um.es/rie/article/view/96831

Powell, C. (2003). The Delphi technique: Myths and realities. *Journal of Advanced Nursing,* 41(4), 376-382. https://doi.org/10.1046/j.1365-2648.2003.02537.x

Rodríguez-Pavón, P.R., Morales-Salas, R.E., Infante-Moro, A., & Infante-Moro, J.C. (2024). The Nominal Group Technique as a tool to select items that measure the level of digital skills in postgraduate students. *IJERI: International Journal of Educational Research and Innovation,* 21, 1-17. https://doi.org/10.46661/ijeri.9393

Roth, T., & Bogner, F.X. (2024). The trade-off between STEM knowledge acquisition and language learning in short-term CLIL implementations. *International Journal of Science Education,* 46(4), 338-361. https://doi.org/10.1080/09500693.2023.2232502

Schoettler, S.D. (2015). STEM Education in the Foreign Language Classroom with Special Attention to the L2 German Classroom. *Dissertations and Theses,* Paper 2313. https://doi.org/10.15760/etd.2310

Schwab, K. (2016). *The Fourth Industrial Revolution.* Ginebra: World Economic Forum.

Skumolski, G.J., Hartman, F.T., & Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education: Research,* 6(1), 1-21. https://doi.org/10.28945/199

Steurer, J. (2011). The Delphi method: an efficient procedure to generate knowledge. *Skeletal Radiol,* 40, 959-961. https://doi.org/10.1007/s00256-011-1145-z

Tezer, M. (2019). The role of mathematical modeling in STEM integration and education. In Fomunyam, K.G. (Ed.), *Theorizing STEM education in the 21st century* (1-14). IntechOpen. https://doi.org/10.5772/intechopen.88615

Trevallion, D., & Trevallion, T. (2020). STEM: Design, implement and evaluate. *International Journal of Innovation, Creativity and Change,* 14(8), 1-19.

Varela-Ruiz, M., Díaz-Bravo, L., & García-Durán, R. (2012). Descripción y usos del método Delphi en investigaciones del área de la salud. *Investigación en Educación Médica,* 1(2), 90-95. https://doi.org/10.22201/fm.20075057e.2012.02.00007

Vinuesa, V. (2017). *Content and language integrated learning (CLIL): an effective methodological approach to teach foreign languages in mainstream education.* Madrid: Dykinson.

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge: Harvard University Press.

World Economic Forum (2016). *The future of jobs. Employment, Skills, and workforce strategy for the Fourth Industrial Revolution.* Geneva. Available at: http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

Published by OmniaScience (www.omniascience.com)

Journal of Technology and Science Education, 2025 (www.jotse.org)