

الجمهورية الجزائرية الديمقراطية الشعبية

**PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA**

وزارة التعليم العالي والبحث العلمي

**Ministry of Higher Education and Scientific Research**

جامعة أبي بكر بلقايد - تلمسان

Aboubakr Belkaïd University – Tlemcen –

Faculty of Letters and Languages

Department of English



**THESIS**

Submitted in *Candidacy for the Degree of DOCTORAT*

**In:** Assessment and Testing in English Language Education

**By:** Abdelbassed Anes BOUBRIS

**THE IMPACT OF ENGLISH FOR SPECIFIC PURPOSES ASSESSMENT ON THE  
IMPROVEMENT OF LANGUAGE SKILLS AMONG 1<sup>ST</sup> YEAR COMPUTER SCIENCE  
STUDENTS AT TLEMCEN UNIVERSITY**

**BOARD OF EXAMINERS**

Prof. Abderrahmane BASSOU	Univ. Tlemcen	Chairman
Prof. Faiza Haddam Bouabdallah	Univ. Tlemcen	Supervisor
Dr. Nabil Aziz HAMADI	Univ. Sidi Bel Abbas	External Examiner
Prof. Hayat MESSEKHER	Higher Normal School of Bouzareah	External Examiner
Prof. Nawel BENMOSTEFA	Univ. Tlemcen	Internal Examiner

**Academic Year: 2023-2024**

## Declaration Of Originality

I affirm that the content presented in this submission is entirely my own creation. It does not incorporate any previously published or authored material from another individual, nor does it include content that has been acknowledged for the attainment of any other academic degree or diploma from a university or any other institution. I further attest that this work is free from plagiarism and is the outcome of my personal research, unless explicitly stated otherwise.

Name of the candidate:

Mr. Abdelbassed Anes BOUBRIS

Date: 18/12/2023

Signature:

A handwritten signature in blue ink, consisting of stylized initials 'BA' followed by a long, sweeping horizontal stroke that curves upwards at the end.

## **Dedication**

To my dearest Father and wonderful Mother,

To my beloved Wife and my precious Daughter,

To the best Sister ever.

## **Acknowledgements**

I am overwhelmed with gratitude for the incredible support and encouragement that has been showered upon me throughout this remarkable journey. My heartfelt acknowledgments go out to those exceptional individuals who have played a pivotal role in making this endeavor a reality.

I want to express my deepest thanks to my cherished supervisor, colleague, and friend Prof. Faiza Haddam Bouabdallah. Her enthusiasm, invaluable assistance, and pertinent suggestions have been the bedrock of this success.

A profound debt of gratitude is owed to the esteemed members of the jury. Their willingness to delve into the intricacies of my thesis and provide invaluable comments has left an indelible mark on this research journey.

A special note of thanks is extended to all the teachers in the English department for their uplifting comments and unwavering support, with a special mention to Prof. Wassila Hamza Reguig Mouro, Prof. Daoudi Frid, and Prof. Noureddine Mouhadjer.

My heartfelt gratitude to Emeritus Professor Jame E. Purpura, Prof. Donna Brinton, and Dr. Erik Voss, along with all the dedicated staff at Teachers' College –Columbia University. Your insights and support have left an indelible mark on my academic pursuit, and I am truly thankful for the privilege of benefiting from your expertise.

Special thanks also go to the Faculty of Sciences staff and students who have contributed in various ways to the completion of this work. Their assistance has been nothing short of instrumental in bringing this research work to fruition.

## **Abstract**

Language assessment produces reliable and organized data about learners' abilities, achievements, needs, and difficulties. The Algerian Higher Education Institutions recognize the crucial role of assessment in shaping teaching methodologies, curriculum design, student's motivation, and overall learning experiences. However, a lack of assessment literacy among Higher Education teachers, coupled with an overreliance on traditional assessment methods, presents significant challenges. This research delves into the English for Specific Purposes Language Assessment and its impact on the development of domain-related language skills among Computer Science students, at Tlemcen University. The study adopts a two-phased sequential exploratory case study design. The qualitative instruments encompass the analysis of 12 selected assessment artifacts and the guided observation of 3 English for Specific Purposes practitioners based on convenience sampling. The quantitative instruments include a satisfaction survey involving 367 undergraduate Computer Science students and two performance-based tests administered to 76 1<sup>st</sup> year Computer Science Students. Findings reveal critical gaps in English for Specific Purposes assessment practices, including issues with alignment, feedback provision, and lack of diversity in evaluative practices and materials, negatively impacting the development of domain-related language skills. Furthermore, the research sheds light on the complexities surrounding English for Specific Purposes assessment, such as the shortage of qualified practitioners, misalignment with learners' professional needs, and dysfunctional assessment designs. The study emphasizes the necessity of addressing these challenges to foster a higher quality assessment environment. The recommendations provided offer a roadmap for advancing research on English for Specific Purposes assessment in Algerian Higher Education, by embracing diverse research methodologies and exploring the long-term impact of evaluative practices, to actively participate in framing educational policies seeking to enhance the socio-economic progress of Algeria.

*Keywords:* LA, Computer Science, ESP, HEIs, impact, ICL, Integrated Skills, LOA, Algeria

# Table Of Contents

Declaration Of Originality.....	i
Dedication .....	ii
Acknowledgements.....	iii
Abstract.....	iv
Table Of Contents .....	v
List of Tables.....	x
List of Figures.....	xi
List of Acronyms .....	xii
<b>GENERAL INTRODUCTION .....</b>	<b>1</b>
<b>1.1 Introduction .....</b>	<b>11</b>
<b>1.2 An Overview of Assessment.....</b>	<b>12</b>
<b>1.3 Classroom-based Assessment .....</b>	<b>13</b>
<b>1.3.1 Classroom Assessment Techniques.....</b>	<b>16</b>
<b>1.3.2 Validity in Classroom Assessment .....</b>	<b>20</b>
<b>1.3.3 Reliability in Classroom Assessment .....</b>	<b>22</b>
<b>1.3.4 Fairness in Classroom Assessment.....</b>	<b>25</b>
<b>1.3.5 Classroom-based Assessment Practicality .....</b>	<b>27</b>
<b>1.4 Language Assessment Impact .....</b>	<b>28</b>
<b>1.4.1 Washback and Impact Status .....</b>	<b>28</b>
<b>1.4.2 Research into LA Impact.....</b>	<b>30</b>
<b>1.4.3 Aspects of Washback/Impact .....</b>	<b>35</b>
<b>1.4.4 The Impact of CATs .....</b>	<b>42</b>
<b>1.4.5 Assessment Impact in the Algerian Higher Education Context .....</b>	<b>44</b>
<b>1.5 Designing Language Tests for Specific Purposes .....</b>	<b>47</b>
<b>1.5.1 Achievement Tests .....</b>	<b>49</b>
<b>1.5.2 Classifying Tests .....</b>	<b>50</b>
<b>1.5.3 Common Task Types .....</b>	<b>52</b>
<b>1.5.3.1 Selected-response Task Formats .....</b>	<b>56</b>
<b>1.5.3.2 Limited-production Tasks .....</b>	<b>57</b>
<b>1.5.3.3 Extended-production Tasks .....</b>	<b>59</b>
<b>1.5.4 Task Specifications .....</b>	<b>60</b>
<b>1.5.4.1 Specifications for Item-based Tasks .....</b>	<b>62</b>
<b>1.5.4.2 Specifications for Reading and Listening Tasks .....</b>	<b>64</b>
<b>1.5.4.3 Specifications for Speaking and Writing Tasks .....</b>	<b>67</b>

1.5.4.4	<i>Specifications for Grammar Tasks</i> .....	68
1.6	<b>Test Development</b> .....	68
1.6.1	<i>Test Design</i> .....	70
1.6.2	<i>Test Operationalization</i> .....	73
1.6.3	<i>Test Administration</i> .....	74
1.7	<b>Integrated Content and Language Assessment</b> .....	75
1.7.1	<i>Learning-oriented Assessment</i> .....	77
1.7.2	<i>Meaning-oriented Model of L2 Proficiency</i> .....	78
1.8	<b>Language Assessment Literacy</b> .....	80
1.8.1	<i>Language Assessment Literacy Standards</i> .....	82
1.8.2	<i>Impact of Training and Education programs on Teachers' Assessment Literacy</i> .....	84
1.9	<b>Conclusion</b> .....	85
2.1	<b>Introduction</b> .....	90
2.2	<b>Situation Analysis</b> .....	90
2.2.1	<i>ESP in Algeria</i> .....	90
2.2.2	<i>ESP assessment in the Algerian HEIs</i> .....	93
2.3	<b>Theoretical Framework of the Research</b> .....	97
2.3.1	<i>Research Design</i> .....	97
2.3.2	<i>Research Objectives</i> .....	101
2.4	<b>Research Instruments</b> .....	103
2.4.1	<i>An Overview</i> .....	103
2.4.2	<i>Document Analysis</i> .....	104
2.4.2.1	<i>Purpose</i> .....	106
2.4.2.2	<i>Procedure</i> .....	107
2.4.2.3	<i>Summary</i> .....	111
2.4.3	<i>Classroom Observation</i> .....	112
2.4.3.1	<i>Purpose</i> .....	114
2.4.3.2	<i>Procedure</i> .....	115
2.4.3.3	<i>Observational Guide</i> .....	117
2.4.3.4	<i>Summary</i> .....	120
2.4.4	<i>Learners' Survey</i> .....	121
2.4.4.1	<i>Purpose</i> .....	123
2.4.4.2	<i>Description</i> .....	124
2.4.4.3	<i>Administration</i> .....	125
2.4.4.4	<i>Summary</i> .....	126
2.4.5	<i>The Test</i> .....	127

2.4.5.1	<i>Purpose</i> .....	129
2.4.5.2	<b>ICL Test Design</b> .....	130
2.4.5.3	<i>Description</i> .....	137
2.4.5.4	<i>Administration</i> .....	138
2.4.5.5	<i>Summary</i> .....	140
2.4.6	<i>Conclusion</i> .....	141
2.5	<b>Sampling Procedures</b> .....	142
2.5.1	<i>Defining the Target Population</i> .....	144
2.5.2	<i>Defining the Source Population</i> .....	144
2.5.3	<i>The Sampling Frame</i> .....	145
2.5.4	<i>Sampling Method</i> .....	145
2.5.5	<i>Sample Size</i> .....	146
2.5.6	<i>Conclusion</i> .....	147
2.6	<b>Data Analysis</b> .....	147
2.6.1	<i>Qualitative Data Analysis</i> .....	147
2.6.2	<i>Quantitative Data Analysis</i> .....	149
2.7	<b>Ethical Considerations</b> .....	151
2.8	<b>Conclusion</b> .....	153
3.1	<b>Introduction</b> .....	156
3.2	<b>Qualitative Phase</b> .....	157
3.2.1	<b>Assessment Artifacts Analysis</b> .....	157
3.2.1.2	<b>Findings</b> .....	158
3.2.1.3	<b>Discussion of the Findings</b> .....	170
3.2.1.4	<b>Summary</b> .....	174
3.2.2	<i>Classroom Observation</i> .....	174
3.2.2.1	<i>Findings</i> .....	175
3.2.2.2	<i>Discussion of the Findings</i> .....	177
3.2.2.3	<i>Summary</i> .....	180
3.2.3	<i>Conclusion</i> .....	181
3.3	<b>Planning the Second Phase</b> .....	182
3.3.1	<i>Informing the Design of Learners' Survey</i> .....	183
3.3.2	<i>Informing the Design of the Tests</i> .....	184
3.3.3	<i>Conclusion</i> .....	187
3.4	<b>Quantitative Phase</b> .....	188
3.4.1	<i>Learners' Survey</i> .....	188
3.4.1.1	<i>Findings</i> .....	188



3.4.1.2	<i>Discussion of the Findings</i> .....	197
3.4.1.3	<i>Summary</i> .....	205
3.4.2	<i>The Test</i> .....	206
3.4.2.1	<b>Findings</b> .....	207
3.4.2.2	<b>Discussion of the Findings</b> .....	211
3.4.2.3	<i>Summary</i> .....	217
3.4.3	<i>Conclusion</i> .....	218
3.5	<b>Discussion of the Main Findings</b> .....	219
3.6	<b>Conclusion</b> .....	224
4.1	<b>Introduction:</b> .....	225
4.2	<b>Incorporating Formative Assessment</b> .....	226
4.2.1	<i>Formative Assessment Potential</i> .....	226
4.2.2	<i>Enriching Classroom Assessment</i> .....	228
4.2.3	<i>Boosting Classroom Assessment Effectiveness</i> .....	231
4.3	<b>Incorporating Learning-oriented Assessment</b> .....	233
4.3.1	<i>Developing LOA in the Algerian HE Context</i> .....	233
4.3.2	<i>A Focus on Performance Moderators</i> .....	236
4.3.3	<i>Feedback in LOA</i> .....	238
4.4	<b>Diversifying Assessment Tools</b> .....	240
4.4.1	<i>ICL Assessment in the Algerian ESP Context</i> .....	240
4.4.2	<i>The Need for Listening and Speaking</i> .....	242
4.4.3	<i>Improving ESP Classroom Logistics</i> .....	246
4.4.4	<i>Designing Useful LA Tests</i> .....	249
4.4.5	<i>Rethinking Assessment Delivery in the Age of AI</i> .....	253
4.5	<b>Enhancing Teacher Educational Development</b> .....	257
4.5.1	<i>Training Programs</i> .....	257
4.5.2	<i>The Algerian ESP Centre</i> .....	260
4.6	<b>Fostering Research on ESP Assessment</b> .....	261
4.7	<b>Conclusion</b> .....	264
	<b>GENERAL CONCLUSION</b> .....	266
	<b>BIBLIOGRAPHY</b> .....	274
	<b>APPENDICES</b> .....	291
	<b>Appendix A: Selected Assessment Artifacts</b> .....	292
	<b>Appendix B: Observation Guide</b> .....	308
	<b>Appendix C: Learners' Appreciation Survey</b> .....	310
	<b>Appendix D: Scoring Rubric for Discount Scheme Dialogue Completion Task</b> .....	317

**Appendix E: Generic Test** ..... 319  
**Appendix F: ICL Test** ..... 324

## List of Tables

Table 1. 1 <i>Research on Washback Methodologies</i> .....	33
Table 1. 2 <i>Washback, Impact, and Effect in Research on LA</i> .....	40
Table 1. 3 <i>Common Task Formats</i> .....	55
Table 1. 4 <i>Specifications for Item-based Task Formats</i> .....	63
Table 2. 1 <i>Selected Assessment Artifacts</i> .....	109
Table 2. 2 <i>Artifact Analysis Categories</i> .....	110
Table 2. 3 <i>Observation Guide</i> .....	117
Table 2. 4 <i>LOA Dimensions of the ICL Test</i> .....	132
Table 2. 5 <i>ICL Test Blueprint</i> .....	135
Table 2. 6 <i>Observation Sample Frame</i> .....	145
Table 3. 1 <i>Learners' Satisfaction with the English Final Exam Procedures</i> .....	190
Table 3. 2 <i>Exam Design and Learners' Reading Skill Development</i> .....	192
Table 3. 3 <i>Exam Design and Learners' Writing Skill Development</i> .....	193
Table 3. 4 <i>Exam Design and Learners' Speaking Skill Development</i> .....	194
Table 3. 5 <i>Exam Design and Learners' Listening Skill Development</i> .....	196
Table 3. 6 <i>Exam Design and Language Skills in Domain-specific Situations</i> .....	197
Table 3. 7 <i>ICL Test Students' Scores</i> .....	207
Table 3. 8 <i>Generic Test Students' Scores</i> .....	208
Table 3. 9 <i>Test Scores Central Tendency and Variability</i> .....	210

## List of Figures

Figure 2. 1 <i>Mixed-methods Design</i> .....	142
Figure 3. 1 <i>Learners' Satisfaction with Exam Procedures</i> .....	198
Figure 3. 2 <i>Learners' Receptive Skills Assessment</i> .....	200
Figure 3. 3 <i>Learners' Productive Skills Assessment</i> .....	202
Figure 3. 4 <i>Learners' Satisfaction and Domain-related Linguistic Needs</i> .....	203
Figure 3. 5 <i>Symmetrical Distribution of GTS, ITS, and Net ITS</i> .....	212
Figure 3. 6 <i>The Impact of MCQ on RC GT and RC IT Score Distribution</i> .....	213
Figure 3. 7 <i>The Impact of MCQ on GR GT and GR IT Score Distribution</i> .....	214
Figure 3. 8 <i>Algo RC Scores impact on GTG, ITG, and Net ITG</i> .....	215
Figure 3. 9 <i>Students' Progress across Tasks</i> .....	216

## List of Acronyms

**AaL:** Assessment as Learning

**AfL:** Assessment for Learning

**AI:** Artificial Intelligence

**AoL:** Assessment of Learning

**CAT:** Classroom Assessment Techniques

**CEIL:** Centre d'Enseignement Intensif des Langues

**CLT:** Communicative Language Teaching

**CRT:** Criterion-referenced Testing

**DCT:** Dialogue Completion Task

**EFL:** English as a Foreign Language

**EGP:** English for General Purposes

**EMI:** English Medium Instruction

**ESP:** English for Specific Purposes

**FA:** Formative Assessment

**FE:** Final Exam

**FL:** Foreign Language

**FLT:** Foreign Language Teaching

**GR:** Grammar

**GT:** Generic Test

**GTG:** Generic Test Grade

**GTS:** Generic Test Score

**HE:** Higher Education

**HEI:** Higher Education Institutions

**I:** Informatiques

**ICL:** Integrated Content and Language

**ICLHE:** Integrated Content and Language in Higher Education

**IT:** ICL Test

**ITG:** ICL Test Grade

**ITS:** ICL Test Score

**KSA:** Knowledge, Skills, Abilities

**K-W-L:** Know-Want-Learned

**L1:** First Language

**L2:** Second/Foreign Language

**LA:** Language Assessment

**LAL:** Language Assessment Literacy

**LMD:** License-Master-Doctorat

**ME:** Makeup Exam

**MCQ:** Multiple-Choice Questions

**MI:** Mathématiques-Informatiques

**MOM:** Meaning-oriented Model

**NRT:** Norm-referenced Testing

**RC:** Reading Comprehension

**SLA :** Second Language Acquisition

**SM:** Sciences de la Matière

**ST:** Sciences et Technologies

**STEM:** Science, Technology, Engineering, and Mathematics

**TLU:** Target Language Use

**WT:** Writing Task

# **GENERAL INTRODUCTION**

Assessment is inherent to any teaching situation as it may occur before, during, or after teaching to provide insights about the teaching and learning experiences. Moreover, it produces reliable and organized data about learners' abilities, achievements, needs, and difficulties. Assessment encompasses the different processes used for gathering, measuring, and analyzing this data. These can be exploited reflectively and used to form informed judgments about the teaching-learning activity. Assessment, hence, fulfills a major function in the teaching-learning process and holds a prominent place in education. A number of significant decisions, with both microscopic and macroscopic impacts, can be reached based on continuous and periodic evaluation of learners, courses, and assessments among others. The fact of the matter is that due to its intrinsic importance, assessment has witnessed a strong and active interest in the scope of Foreign Language Teaching (FLT), encouraging its theoretical and methodological developments.

The two terms, assessment and evaluation, are often used interchangeably, though they differ to a great extent that they could be seen as diametrically opposed. Assessment is a modern and wider concept which encompasses, but is not limited to, tests and evaluations. The developments within the field of assessment have first focused on the disambiguation of the concept, through the study of its nature, functions, and methods. Language assessment, regardless of its type, is implemented to collect feedback upon which decisions are reached and actions taken. The significant body of research on language assessment calls attention to its importance to education in general and language teaching in particular. Several research findings have shown that effective teaching is narrowly linked to the effective implementation of assessment strategies and tools. A conceptual understanding of assessment functions, methods, and effects is indispensable to the success of the teaching-learning experience. The understanding, awareness, and skills required in the process of designing, implementing, and



interpreting assessment is referred to as assessment literacy. Basic assessment literacy can foster the natural relationship between assessment and teaching – informing and improving each other.

Language Assessment (LA) is a crucial component of the teaching and learning process, as it allows educators to gauge students' progress and identify areas where they need additional support. However, research on LA in the Algerian context is scarce, secondary-education oriented, and inconsistently documented. This is particularly concerning given the country's multilingual and multicultural population. LA importance, inherent to the importance of language as a medium of thought and communication in all domains, is amplified by the the political will and efforts to make a shift from French to English Medium Instruction (EMI) within Higher Education (HE) Science, Technology, Engineering, and Mathematics (STEM) contexts.

One of the most noticeable gaps in research on language assessment in Algeria is in the area of assessment impact. Assessment impact refers to the ways in which assessment can influence teaching and learning. For example, well-designed assessments can motivate students to learn and help teachers to identify and address students' needs. However, poorly designed assessments can have a negative impact on students' motivation and learning outcomes. Research on assessment impact in Algeria is limited. However, some studies have suggested that the current assessment system in the Algerian Higher Education Institutions (HEIs) may be having a negative impact on students' motivation and learning. Research has found that Algerian students often view assessments as punitive rather than formative due to the fact that assessments are often high-stakes and have a significant impact on students' grades and placement.

There is also a lack of research on the design, delivery, and scoring of language assessments in the Algerian HE context. This includes research on the development of valid and reliable assessments, as well as research on the use of appropriate scoring procedures. The lack

of research on language assessment in Algeria is a major concern. It is essential to conduct more research in this area in order to develop a better understanding of the current situation and its impact on students' learning and overall development. This research can be used to inform the development of more effective and useful assessment practices in Algeria.

In addition to the research gaps identified above, it is also important to note that research on LA in Algerian HEIs has rarely been conducted in the context of English for Specific Purposes (ESP). ESP courses are gaining prominence in the Algerian HE landscape, responding to the global demand for English proficiency in technical, scientific, and medical fields. This growing trend necessitates thorough examination of how effective LA approaches can be tailored to support the development of learners' ESP competencies, encompassing both linguistic and topical Knowledge, Skills, and Abilities (KSAs). The absence of research specifically focused on ESP contexts poses a significant gap in understanding the unique assessment needs of ESP learners. Unlike English for General Purposes (AGP) assessments, ESP assessments must not only evaluate general language proficiency but also assess the ability to comprehend and utilize specialized vocabulary, grammar, and discourse structures relevant to learners' specific disciplines. Without dedicated research on ESP assessment, there is a risk of inadvertently applying AGP assessment methods to ESP contexts, potentially leading to inadequate evaluation of learners' true ESP proficiency.

Moreover, the integration of ESP courses into Algerian HEIs calls for a critical examination of the role of LA in supporting the learning process and informing curriculum development. LA can serve as a valuable tool for ESP instructors to gauge the effectiveness of their teaching methods and identify areas where instruction may need to be adapted to better meet the specific needs of ESP learners. Furthermore, LA data can inform the development of ESP curricula that are tailored to the specific language demands of specific disciplines. Addressing this research gap is essential for ensuring that LA practices in Algerian HEIs are

effectively supporting the development of ESP learners' competencies and contributing to the overall success of ESP programs.

However, despite the increasing importance of ESP in the Algerian HE context, there still exists a significant misalignment between the intended goals of ESP assessments and their actual implementation, particularly in the Faculty of Sciences, Tlemcen University. This misalignment raises critical questions about the effectiveness of ESP assessment in addressing learners' specific needs and preparing them for future careers. The impact of evaluative practices on learners, society, and economy should become a national concern. One of the primary challenges confronting ESP assessment in Algeria is the scarcity of qualified instructors. ESP courses necessitate proficient educators not only in English but also in the intricacies of the relevant field. However, the prevailing practice of appointing newly graduated EFL students as part-time teachers without adequate training or clear syllabi undermines the pedagogical effectiveness of ESP courses. This situation prompts an urgent examination of the qualifications, training, and readiness of ESP teachers in the Faculty of Sciences.

ESP assessments encounter obstacles as many courses are not tailored to the specific needs of learners due to the absence of needs analysis in course design. The disconnection between instruction and real-world professional requirements necessitates a comprehensive investigation into the curriculum development process. Effective ESP assessment requires seamless collaboration between language tutors and subject matter specialists to ensure alignment with international job industry requirements. The lack of cooperation poses challenges to the relevance, usefulness, and effectiveness of assessments. Moreover, the limited time allocated to ESP modules, often only one or two hours per week, poses challenges for effective teaching and assessment. The constraints in time allocation require a closer examination to understand their impact on assessment methodologies and coverage of necessary material. The

proficiency level of students undertaking ESP courses is a significant factor influencing the success of assessments.

The notable scarcity of research dedicated to ESP assessment within the Algerian context is a critical issue. The absence of a robust body of research limits the availability of well-informed guidance for ESP educators, hindering the establishment of clear assessment procedures. Addressing this research gap is vital for ensuring the continued improvement of ESP evaluative practices in Algeria and the overall success of ESP learners. The present research is essential to shedding light on the current evaluative practices, developing more relevant, useful, and effective assessments that support the development of the Algerian STEM students' ESP competencies and adhere to the Algerian shift towards EMI.

The challenges facing ESP assessment in the Computer Science Department, at the Faculty of Sciences, encompass issues related to teachers' qualifications, needs analysis, collaboration, course organization, cohorts' size, students' characteristics, and resources, in addition to the lack of official guidelines regarding ESP instruction and assessment. These challenges underscore the critical need for the identification of ESP teachers' evaluative practices, the exploration of their impact, and the planning of targeted interventions. This research aims at understanding the impact of ESP assessment on first-year Computer Science learners' language skills development, to contribute to the socio-economic development of Algeria by producing graduates who are competitive in the global job market.

The significance of the present study lies in providing a model of ESP achievement tests that is learning-oriented, integrating both content and language, and aligning with the demands of the computer science profession. Moreover, the study offers an internal view of current practices regarding ESP assessment, collecting data that could be used by other researchers. The study also aims to improve students' overall linguistic performance and achievements through

evaluation by adopting a novel approach to Integrated Content and Language (ICL) through Learning-oriented Assessment (LOA).

Drawing on the previously mentioned problematic, and in order to reach the stated objectives, the researcher strives to provide answers to the following questions:

1. How do ESP teachers at the Faculty of Sciences assess 1st year Computer Science students?
2. How do undergraduate Computer Science students experience ESP assessment?
3. What impact does ESP assessment have on 1<sup>st</sup> year Computer Science students at the Faculty of Sciences?

These questions represent the framework leading the present research, thus, in order to thoroughly examine them, we hypothesize the following:

1. ESP teachers mainly rely on poorly designed summative assessments that are not aligned with 1<sup>st</sup> year Computer Science learners' needs.
2. Computer Science students are not satisfied with the actual ESP evaluative practices.
3. First-year Computer Science ESP assessment has an intense negative impact on learners' linguistic and topical skills development.

To fulfill the outlined objectives, address the research questions, and test the hypotheses, a sequential exploratory research design is employed. The initial phase involves collecting and analyzing qualitative data derived from twelve selected ESP assessment artifacts (encompassing final and makeup 1<sup>st</sup> year Computer Science English exams) and the guided observation three ESP teachers' classroom evaluative practices. This qualitative phase informs the development of research instruments used in the subsequent phase, which focuses on gathering and analyzing

quantitative data. The second, or quantitative, phase employs a satisfaction survey addressing all Computer Science undergraduate students (1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> year), in addition to two performance-based tests designed to assess the impact of ESP achievement tests on 1<sup>st</sup> year learners' language skills development within the Computer Science context. A comprehensive exploration of the research problem requires an examination of the theoretical foundations underlying each aspect of the study. Consequently, the work is organized into four chapters.

The first chapter establishes the theoretical framework for language assessment, outlining the concept of assessment, its categorizations, and standards. It provides an extensive review of formative, summative, and learning-oriented assessments, covering aspects such as assessment design, operationalization, delivery, and scoring. The chapter delves into the utility of assessments, emphasizing concepts like validity, reliability, authenticity, impact, fairness, and practicality. Notably, it introduces ICL assessment and a practical LOA framework based on Turner and Purpura (2016). The chapter concludes by highlighting the crucial role of teachers' assessment literacy in ensuring effective evaluative procedures.

The second chapter focuses on the theoretical background specific to this research design. It reviews the literature on the Algerian ESP situation, with particular attention to ESP assessment and the associated challenges. The chapter also establishes the practical foundations driving the research, outlining the research design, objectives, as well as data collection and analysis procedures.

The third chapter is dedicated to data analysis and discussions. It begins by presenting, analyzing, and discussing the data from the first phase. The chapter then explains how these insights informed the design of the second phase before detailing and discussing its main findings. It concludes with a general discussion that integrates data from both phases.

The fourth and final chapter is centered around recommendations based on the findings, discussions, and the review of literature. It primarily focuses on advocating for useful and learning-oriented ICL assessment strategies within the Algerian ESP context. The chapter includes recommendations for stakeholders, encompassing training programs and the establishment of an ESP center. Additionally, it seeks to provide insights into how language assessment may evolve in the age of Artificial Intelligence (AI).

## **CHAPTER ONE: Higher Education Language Assessment**

### 1.1 Introduction

### 1.2 An Overview of Assessment

### 1.3 Classroom-based Assessment

#### *1.3.1 Classroom Assessment Techniques*

#### *1.3.2 Validity in Classroom Assessment*

#### *1.3.3 Reliability in Classroom Assessment*

#### *1.3.4 Fairness in Classroom Assessment*

#### *1.3.5 Classroom-based Assessment Practicality*

### 1.4 Language Assessment Impact

#### *1.4.1 Washback and Impact Status*

#### *1.4.2 Research into LA Impact*

#### *1.4.3 Aspects of Washback/Impact*

#### *1.4.4 The Impact of CATs*

#### *1.4.5 Context Assessment Impact in the Algerian Higher Education (HE)*

### 1.5 Designing Language Tests for Specific Purposes

#### *1.5.1 Achievement Tests*

#### *1.5.2 Classifying Tests*

#### *1.5.3 Common Task Types*

##### *1.5.3.1 Selected-response Task Formats*

##### *1.5.3.2 Limited-production Tasks*

##### *1.5.3.3 Extended-production Tasks*

#### *1.5.4 Task Specifications*

##### *1.5.4.1 Specifications for Item-based Tasks*

##### *1.5.4.2 Specifications for Reading and Listening Tasks*

##### *1.5.4.3 Specifications for Speaking and Writing Tasks*

##### *1.5.4.4 Specifications for Grammar Tasks*

### 1.6 Test Development

#### *1.6.1 Test Design*

#### *1.6.2 Test Operationalization*

#### *1.6.3 Test Administration*

### 1.7 Integrated Content and Language Assessment

#### *1.7.1 Learning-oriented Assessment*

#### *1.7.2 Meaning-oriented Model of L2 Proficiency*

### 1.8 Language Assessment Literacy

#### *1.8.1 Language Assessment Literacy Standards*

#### *1.8.2 Impact of Training and Education programs on Teachers' Assessment Literacy*

### 1.9 Conclusion



**1.1 Introduction**

Language Assessment (LA) stands as a pivotal component in the broader realm of language education, serving as the compass that guides the language learning and proficiency development. At its core, Language Assessment encompasses a complex process aimed at evaluating individuals' language skills and abilities, offering insights into their linguistic competencies. This chapter delves into the intricate landscape of Language Assessment, through its conceptual foundations, types, standards, and the design techniques employed to ensure its effectiveness.

The significance of Language Assessment cannot be overstated, as it pervades the entire teaching and learning continuum, leaving an indelible mark on students' academic and professional future. Language Assessment plays a transformative role in shaping educational experiences, influencing curriculum design, and providing educators with valuable feedback to tailor instructional approaches. Beyond the classroom, the impact of language assessment reverberates into students' professional endeavors, influencing their communicative competencies in real-world and domain-related contexts. As an integral part of language education, assessment acts as a catalyst for continuous improvement, fostering an environment where learners can refine their language skills and educators can tailor their teaching methodologies to meet evolving needs.

From formative assessments that guide ongoing instruction to summative assessments shaping final evaluations, each type serves a unique purpose in the evaluation landscape. The chapter delves into the standards that underpin language assessments, ensuring fairness, validity, and reliability. Furthermore, the discussion extends to the design techniques that make assessments effective tools for gauging language proficiency. In addition to these key aspects, the chapter will illuminate the concepts of Integrated Content and Language (ICL) assessment and assessment literacy. Through this comprehensive exploration, the chapter aims to provide a

thorough understanding of Language Assessment, shedding light on its dynamic and indispensable role in language education.

### **1.2 An Overview of Assessment**

Language assessment is a pivotal element in language learning environments, serving to evaluate learners' progress during and after instruction and influencing the methodologies employed for language acquisition (Brown, 2003). Before constructing language tests, it is imperative to establish a clear purpose for assessment. In other words, when developing or utilizing language tests, it is crucial to recognize that tests and assessments serve as tools. The purpose of using a test or assessment should be clear, rather than using it arbitrarily.

The purpose and the audience of tests should be first considered, before any planning or design. Usually, the purpose of administering tests is to come to informed decisions about students or other individuals, such as those seeking language proficiency certifications. These decisions should influence the design of our tests. Recognizing a test as a measurement instrument can significantly aid in figuring out the most efficient ways to employ it, as various instruments are appropriate for distinct purposes (Carr, 2011). In other words, a specific test can be effective, considering its purpose, while proving unsuitable for another. Poorly designed tests with limited usefulness feature irrelevant and unsuitable tasks due to an inadequate planning. Conversely, well-designed tests might be highly specialized.

Assessment is integral to any teaching situation, occurring before, during, or after instruction (Boubris & Haddam, 2020). This reciprocal relationship requires educators to align assessment strategies with learning outcomes, learners' needs, and the learning environment. The reflective use of assessment data provides valuable insights into teaching and learning experiences. Research underscores the close relationship between quality teaching and the implementation of valid and reliable assessment strategies. Continuous and periodic evaluation of learners, courses, and quality is emphasized as a key aspect of language assessment in FLT.

Furthermore, the research has distinguished between assessment and evaluation, highlighting assessment as a broader concept that encompasses tests and evaluations (Dixson & Worrell, 2016). Understanding the nature, functions, and methods of assessment is crucial for its effective implementation. This disambiguation is essential for practitioners, as assessment, regardless of its type, aims to collect feedback for decision-making and action (Desheng & Varghese, 2013).

Assessment literacy is a critical aspect of the language assessment process. The understanding, awareness, and skills required in designing, implementing, and interpreting assessments contribute to the success of the teaching-learning experience (Malone, 2013). Assessment literacy fosters a natural relationship between assessment and teaching, allowing for mutual enhancement.

Language assessment is integral to language education, influencing teaching methods, shaping learning experiences, and contributing to the development of the field. As language assessment continues to evolve, a nuanced understanding of its purpose, design, and integration with teaching methods becomes increasingly essential for educators and researchers alike.

### **1.3 Classroom-based Assessment**

Language testing research has been imbalanced, with a disproportionate focus on large-scale international tests, overlooking the more common practice of classroom-based language assessment. Despite the prevalence of teachers conducting assessments, grading tests, providing feedback, and preparing students for high-stakes exams, there are still aspects of assessment practices in everyday classroom contexts that remain unclear or inadequately understood.

Classroom assessment encompasses a range of methods and techniques employed by teachers and students to gather, evaluate, and report on students' academic progress. It serves as a valuable tool for teachers to gather data and information on students' knowledge and skill development. Classroom assessment serves various purposes, including documenting students'

knowledge and abilities, identifying strengths, weaknesses, and misconceptions, enhancing learning outcomes, motivating students, assigning grades, and providing feedback to parents. It extends beyond mere testing, although tests have been commonly used as one form of classroom assessment (McMillan, 2015).

It is important to differentiate classroom assessment from other types of assessments in education. Classroom assessments are primarily administered and controlled by teachers, tailored to specific classes. They come in various forms, ranging from daily assessments to those conducted over extended periods. On the other hand, standardized assessments follow specific procedures, formats, and scoring criteria, and are used for multiple students across different classes, schools, and states. These assessments, often referred to as large-scale tests, have been in use for many years and include standardized achievement tests.

Classroom assessment encompasses five essential components: purpose, learning objectives, measurement, interpretation, and use. The purpose of assessment varies and serves functions such as diagnosing, providing feedback, documenting achievement, motivating students, grading, reporting to parents, and preparing for accountability tests. Learning objectives, also known as standards, specify what students should learn at different grade levels and should be clear and comprehensive, encompassing both content and cognitive complexity (Marzano & Kendall, 2007).

Measurement involves quantifying student performance and differentiating levels of knowledge and skills. Assessment methods include selected-response tests (e.g., multiple-choice) and constructed-response assessments (e.g., essays), with scores or grades as the outcome. Interpreting assessment results requires professional judgment, utilizing benchmarks and contextual information to determine the meaning of students' performance (Anderson & Krathwohl, 2001; Marzano & Kendall, 2007). While objective assessments, such as multiple-choice tests, are scored objectively, their interpretation still relies on teachers' informed judgments.

The use of assessment results should align with the intended purpose. Teachers employ the results in various ways, such as grading, reporting to parents, planning subsequent instruction, and providing feedback to students. By considering the purpose, teachers ensure that assessment outcomes serve as valuable tools to inform their decisions and support student learning.

In the past two decades, assessment for learning, also known as formative assessment, has emerged as a significant advancement in classroom assessment (Andrade & Cizek, 2010). It differs from summative assessment, which occurs at the end of a learning period to document student performance. Formative assessment is an ongoing and informal process integrated with instruction, with the primary goal of enhancing student learning.

The process of assessment for learning involves several essential steps. It starts by establishing clear criteria for evaluating student learning and understanding, which serve as the foundation for assessing progress toward learning targets (McMillan, 2015). Evidence of student understanding is collected through various methods, including pretests, questioning, short tasks, and interviews. Providing effective feedback is a crucial component, which should be timely, specific, individualized, and connected to the established criteria (McMillan, 2014).

Assessment for learning promotes the integration of curriculum, instruction, and assessment to positively impact student achievement and engagement. It emphasizes the mastery of knowledge and skills, encourages self-monitoring and self-determination, and cultivates student dispositions necessary for success in both careers and life. Embracing this philosophy requires teachers' commitment to improving student learning and motivation (McMillan, 2015).

Quality indicators for successful classroom assessment include establishing clear objectives and learning targets that are understood by students, parents, and teachers. By following these principles, assessment for learning becomes a powerful tool in facilitating student growth and academic success.

***1.3.1 Classroom Assessment Techniques***

Classroom assessment techniques (CATs) encompass a wide array of activities that can be categorized as either summative or formative (Hanson & Florestano, 2020). Summative assessments, such as tests and student ratings of instruction, serve an evaluative purpose and are typically administered at the conclusion of a learning period to determine the level of retained knowledge. In contrast, formative assessments are reflective and student-centered, serving as an ongoing process aimed at enhancing learning. These assessments allow for adjustments and improvements to be made by both instructors and students prior to summative evaluation (Adams, 2004).

The concept of formative classroom assessment techniques has gained significant attention since the 1980s (Simpson-Beck, 2011). CATs consist of specific teaching strategies designed to provide formative assessments by engaging students in reflective evaluations of course material and collecting their reflections on learning. These student reflections offer valuable feedback to instructors on the extent and quality of students' learning, thus facilitating improvements in teaching and learning practices.

Classroom assessment shifts the primary focus of teachers and students from improving teaching to observing and enhancing learning. It recognizes that fostering changes in students' study habits and metacognitive skills (i.e., thinking about their own thinking and learning) can often be more effective in improving learning outcomes than modifying teaching methods (Hanson & Florestano, 2020; Simpson-Beck, 2011). Ultimately, for students to become independent lifelong learners, they must take full responsibility for their own learning. Classroom assessment plays a crucial role in guiding both teachers and students in making necessary adjustments to enhance learning (Adams, 2004).

Implementing CATs involves a series of steps, beginning with planning, where instructors consider the potential benefits of CATs, select an appropriate technique that aligns

with course goals, provides valuable feedback, suits their teaching style, and can be easily implemented in the classroom. Subsequently, the technique is implemented by clearly explaining its purpose to students, collecting their responses, and promptly analyzing the collected data. Finally, instructors respond to students by sharing the insights gained from the assessment and explaining how that information will impact future instruction (Simpson-Beck, 2011).

A variety of CATs are available, differing in complexity and the time required for preparation, administration, and analysis (Marzano & Kendall, 2007). The choice of CATs depends on instructional needs, but some commonly mentioned techniques in the literature include the Minute Paper, where students spend a minute writing about the most significant point covered in class; the one-sentence summary; the Memory Matrix, which involves filling in a diagram based on provided labels; and the Muddiest Point, where students reflect on the most confusing aspects of a lesson or topic (Simpson-Beck, 2011).

Several decades of qualitative literature and anecdotal comments have asserted the effectiveness of Classroom Assessment Techniques (CATs) in improving student learning. However, the existing empirical evidence supporting this claim is limited (Simpson-Beck, 2011). There seems to be some confusion in the literature regarding the causal relationship between CATs and learning, particularly whether CATs have a direct effect on learning or on teaching. It is conceivable that CATs primarily enhance teaching practices, such as creating a positive learning environment, rather than directly impacting learning outcomes. Furthermore, CATs may assist students in becoming better learners by improving their ability to identify gaps in their understanding or misconceptions. In any case, CATs serve as a tool to provide a break from traditional lectures, promote student self-assessment, and offer ungraded and anonymous formative feedback (Marzano & Kendall, 2007).

According to Simpson-Beck (2011), CATs have become synonymous with assessment for learning (AFL) and formative assessment (FA). The aim of formative assessment and CATs is to enhance student learning and motivation through a series of five crucial steps. According to

McMilan (2015), FA should establish clear criteria for evaluating student learning and comprehension. These criteria serve as the key indicators of success, encompassing the aspects of student responses or the dimensions of products and performance that determine the level of understanding. FA should collect evidence that demonstrates student learning and understanding. This evidence can be obtained through various means, such as pretests or pre-assessments conducted before instruction begins. Other approaches, like observing student performance during instructional activities or using effective questioning techniques, can also be valuable. Both verbal and nonverbal indicators can be assessed to gauge student progress accurately.

Providing feedback to students based on their performance is the third step of the process. Extensive research has been conducted on effective feedback that promotes student learning. The feedback should be timely, specific, and individualized, avoiding general evaluative statements that lack specificity and may discourage students. Instead, the feedback should be directly linked to the criteria and specific responses of each student, enabling them to recognize their progress toward the learning target. It is also essential to help students understand that their lack of achievement is primarily due to insufficient effort, fostering a mindset focused on future improvement.

CATs should engage students in self and peer assessments. Encouraging students to reflect on their own learning and involve peers in the assessment process can be beneficial. This practice promotes self-monitoring and self-determination, as students gain a deeper understanding of their strengths and weaknesses. Students can generate their own ideas on how to improve their understanding, which contributes to their ownership of the learning process.

Ultimately, FA should help at implementing instructional adjustments or correctives. In this step, teachers guide students on additional learning activities that can further enhance their understanding or mastery of the material. These activities should be distinct from previous ones and may involve students' own ideas for improvement. By encouraging self-directed learning and



providing targeted support, teachers empower students to bridge the gap between their current knowledge and the desired learning outcomes.

Assessment for learning is based on the premise that assessments should be constructed and utilized in ways that positively influence both student achievement and motivation. This educational approach integrates curriculum, instruction, and Classroom Assessment Techniques (CATs) to shape the content of learning and the extent of student engagement (Hanson & Florestano, 2020; McMillan, 2015; Simpson-Beck, 2011; Thomas & Hornsey, 2014). It places a strong emphasis on mastering knowledge and skills, moving beyond the narrow focus on test scores to foster students' capacity to assess their own understanding, identify gaps in their knowledge, and design appropriate learning strategies to reach their educational goals. For educators, assessment for learning embodies a mindset centered on enhancing both student learning outcomes and motivation, thereby nurturing the development of crucial qualities for success in both professional careers and personal endeavors.

Amidst various CATs that can be harnessed to enrich students' competencies, one prominent method is the Know-Want-Learned (K-W-L) chart technique. This collaborative approach encourages active interaction with informational materials (Amaliani, 2017). It leverages the preexisting knowledge of students to collectively establish a more robust foundation for learning and to share insights gained through instruction. The K-W-L acronym signifies three essential phases: "what I know," "what I want to know," and "what I have learned." The primary objective of employing the K-W-L chart technique is to activate and build upon prior knowledge, establish a purpose, and facilitate the synthesis of newfound knowledge. Consequently, this strategy not only evaluates students' learning experiences but also provides educators with an evaluative tool (Conderman & Hedin, 2012). By guiding students through their reading materials, the K-W-L technique furnishes a framework to navigate their learning process.

***1.3.2 Validity in Classroom Assessment***

Validity in educational assessment refers to the appropriateness of its use and the reasonable interpretations, inferences, and consequences derived from it. It also involves evaluating whether an assessment encourages surface-level studying or deep content mastery, aligning with desired outcomes (Kane & Wools, 2019; McMillan, 2015).

When establishing validity for classroom assessments, practical reasoning, comparisons, logical analysis, and careful examination of assessment characteristics are essential, as standardized tests have well-defined technical approaches. One systematic approach to establishing validity is creating a test blueprint or table of specifications, which provides evidence for the alignment between assessed content and intended coverage (Kane & Wools, 2019).

Content-related evidence is commonly used in establishing validity for classroom assessments and involves evaluating the match between assessment content and instructional material. Another form of evidence, instructional sensitivity, compares test content with classroom instruction. External reviews of assessments in relation to instructional material can be utilized to gather content-related evidence and assess instructional sensitivity (Bonner, 2013).

Furthermore, employing multiple measures of the same concept contributes to validity evidence when they consistently lead to similar conclusions. This can be achieved by using various assessment methods such as quizzes, homework, tests, and projects, ensuring alignment in assessing student performance. Utilizing multiple measures and diverse measurement methods is recommended whenever possible.

Bonner (2013) proposed five principles crucial for establishing validity in educational interpretations based on classroom assessments. These principles offer a comprehensive framework that enhances the validity of classroom assessments and apply to learning, achievement, or their combination without requiring the development of new measurement

theories. They emphasize alignment between assessment content and instructional material, minimizing bias in assessment design and administration, assessing for student thinking processes, evaluating the interpretations and consequences of assessment-based decisions, and involving multiple stakeholders in the validation process.

Alignment is a critical aspect of validity, focusing on the connection between assessment tasks and curriculum standards as well as instructional practices. It involves considering learning objectives, educational philosophy, teaching style, and contemporary theories of human learning and motivation. The trade-offs between content validity and instructional validity are worthy of research attention.

Minimizing bias in assessments is crucial, both in task design and during test administration and scoring. Methods for detecting biases include subject matter expert analysis, teacher collaboration, debriefing with students, and item analysis techniques. Collaboration among teachers in addressing bias requires an environment conducive to inquiry into assessment validity.

Assessing for elicitation involves understanding student thinking processes and task-related behaviors, which can provide insights into cognitive processes, identify biases, and improve assessment design, scoring, and feedback. Timely feedback is vital for students to understand the relevant processes influencing their performance.

Evaluating interpretations involves justifying the consequences of assessment-based decisions through strong logical arguments and considering the impact on learning, motivation, and achievement. Frequent data collection and monitoring in the classroom environment aid in this evaluation, along with direct investigation methods such as student perceptions and opportunities for reassessment.

Validating interpretations requires involving multiple stakeholders to question and critically appraise assessment processes, interpretations, and decisions. External perspectives help identify potential validity issues, and a collaborative validation process is crucial to avoid

confirmation bias and improve the overall validity of classroom assessments. In summary, these principles provide valuable guidance for researchers, teachers, students, and other stakeholders involved in classroom assessments, fostering reflection and enhancing the validity of assessment practices in educational contexts.

### ***1.3.3 Reliability in Classroom Assessment***

Researchers in the fields of education and psychology have long been interested in the reliability of scores obtained from classroom assessments. Although classroom teachers may not find this topic as captivating, the body of research in this area has had significant impacts on classroom assessment practices, educational measurement as a whole, and the expectations placed on teachers regarding assessment knowledge. Reliability extends beyond its practical application in the classroom and encompasses philosophical, theoretical, and methodological dimensions. Therefore, it is crucial for teachers, administrators, policymakers, and researchers, regardless of their awareness, to recognize the importance of reliability (McMillan, 2015).

Every classroom assessment contains a certain amount of error or imprecision, which introduces variability and interferes with obtaining an entirely accurate result. In the field of measurement, reliability is employed to define and estimate the degree of error present in a single assessment. Essentially, reliability reflects the stability, consistency, or dependability of scores obtained from an assessment (Brookhart, 2009). Higher reliability indicates less error in the measurement. Determining the extent of error in a classroom assessment, similar to validity, often relies on professional judgment (Parkes, 2013). Various factors contribute to this error, including the nature of the assessment and scoring procedures (e.g., ambiguous directions, poorly worded items, biased scorers, fatigue), as well as individual student factors (e.g., fatigue, motivation, anxiety, luck in answering questions).

Teachers are typically unaware of how these error sources impact results, underscoring the importance of a general understanding of both the testing conditions and the students.

Enhanced reliability can be achieved by posing multiple questions that target the same skill, such as including several two-digit multiplication items in a math test. If a student who clearly possesses the skill answers all items correctly, while another student who clearly lacks the skill answers them incorrectly, it suggests that the test exhibits minimal error. However, it is essential to acknowledge that other sources of error may still be present. Therefore, even with well-designed tests, final scores should be interpreted with a margin of error, similar to how it is done in political polling. This means that a score of 85% should be understood as more accurately representing a range, such as 82-87% or 83-86%, within which the students' actual or 'true' knowledge or capability lies (McMillan, 2015).

To enhance reliability, assessment practices should include clear guidelines for answering and scoring, multiple items or tasks targeting the same construct, independent raters or scorers whenever feasible, minimizing disruptions during test administration, utilizing exemplar answers and papers for scoring, employing specific scoring rubrics, and opting for more frequent administration of shorter assessments rather than infrequent administration of longer assessments (Brookhart, 2009; McMillan, 2015; Parkes, 2013).

Grading and reporting practices are deeply intertwined with reliability, as the results of assessments are frequently used to determine grades. Grades serve as an indicator of a student's achievement or performance (Brookhart, 2009). Letter grades, such as A to F, along with percentage ranges, are commonly used to convey predetermined meanings. Descriptor labels like "excellent," "good," "fair," "poor," or "failing" are often attached to grades. Traditionally, letter grades are not evenly distributed among students, with fewer students typically receiving As, Ds, and Fs, while most students receive Bs or Cs. This distribution is often left to the discretion of the teacher. Grading involves a level of subjectivity and professional judgment, which can vary among teachers even within the same course or group of students. Factors like test difficulty, extra credit opportunities, improvement, and effort can influence grading decisions. In higher-

level classes, grades may receive higher point values compared to lower-level classes (Randel & Clark, 2013).

A more recent trend in grading is standards-based grading, which focuses on absolute levels of performance. This approach is similar to criterion-referenced grading, as it assigns grades based on pre-established performance levels that are clearly described and understood. The standards may be state standards or descriptions of different performance levels in classroom assessments. Rubrics with descriptors indicating increasingly sophisticated understanding or skills are used to differentiate the levels. In contrast, norm-referenced grading compares students' performance to that of their peers, with only the highest-achieving students receiving the highest grades. Norm-referenced grading is commonly used for ranking students and has implications for honors, college admissions, and employer evaluations (McMillan, 2015).

Another less common type of grading is self-referenced grading, which compares students' final performance to their past performance or aptitude. The focus here is on growth in performance rather than just the level of performance. Students who show significant growth from a weak initial understanding to an adequate or proficient level may receive higher grades, while those who begin at a proficient level but show limited progress would receive lower grades. Evaluating growth through classroom assessments or other assessments requires additional testing time, reliable measures, student motivation during pretests, and accurate measurement of aptitude if used as a pre-measure (Randel & Clark, 2013).

Final semester or course grades are typically determined by combining the results of various classroom assessments. Web-based and other software grading programs are often employed to aggregate grades, points, and percentages into a final score, which is then converted to a grade. Teachers decide the weight given to each assessment and whether scores are averaged or added using methods such as the total points method. Exams generally carry more weight than

quizzes and homework. Students and parents can usually access grades and final calculations online.

In addition to reporting grades, teachers communicate students' strengths, weaknesses, and areas for improvement to students and parents. This can be accomplished through parent-teacher conferences, student-led conferences where students review their work, and written feedback. Student self-assessment, often facilitated through portfolios containing examples of student work and reflections on their meaning, is a recent development in classroom assessment. Peer assessment is also gaining prominence.

#### ***1.3.4 Fairness in Classroom Assessment***

In order to ensure fairness, assessments should provide equal opportunities for all students, regardless of their personal characteristics and circumstances, to demonstrate their achievement. Fair assessments are characterized by their lack of bias and discrimination. A biased assessment is one that is influenced by factors unrelated to the knowledge or skills being measured. This bias can manifest in various ways, such as through assessment content that unfairly penalizes certain students or contains offensive material based on culture, income, ethnicity, or gender (Tierney, 2013).

According to McMillan (2015), unfair penalization occurs when certain students, due to their status, experiences, or characteristics, face disadvantages during test-taking, while others may have advantages. For instance, if a math test question uses a sailing example that favors students with sailing experience and poses challenges for those without such experience, it would be considered a biased item. Cultural differences, including vocabulary, values, and conventions, can also unfairly influence some students' ability to demonstrate their knowledge and skills. Additionally, students' learning style preferences can impact their performance on certain types of assessments. It is important to avoid culturally-based content in assessments that may be demeaning or perpetuate stereotypes, such as depicting a particular race as having only

low-status jobs. Lastly, the fairness of assessments is determined by the inclusion of appropriate accommodations for students with special needs and English Language Learners, thus providing them with equal opportunities to showcase their abilities.

The concept of fairness in educational assessment is subject to various and contradictory interpretations within the existing literature (Tierney, 2013). Perspectives on fairness are influenced by individuals' positions and roles within specific contexts, leading to differing notions of what is considered fair. Test developers, for instance, focus on addressing differential item functioning, whereas tertiary students desire respectful communication regarding grading. There are notable distinctions between fairness in standardized testing and fairness in classroom assessment. In classroom settings, where assessment primarily aims to support learning, the opportunity to learn holds greater significance. Fairness becomes exceptionally intricate in formative assessment due to its reliance on dynamic human interactions within constantly evolving circumstances. While retrospective evidence of fairness obtained through investigation or argumentation is valuable from a theoretical standpoint, it falls short in addressing the immediate needs of classroom practice, particularly for the learners affected by assessments.

Moreover, the complexity of fairness is compounded by the multifaceted nature of classroom assessment, which varies across different contexts. Consequently, fairness cannot be universally prescribed for all students in all classrooms. Sustained dialogue among stakeholders is of utmost importance to foster the fairest possible classroom assessment practices. Although notions of fairness may seem contradictory in specific situations, the underlying essence often remains consistent. This shared understanding provides a solid foundation for conceptualizing and implementing fair assessment practices within the educational community (Bonner, 2013; Tierney, 2013; McMillan, 2015).



***1.3.5 Classroom-based Assessment Practicality***

As educators, finding the right balance between comprehensive assessments and practicality is an ongoing challenge. Teachers face time constraints, and it is crucial to utilize assessment strategies that maximize the limited time available while still yielding meaningful outcomes. According to McMillan (2015), the ultimate objective is to employ assessments that lead to valid conclusions and have positive consequences for students, such as enhanced motivation, self-reflection, and overall academic achievement.

When selecting assessment methods, teachers must consider the time investment required for various assessment tasks. For instance, constructed-response assessments, which allow students to provide open-ended answers, can be highly effective in promoting critical thinking skills and deeper understanding. However, it is essential to carefully evaluate the practicality of scoring and providing feedback for such assessments. Constructed-response items often demand considerable time and effort to evaluate accurately, especially if they require subjective judgment. Teachers need to weigh the benefits of rich, open-ended responses against the time it will take to assess them thoroughly.

On the other hand, objective tests, such as multiple-choice assessments, offer advantages in terms of efficiency in scoring. These tests can be quickly and objectively graded, saving valuable time. However, creating high-quality multiple-choice questions that effectively assess students' knowledge and understanding is not a simple task. It requires careful consideration to ensure that the answer choices are well-constructed and that one or two options are not obviously incorrect. Crafting such questions demands upfront time investment to develop a reliable and valid assessment tool.

Considering the time limitations faced by teachers, it is generally advisable to use assessments that are as concise as possible while still providing valid, reliable, and fair results. Shorter assessments minimize the time needed for administration and scoring, allowing teachers

to dedicate more instructional time to classroom activities. However, it is crucial to strike a balance between brevity and the ability to capture essential information about students' knowledge and skills. The assessments must align with the learning objectives and provide a comprehensive understanding of students' performance without compromising the validity of the results.

In conclusion, teachers should carefully consider the time required to construct, administer, and score assessments while ensuring that the chosen methods are practical and efficient. The selection of assessment strategies should prioritize validity, reliability, and fairness. By finding the right balance between comprehensive assessments and practicality, teachers can optimize their use of assessment time and promote positive educational outcomes for their students.

#### **1.4 Language Assessment Impact**

Impact is a modern concept in the field of language assessment which “appeared in the literature as an extension of washback” (Saville, 2010, p. 1). Even if the review of literature on language assessment effects and consequences in academic contexts shows that both terms can be used interchangeably, theory presents them as two distinguishable concepts. Impact represents a “superordinate concept covering the effects and consequences of tests and examinations throughout society,” while washback is supposed to be limited “to the influence of tests and examinations in teaching and learning contexts” (Ibid).

##### ***1.4.1 Washback and Impact Status***

According to (Yi-Ching, 2009), there has been many attempts to define washback, also referred to as backwash (Nodoushan, 2021). This is why various definitions of the concept have been offered, in addition to a number of similar concept terms. The prevalence of both washback and impact in language assessment literature draws attention to the real and potential effects of evaluative practices on teaching and learning.

Hughes (1989) was the first to attach such a critical importance to ‘backwash’ in his textbook *Testing for Language Teacher*, introducing it as the side-effects of tests on the teaching and learning processes. Based on the various definitions that have followed Hughes’ (1989), washback, or backwash, encompasses the side-effects of evaluations at both a microscopic (the classroom) and macroscopic (beyond the classroom) levels (Yi-Ching, 2009).

Washback can be seen as the dynamics between evaluative practices and learning that result in decisive influences on the attitudes and motivation of teachers and learners, which ultimately promote or hinder learning (Green, 2013; Cheng & Curtis, 2012; Yi-Ching, 2009; Cheng, 2005; Bailey, 1996; Alderson & Wall, 1993). According to Bachman and Palmer (2010), washback is a subset of impact, in the sense that impact is a holistic washback that has influences educational systems and society as well (McNamara, 2014; Hill & McNamara, 2012).

The term ‘washback’ is more often associated with testing, standardized evaluations, and summative assessments (Ahmmed & Rahman, 2019; Sultana, 2018; Toksöz & Kiliçkaya, 2018; Green, 2013; Cheng & Curtis, 2012; Bachman & Palmer, 2010). ‘Impact,’ on the other hand, is mainly used in studies that focus on formative and alternative types of assessment (Monib et al., 2020; Alam & Aktar, 2019; Cong-Lem, 2019; Torres, 2019; Owen, 2016; McNamara, 2014; Wiliam, 2011). Research works that consider both formative and summative assessment tend to employ the term ‘impact,’ even if limited to certain aspects at a microscopic level (Torres, 2019; Qu & Zhang, 2013; Stoyhoff, 2012; Cilliers et al., 2010). Yi-Ching’s (2009) review of research into washback showed that the term may refer to the language assessment side-effects, experienced both within the classroom and beyond. Kiliçkaya (2016) sees that: “any assessment made, be it formative or summative, or teacher made and nation-wide, has an effect on both learners and teachers” (p. 117), and may extend its influence to educational systems and policies (Yi-Ching, 2009).

Since the early 90s, the idea that evaluations exerted a discernable effect on individuals and educational systems has held a prominent place in research into language assessment and

testing. Alderson and Wall (1993), among other pioneers of research into washback, debated the issue of washback existence, as clearly indicated by the titles of their work. This is no longer a central issue, since the overwhelming amount of literature on washback has proved its existence. The two last decades of research into washback have been marked by a great interest in the measurement of language evaluative practices impact in and beyond the classroom, its categorization, as well as its anticipation. In a nutshell, “The quality of instruction in any ... classroom turns on the quality of the assessments used there” (Stiggins, 1999, as cited in Mertler & Campbell, 2005, p. 20). Language evaluative practices definitely have an impact; however, “it is hard to tell exactly what washback looks like as there may not always be a linear relationship between tests and the teaching/learning process” (Dawadi, 2021, p. 2).

#### ***1.4.2 Research into LA Impact***

Educational research has started to put emphasis on washback in the 90s, particularly with the publication of Hughes’ (1989, 2003) *Testing for Language Teacher* in addition to Alderson and Wall’s (1993) *Does Washback Exist?* These foundational works have paved the way for future research on washback nature and consequences. Green (2013, p. 2) stated that: “a shift in views of test validity to embrace the use of tests as instruments of social policy” has promoted research into washback. It began as analyses of “the ongoing effects of established testing programmes or looked into how changes in systems of assessment affect educational practice” (Green, 2013, p. 4).

Alderson and Wall’s (1993) approach to the investigation of washback laid the ground for the body of research that followed after. Based on their pioneering work, the majority of washback-related research distinguished between its effects on content, methods, and attitudes (Toksöz & Kiliçkaya, 2017; McNamara, 2014; Green, 2013; Yi-Ching, 2009). Studies on washback put emphasis on the existence and the mechanisms of any eventual side-effects of the

evaluation (Ahmmed & Rahman, 2019; Toksöz & Kiliçkaya, 2017; Green, 2013; Cheng & Curtis; 2012; Cilliers et al., 2010).

The deployment of quantitative and qualitative research tools such as surveys and interviews is what, according to Green (2013), has allowed for what we now know about washback, or impact, in both high-stakes exams and low-stakes classroom assessment. Along similar lines, Rahman (2017) sees that there is no compelling reason to defend one approach to research against the other. The use of qualitative methods in LA research has strengths

such as, eliciting deeper insights into designing, administering, and interpreting assessment and testing; and exploring test-takers' behavior, perceptions, feelings, and understanding. Some weaknesses are, for instance, smaller sample size and time consuming. Quantitative research methods, on the other hand, involve a larger sample, and do not require relatively a longer time for data collection. Some limitations are that quantitative research methods take snapshots of a phenomenon: not in-depth, and overlook test-takers' and testers' experiences as well as what they mean by something. Among these two research paradigms, the quantitative one is dominant in the context of language testing and assessment research (Rahman, 2017, p.102).

Qualitative research has the particularity of recording and examining data without resorting to statistics theory. Qualitative research is characterized by the adoption of interpretive and naturalistic approaches to the collection and analysis of non-standardized data (Flick, 2014; Denzin & Lincoln, 2001). It is fundamentally opposed to quantitative research which grounds its methods in quantifications, often to present numerical amounts or rates (Rasinger, 2013). Unlike qualitative research, it emphasizes the systematic measurement and interpretation of variables using statistics theory (Bryman, 2012).

According to Rahman (2017), there are a lot of advantages specific to each of these opposed approaches. However, in his review of quantitative and qualitative methods use in LA research, he also draws attention to their major shortcomings. One has the merits of “producing descriptions of participants' feelings, opinions, and experiences; and interpret the meaning of their actions” seeking “to achieve deeper insights into issues

related to designing, administering and interpreting language assessment” (p. 104). Moreover, qualitative research methods emphasize the context in which the language assessment is administered and call attention to the complex features of language assessment (Tscushima, 2015).

Quantitative research, on the other hand, allow for data extrapolation based on random sampling techniques. Statistics theory is used to collect and analyze data, sometimes using software to process huge amounts of data in no time. The positivist paradigm is a pillar of strength to quantitative research methods –based on the measurement of objective and observable variables. Rahman (2017) concluded that quantitative approaches to language assessment research fail at uncovering deeper meanings and implicit relationships. Therefore, even if they can provide meaningful estimations about the proficiency, language skills, and washback, they are incapable of entering into an explanation of how social reality of a phenomenon is formed, maintained, and interpreted. Moreover, he argued that approaches based on the quantification of variables at a given moment do not offer more than a preview of the phenomenon, omitting participants’ experiences, and measuring in an abstract objectivity what sometimes should not be.

Qualitative research methods have shortcoming as well. These methods are not sufficient to form generalizations, neither to extrapolate findings to different contexts, because of the small and wholly representative sample size they include. Examining matters related to language assessment and testing through an interpretivist or constructivist paradigm cannot be done by software and does not output easy-to-read amounts and estimations. Hence, policy-makers have a tendency to prefer quantitative methods over qualitative ones (Rahman, 2017; Sallee & Flood, 2012). This is why, in spite of its drawbacks, quantitative methods in research about language assessment, and in language assessment itself, are the most common –the psychometric tradition (Purpura, 2011). Quantitative data is, against the common assumption, not as easily generalized because of the different contextual variables that may shape the findings.

The consensus view seems to be that the combination of both numerical and descriptive analyses to the study of washback will deepen our understanding of the phenomenon. It can offer, through the use of phenomenographic research methods for instance (Wilson, 2014), new insights into the washback phenomenon. In spite of the psychometric tradition that has profoundly influenced language assessment research and practices, qualitative approaches have contributed to research and policy-making by bridging the gap between educational phenomena and numbers. Recently, researchers have been showing a growing interest in mixed methodologies to approach language assessment research Toksöz and Kiliçkaya (2017). However, research into language assessment is dominated by quantitative methods (Rahman, 2017; Jang et al., 2014).

In their paper, *review of journal articles on washback in language testing in Turkey (2010-2017)*, Toksöz and Kiliçkaya (2017) clearly exposed the paradigm shift in research methodologies. The papers, they have reviewed in order to clarify the findings of research regarding washback in Turkish high- and low-stakes examinations, were mostly based on qualitative methodologies, and mixed methods (Table 1.1). This shows that even if a macroscopic view would still show the dominance of qualitative research methods, the microscopic view shows the gradual shift that is happening in language assessment research paradigms.

**Table 1.1** *Research on Washback Methodologies (Toksöz & Kiliçkaya, 2017, p. 186-189)*

Author(s)	Year	Methodology
Cinkara & Tosun	2017	Mixed-method research design, response sheets and interviews, 49 participants from Gaziantep University, Turkey.
Akin	2016	Qualitative research design, document analysis, 2013 spring and 2015 fall YDS questions.
Hatipoğlu	2016	Mixed-method research design, survey questionnaire and interview, 50 preservice English language teachers at Middle East Technical University, Turkey.

Kiliçkaya	2016	Qualitative research design, content analysis, 30 teachers at secondary schools, Burdur, Turkey.
Külekçi	2016	Qualitative research design, descriptive analysis, <i>YDS</i>
Sayın & Aslan	2016	Quantitative research design, survey questionnaire, 74 Turkish freshman ELT students studying at two universities in Samsun, Turkey.
Yeşilyurt	2016	Qualitative research design, metaphor elicitation, 110 academics at Atatürk University, Turkey.
Akpınar & Çkildere	2013	Survey questionnaire, descriptive statistical analysis, 103 academics at Nevşehir University, Turkey.
Paker	2012	Qualitative research design, document analysis, achievement tests given at 13 universities in Turkey.
Yavuzer & Göver	2012	Mixed-method research design, survey questionnaire, 121 academics at Nevşehir University, Turkey
Özmen	2011a	Mixed-method research design, comparative analysis, 164 pre-service teachers at 9 universities in Ankara, Turkey.
Özmen	2011b	Qualitative research design, comparative analysis, 8 candidate academics, and 4 graduate students in Ankara, Turkey.
Sarıçoban	2011	Qualitative research design, document analysis, two exams conducted at a high school, Ankara, Turkey.
Şenel & Tütüniş	2011	Mixed-method, experimental research design, 42 students at Aydın University, Istanbul, Turkey.
Yıldırım	2012	Mixed-method research design, descriptive statistics, 6 EFL instructors at a state university and 70 EFL students.

The main findings of their review were that state and private exams do not attach much importance to productive and receptive language skills, speaking and listening in particular. The main emphasis is placed on recognition skills, through Multiple Choice Questions (MCQs), which appears to interpose to learners' professional future. Participants, in some of the reviewed research works, have shared that they have to cope with fear and anxiety related to language tests. State and private language assessment give a high priority to reading, vocabulary, and



grammar, which, in turn, may have influenced curriculums and promoted teaching to the test; “students appear to master the format of these tests instead of acquiring the necessary skills for language learning” (Toksöz & Kiliçkaya, 2017, p. 201).

The researchers are not alone in his view that mixed methods would be beneficial to answer such questions in the sense that it neutralizes the shortcomings of both methods. Hence, the combination of introspections, retrospections, and observations with statistical estimations is gradually becoming the new ground of research into language assessment in general, and washback in particular (Rahman, 2017; Toksöz & Kiliçkaya, 2017; Manias & McNamara, 2015; Donmoyer, 2012). This methodological shift is particularly observable in recent research on language alternative assessment and its impact on the teaching-learning process and environment (Perrera, 2020; Shohamy, 2020; Cong-Lem, 2019; Afshar et al., 2018; Göçtü, 2013; Charvade et al., 2012; Bachman & Palmer, 2010).

### ***1.4.3 Aspects of Washback/Impact***

Language assessment and testing influences, at both microscopic and macroscopic levels, have been referred to as *test impact* and *curriculum alignment* as well (Yi-Ching, 2009). The former underscores the effects of testing on teachers, learners, and practices within the classroom and beyond, reaching educational systems, policies and society (Yi-Ching, 2009; McNamara 2004). The latter centers its attention on curricula being edited based on tests results. While *test impact* is a broad, and, to a certain extent, evasive term, *curriculum-alignment* (Yi-Ching, 2009; Shohamy et al., 1996) narrows considerably narrows the influences language assessment and tests may exert, and cannot pretend to be a similar concept term.

There have been many attempts to delimit the extent to which language assessment and testing influence its actors and environment. Most of these attempts argued in favor of a certain perspective of influence and provided terms such as *consequential validity*, or *side-effects*

(Messick, 1989) and *systemic validity* (Fredericksen & Collins, 1989, as cited in Yi-Ching, 2009). However, the terms *impact* and *washback* are the most established. Even though *impact* is theoretically regarded as a holistic view of language assessment and testing influence that encompasses washback, both terms appear to be used interchangeably in research (Schissel, 2018).

Research on language assessment and testing impact, including washback, has mainly focused on delimiting the extent to which side-effects can be observed and their measurement. Sultana (2018) has reported that the body of research on impact in South Asian countries concentrated on teachers and teaching, learners and learning, and assessment design, in addition to educational systems. Toksöz and Kiliçkaya's (2017) review of studies on washback in the Turkish context also revealed the same focal points.

The main theoretical premise behind research into washback, or impact, is that language assessment and testing have an influence on teachers, learners, and test-takers, in addition to instructive and evaluative practices, curricula, and educational policies. Current research appears to validate the view that washback, or impact, can be described and analyzed in terms of negativity and positivity, mostly known as *washback direction* (Alderson & Wall, 1993) or *washback value* (Hughes, 2003). Besides negative and positive washback, researchers have also looked at its *extent* (Bachman & Palmer, 1996), or *intensity* (Cheng, 2005; Cheng et al., 2004). Research on washback has also tried to conceptualize its effects in different dimensions, other than value and extent (Schissel, 2018; Aftab et al., 2014; Green, 2013; Sukyadi & Mardiani, 2011; Yi-Ching, 2009).

The first body of research on language assessment and testing *consequences* (Schissel, 2018, p. 1) propounded the view that washback extent, or intensity, are proportional to the importance attached to the test, or assessment, itself, meaning the two variables have a constant ratio (Liu & Yu, 2021; Green, 2013; Cheng, 2005). One question that needs to be asked however is: if learners, or test-takers, attached very little importance to the evaluation, would that reduce,

and eventually neutralize, its side-effects? Research on washback intensity has hardly ever considered the problematic from this perspective. Therefore, the majority of research on washback extent and direction focused on high-stake tests –based on their importance.

It is reasonable to think, nonetheless, that if learners/test-takers were to attach very little importance to a given language exam, this would eventually have some consequences, particularly over time, at both micro- and macroscopic scales. Cheng, Watanabe, and Curtis (2004) distinguished between weak and strong washback, in terms of intensity. As Aftab et al., (2014) reported: “Washback is said to be strongest or most intense where participants: (1) Value success on the test above developing skills for the target language use domain (2) Consider success on the test challenging, but both attainable and amenable to preparation (3) Work in a context where these perceptions are shared (or dictated) by other participants” (p. 152). In other words, strong washback “influences everything that happens in the classroom and where all teachers are teaching in the same way. An example of this scenario might be observed in the form of changes in teaching methodology and classroom activities as a consequence of the modifications in the test format from subjective to multiple choice questions” (Ibid., p. 150).

Nevertheless, it has been shown that participants’ motivation to succeed in the evaluation as well as the importance they place on their success, and on the evaluation itself, is a major indicator of washback (Green, 2013). In this sense, negative washback can at least be anticipated and limited, or even prevented, in similar scenarios (Papakammenou, 2018). Alderson and Wall (1993) distinguished between the beneficial consequences of washback and the damaging ones. In other words, impact can take a positive or negative direction to the extent that it supports or undermines the teaching-learning experience.

In his review of the pedagogical implications of washback, Yi-Ching (2009) argued that at a microscopic level, language assessment and tests can encourage teachers to cover their subjects comprehensively, without necessarily focusing on what is going to be in the test. Therefore, evaluative practices would promote learners’ motivation and attitudes through

appropriate assessment strategies. This, eventually, would be beneficial to the whole teaching-learning process. Conversely, the impact could differ and curricula be narrowed to fit the evaluation, depriving learners from knowledge or skills of paramount importance, and focusing on what is tested. Moreover, research has often shown that this leads to a misrepresentation of learners' performance, mostly because of anxiety. This, overtime, would undoubtedly cause learners' gradual loss of motivation for learning, and encourage them to develop test-taking skills that aim at achieving high scores, without any sense of accomplishment.

Washback direction has become an important component of research into washback, particularly because it can be a determinant component of the whether the influence of language assessment and testing practices is going to be favorable and constructive, or adverse and disruptive. In this sense, negative washback could be anticipated, and positive washback planned (Papakammenou, 2018). Bailey (1996), in his pioneering review of the washback concept in language testing, pointed out the determinant factors of washback occurrence: (1) the purpose of language learning, (2) assessment/test authenticity, (3) students' autonomy, (4) feedback. Moreover, socio-economic factors, such as learners' social background, in addition to English language teaching practices, and the perceived importance of the test can cause washback and greatly affect its aspects, including direction (Dawadi, 2021).

Research on washback has mostly focused on its intensity and direction as they were shown to play a pivotal role in the success of any language assessment or test (Barnes, 2017). Nevertheless, Schissel (2018) called attention to *intended* and *unintended* washback, referring to Messick's (as cited in Cheng et al., 2004) dimension of intentionality. In other words, "judging the validity in terms of whether a test does the job it is employed to do requires evaluation of the intended and unintended social consequences of test interpretation and use" (Sukyadi & Mardiani 2011, p. 97-98). Based on Cheng, Watanabe, and Curtis (2004), the dimension of intentionality involves the anticipated and unanticipated effects of language assessment and tests, at both micro- and macroscopic levels.

In addition to intensity, value, and intentionality, Cheng et al., (2004) put forward two other dimensions: *specificity* (including *general* and *specific* washback) and *length*. General washback refers to the major effects that may be caused by any language assessment or test. On the other hand, specific washback points at the effects related to a specific language assessment, or test. Length refers to the duration of washback, which effects can last for a short, or long, period of time. Moreover, Aftab, Qureshi, and William (2014) have pointed out washback *covertness*. They presented *overt* washback as commonly having a negative direction because it emphasizes the skills used in/for the examination, through teaching and text-books for instance. *Covert* washback is not that explicit in the sense that it refers to the assumptions which teachers, learners, and stakeholders can infer based on the test and its results.

Recent reviews of research into the influences exerted by language assessment and testing have confirmed the prevalence of direction and intensity in literature (Schissel, 2018; Sultana, 2018; Toksöz & Kalıçkaya, 2017; Beikmahdavi, 2016; Yi-Ching; 2009). These two dimensions of impact are more often observed through quantitative methods, which might be another reason for the dominance of these statistical approaches. However, aspects like washback *covertness* require the implementation of qualitative methods. *Covert* impact, for instance, includes teachers' acquisition, or modification, of knowledge based on the whole assessment experience. In other words, *covert* washback partially designates the relationship between teachers' beliefs (about teaching, learning, and assessment) and language assessment and testing. The implementation of qualitative approaches can significantly help if the correspondence between *covert* impact and its causes was analyzed.

Nevertheless, impact *covertness*, in spite of its significance, was scarcely addressed by research. Cheng's et al. (2004) intentionality, specificity, and length have been somewhat set aside as well. The existing body of research on washback, as shown through the present review of literature, has a third common point, besides being focused on mainly two aspects of washback and dominated by statistical theory. It has also emphasized on high-stake language

tests, which have the particularity of being summative, standardized, and as arguably traditional. The special attention given to this type of language evaluation has resulted in associating washback with the negative influences of language tests. Washback has been, henceforth, thought of as intrinsically negative.

On the other hand, research that has focused on low-stake alternative language assessments has attributed a number of positive influences to this type of evaluative practices. The overwhelming majority of these studies do not refer to alternative language assessment effects and influences as ‘washback’. These tend to employ to both ‘impact’ and ‘effect’ instead (Monib et al., 2020; Perrera, 2020; Shohamy, 2020; Alam & Aktar, 2019; Cong-Lem, 2019; Afshar et al., 2018; Göçtü, 2013; Charvade et al., 2012). Impact, in spite of its neutral value, is portrayed through the alternative assessment effects literature as intrinsically positive –in contrast with washback. Göçtü (2013) even declared: “In the thesis I did my best to view portfolio objectively, with all its advantages and disadvantages” (p. iii). Table 1.2 invites to a longer comparison between washback and impact occurrence in literature about language assessment and testing.

**Table 1. 2** *Washback, Impact, and Effect in Research on LA*

<b>Author(s)</b>	<b>Year</b>	<b>Focus</b>	<b>Methodology</b>	<b>Aspect</b>
Aftab et al.	2014	High-stake Summative assessment Washback	Qualitative	Value: negative Intensity: strong Specificity: general
Alam & Aktar	2019	Low-stake Portfolio assessment Impact	Mixed	Value: positive Intensity: weak
Allen	2016	High-stake Summative Washback/Consequential validity	Mixed	Value: positive/negative Intentionality: unintended
Charvade et al.	2012	Low-stake	Quantitative	Value: positive

		Portfolio assessment Impact		Intensity: weak
Elyza	2018	Low-stake Portfolio assessment Impact	Quantitative	Value: positive Intensity: weak
Göçtü	2013	Low-stake Portfolio assessment Impact	Mixed	Value: positive Intensity: strong
Hakim & Srisudarso	2020	Low-stake Portfolio assessment Washback	Qualitative	Value: positive Intensity: weak
Hatipoğlu	2016	High-stake Summative assessment Washback	Mixed	Value: negative Intensity: strong Specificity: specific
Iraji et al.,	2016	Low-stake Self-assessment Effect	Quantitative	Value: positive Intensity: weak
Kalra et al.	2017	Low-stake Portfolio assessment Impact / Effect	Quantitative	Value: positive Intensity: weak
Nassirdoost & Mall-Amiri	2015	Low-stake Portfolio assessment Impact	Quantitative	Value: positive Intensity: weak
Owen	2016	Low-stake Formative assessment Impact	Quantitative	Value: positive Intensity: weak Length: long-term
Perrera	2020	Low-stake Portfolio assessment Impact	Quantitative	Value: positive Intensity: weak
Pratolo & Zahrani	2020	Low-stake Dynamic assessment Effect	Qualitative	Value: positive Intensity: weak
Rahman et al.	2021	High-stake	Mixed	Value: negative

		Summative assessment Washback		Intensity: strong
Sukyadi & Mardiani	2011	High-stake Summative assessment Washback	Qualitative	Value: negative/positive Intensity: strong Length: short-term Specificity: specific

Based on the previous argument, it would be very logical to think that alternative assessment impact is initially expected to have a positive direction, which may at some point cause biases. The very same reasoning could be applied to research on washback, in the sense that it could bypass the positive effects of language evaluations. Overall, the terms ‘impact’ and ‘effect’ have gained in popularity in research on language assessment consequences. However, these two terms are broad and may be encountered in any educational research context where the influence of something on a different thing is explored. Washback and backwash, on the other hand, refer specifically to the effects of high- and low-stake language evaluative practices.

**1.4.4 The Impact of CATs**

Implications arising from the research on CATs have several key points. CATs are recognized as valuable tools due to their ability to engage students, measure their learning outcomes, and facilitate a feedback loop between students and instructors. The body of research on the impact of CATs provides evidence their contribution to an engaging and student-centered learning environment (Adams, 2004; Hanson & Florestano, 2020; Hogan & Daw 2014). By implementing CATs and demonstrating the utilization of feedback, instructors convey to students that their input is valued, fostering a positive relationship between students and instructors. This involvement in the learning process enhances interactions and participation. For instance, the use of CATs in online discussion forums leads to increased engagement in



discussions (Cross & Palese, 2015). The inclusion of students in the learning process also improves overall satisfaction.

CATs primarily focus on formative assessment, allowing students to demonstrate knowledge and skills without the pressure of high-stakes exams. CATs enable instructors to understand what students have learned and assist students in monitoring their own learning. CATs facilitate the creation of connections between students' prior knowledge and new information, supporting students in taking responsibility for their learning. Through reflection and metacognitive practices, CATs enhance students' abilities to monitor their understanding and identify areas for improvement (Adams 2004; McMillan, 2015). CATs also serve as evidence of the learning process, reinforcing metacognition, rehearsal of key ideas, and knowledge organization (Hanson & Florestano, 2020).

Feedback is a crucial aspect of effective teaching that aids both instructors and students in refining the learning process (O'Donovan et al., 2021). When instructors discuss CAT results with students, it helps students understand what they need to learn and how to approach the material. For instance, using CATs to identify students' levels of experience in a subject allows instructors to make adjustments to meet the students' needs and provide targeted feedback (Hanson & Florestano, 2020). Instructors can utilize this valuable feedback to make changes to the course or curriculum.

Moreover, CATs have the potential to improve student performance on assignments and tests work has shown that CATs increase student participation in online discussions, leading to higher quiz scores (Cross & Palese, 2015). CATs also help students process the material, understand instructor expectations, and succeed in future assignments. CATs also produce valuable information to instructors and students regarding areas that require additional attention or clarification. They provide insights into students' understanding, indicating readiness to progress to the next topic. This information enhances the efficiency and effectiveness of the learning process. Instructors can allocate class time more effectively; they can explain complex

topics better, and design assignments and exercises that assist students in exploring challenging concepts (Hogan & Daw, 2014).

Overall, CATs have a positive impact on student engagement, learning outcomes, instructor feedback, and course design. Their implementation enhances the learning experience, empowers students to take ownership of their learning, and enables instructors to provide targeted support and adapt their teaching strategies accordingly.

#### ***1.4.5 Assessment Impact in the Algerian Higher Education Context***

Washback and impact have been a subject of considerable interest and research in the field of assessment. Washback refers to the impact that educational assessments, particularly exams, have on various aspects of classroom practices and learning outcomes. Numerous studies have explored the effects of washback on different components of education, ranging from curriculum design to teaching methodologies, students' attitudes, and overall learning experiences.

In the context of Algeria, like in many other countries, examinations play a pivotal role in shaping educational practices and outcomes. These exams can significantly influence the way teachers design their curricula, select materials, and deliver their lessons. Moreover, they can shape students' attitudes towards learning and have a profound impact on their motivation and development. By examining the findings of studies conducted in the Algerian context, we seek to gain a deeper understanding of how exams influence classroom practices and student learning experiences. Specifically, we will explore the washback effects on curriculum design, the use of teaching materials, instructional methods, student attitudes, and learning outcomes in Algerian educational settings.

Understanding the washback effect in the Algerian context is crucial for educators, policymakers, and stakeholders alike. By identifying the positive and negative implications of

exams on education, we can make informed decisions to enhance the overall quality of the learning environment. Throughout the following, we will review and summarize the key findings of studies conducted on washback in Algeria, providing a comprehensive analysis of the impact of exams on various aspects of classroom practices.

In the context of education, the washback effect of standardized exams has garnered significant attention, including within the Algerian educational landscape. Specifically, research has predominantly focused on the impact of the Baccalaureate exam on teaching practices and student performance. Two notable studies shed light on this subject, each exploring distinct aspects of the washback effect in Algeria.

The first study by Bezziou and Ahmed (2013) aimed to investigate the washback effect of the Baccalaureate Examination on teaching practices in secondary schools in Jijel. This research employed a mixed methods approach, combining qualitative and quantitative data collection methods, including classroom observations and questionnaires. The findings revealed that the exam exerted both positive and negative influences on teaching practices. Notably, teachers employed communicative-oriented tasks to enhance students' communicative competence, reflecting positive washback. Conversely, negative washback was evident in the use of test-like activities aimed at boosting students' scores in the Baccalaureate Exam.

The second study, conducted by Lalaoua et al. (2021), investigated the washback effects of standardized tests, with a specific focus on the Baccalaureate exam's impact on teachers and students. Employing qualitative questions to gather perceptions from instructors and students, the research explored the scope and nature of washback effects, including how participants reacted to changes in the examination. The findings demonstrated that standardized testing, particularly the Baccalaureate exam, significantly influenced schools, teaching and learning processes, and curricula. These effects encompassed both positive and negative aspects.

The research findings from both studies indicated several levels of impact. At the teacher level, educators reacted positively to changes in the Baccalaureate exam, leading to alterations in their teaching methods, such as adopting a more communicative approach. However, they also faced challenges related to time management, curriculum coverage, and student engagement. Similarly, at the student level, washback effects were observed based on motivational strategies, classroom size, and the use of interactive teaching practices. Students' reactions to exam-related conditions, including motivation and anxiety, played a significant role in determining the impact of the Baccalaureate exam. Furthermore, the curricula and schools' level effects highlighted the influence of test preparation activities on narrowing the curriculum and reducing students' higher-order thinking abilities. Additionally, schools' principals could impact student scores, with some focusing more on raising test scores than overall student learning.

Research Findings by Hoadjli (2013) illustrate that the achievement tests employed by English educators in EFL (English as a Foreign Language) classes within the Algerian educational context exhibit significant shortcomings when it comes to achieving comprehensive evaluations. A substantial portion of these achievement tests comprises pre-made assessments, externally administered examinations, or assessments devised by teachers in a hurried and intuitive manner. These evaluations fall short in effectively assessing the advanced skills transmitted through instructional methods. They lack coverage of substantial portions of the syllabus content and do not adhere to systematic developmental stages. More concerning, they lack distinct and precise objectives and overarching goals. In specific scenarios, these tests offer value in evaluating particular segments of the curriculum or confirming general competencies; however, their usefulness diminishes in effectively diagnosing individual student needs and strengths across various learning stages.

To counteract these limitations, the ongoing research underscores the intricate nature of the testing process. It highlights the necessity for a systematic approach and a well-grounded

protocol to collect data on both general and specific language proficiencies using task-based assessments. The importance of crafting tests that can positively influence both educators and learners is emphasized. Hoadjli (2013) places special emphasis on the significance of acknowledging testing as a valuable mechanism to gather constructive feedback for refining teaching methodologies and enhancing the overall learning experience.

In conclusion, the concept of impact, particularly in the Algerian educational context, holds significant importance when considering the effects of washback in assessment practices. The studies discussed shed light on the profound influence that standardized exams, such as the Baccalaureate examination, exert on various aspects of education. It is evident that the impact of these exams extends beyond mere test scores, reaching into the realms of curriculum design, teaching methodologies, student attitudes, and overall learning experiences. As these studies highlight, the positive and negative ramifications of these exams can shape not only classroom practices but also the development of crucial skills and dispositions in students. Recognizing the multifaceted impact of exams in Algeria is essential for fostering an education system that optimizes learning outcomes, aligns teaching methods with student needs, and empowers educators and policymakers to make informed decisions that enhance the quality of education. Therefore, understanding and addressing the impact of exams in the Algerian context is not just an academic pursuit but a necessity for educational progress.

### **1.5 Designing Language Tests for Specific Purposes**

It is more important to think of assessment in terms of decisions to be made, rather than merely techniques or methods (types). In other words, tests, regardless of the method, can provide both type formative and summative information. Bachman and Palmer's (1996, 2010) Usefulness Model considers critical aspects of assessment engineering, which are: validity, reliability, authenticity, impact, practicality, and fairness. As any S/ FL assessment, ESP

assessment adheres to the same standards, as it is only distinguished by its commitment to specific content knowledge (Douglas, 2010; Paltridge & Starfield, 2013).

When creating or utilizing language tests, it is crucial to recognize that these assessments serve as tools, with a predefined specific purpose and audience. Typically, the purpose of administering a test is to make decisions regarding students or other individuals, such as individuals seeking language certification (Carr, 2011). Certain tests may be effective for one purpose but not as suitable for another. Some tests lack organization and come as random sets of questions and tasks put together without careful planning. Similarly, some tools are poorly constructed and serve little practical use. Conversely, well-made tools can be highly specialized, and the same applies to tests. For example, a test designed to assess the English-speaking proficiency of air traffic controllers may excel at its intended purpose but may not accurately measure a doctor's ability to communicate with nurses and patients.

There are often multiple options available when selecting a tool, including both expensive and more affordable alternatives. While the higher-priced options may offer enhanced performance, the additional expense might not justify the marginal improvement over a cheaper alternative that suffices for the task. Finally, it is essential to remember that when someone needs a specific tool, they ask for that tool specifically, not for a different one. Similarly, it is common for teachers to ask colleagues if they know of any suitable tests to use, disregarding the importance of aligning the test with their specific needs (Carr, 2011).

Language tests are primarily used to facilitate decision-making, and there are various types of decisions they can inform. Tests are typically categorized based on the specific decision they are designed to support. Curriculum-related tests, such as achievement and progress tests, are designed with a specific curriculum in mind, providing a reference point during test planning and development. On the other hand, when a test is not linked to a particular curriculum, such as

a proficiency test, it is necessary to determine what basis the test should rely on, which can be seen as either a challenge or an opportunity, depending on one's perspective.

### ***1.5.1 Achievement Tests***

Achievement tests are utilized to assess students' proficiency in meeting course objectives and mastering course content (Boubris & Bouabdallah, 2023). To classify a test as an achievement or progress test, its initial purpose needs to be considered. When a test aims to identify areas for (re)teaching, it is classified as a progress test. Conversely, if it is used to evaluate individual students' understanding of the material they were supposed to learn, it is classified as an achievement test.

For instance, imagine a test administered midway through a course. It serves the dual purpose of assigning grades based on students' comprehension of the material covered in the first half of the course and guiding the teacher's decision on whether any topics require review. In this case, the test functions as both a progress test and an achievement test. Another example involves a test administered at the conclusion of a course solely for assigning grades and evaluating students' overall learning. This test is solely an achievement test. The key consideration when assessing whether a test measures progress, achievement, or both, regardless of its designated label by the teacher or program, is to focus on the types of decisions it facilitates. This becomes especially crucial when the actual application of the test differs from its initial design (Carr, 2011). In the HE context, achievement tests can play a role in deciding whether students should advance to the next level, or make decisions related to graduation, in parallel with end-of-course assessments.

Moreover, there are two other test-based decisions linked closely to language curricula and programs that do not require distinct test types. The first pertains to program evaluation, where tests serve as evidence to gauge the effectiveness of the program. Notably, placement tests

and achievement test results are considered, particularly if the achievement tests align with the course or program goals and objectives, offering insights into program efficacy (Brown 1995). Tests can also contribute to the curriculum planning process. In cases where there is a need to identify learners' needs or evaluate prospective learners for a new program, tests may be employed to assess current knowledge and areas requiring enhancement. Diagnostic, placement, and achievement tests are frequently utilized for this purpose, even though proficiency testing is generally not directly associated with a specific language program.

In this sense, performance language tests were introduced to evaluate the actual performances of relevant tasks, rather than abstract demonstrations of knowledge (Lane, 2015). Carr (2011) distinguishes between two perspectives on second language performance assessments: the strong sense and the weak sense. The distinction is narrowly linked to the criteria used to appraise the performance. The strong sense focuses on the quality of task performance, reflecting upon real-world criteria, equating linguistic accuracy with task performance. In this perspective, adequate S/FL proficiency is essential but not sufficient for the performance task. On the other hand, the main goal of weak performance tests is to gather a proficiency sample for evaluation, considering task completion only in relation to language use. The weak sense of performance assessment is generally prioritized in LA, while the strong sense is more commonly found in ESP contexts (Carr, 2011).

### ***1.5.2 Classifying Tests***

In addition to categorizing tests based on the types of decisions they inform, there are several other perspectives to consider. These include frameworks for interpreting results, the tasks performed by examinees during the test, and the scoring methods used.

One significant perspective involves the distinction between norm-referenced testing (NRT) and criterion-referenced testing (CRT), which offer different ways to interpret test scores.



NRT involves comparing an examinee's results to the performance of others who took the test, often reported as percentile scores. On the other hand, CRT assesses an examinee's performance against predefined standards or criteria, measuring absolute knowledge or ability rather than relative performance. CRT scores are typically reported as percentages correct. Norm-referenced tests provide comparative information, whereas criterion-referenced tests focus on absolute performance (Hussain et al., 2015; Oo et al., 2021).

Norm-referenced tests rely on the scores of a norming sample (a group that took the test before its operational use) to link between these scores and percentiles. The reliability of the norming process relies on the norming sample size. However, norm-referenced tests do not provide information about absolute ability; they only indicate how well a test taker performed in relation to others. Criterion-referenced tests assess performance against predetermined standards and report scores based on the percentage of correct answers. CRTs aim to assess the degree to which learners have achieved specific standards or objectives (Hussain et al., 2015).

Furthermore, tests can be classified as summative or formative assessments. Summative assessments are generally used at the end of a unit, course, or program and provide information about overall learning. They are closely linked to achievement tests and assess learner's achievement (Moss, 2013). In contrast, formative assessments occur during the learning process and monitor progress to guide instruction. They are related to progress assessment and inform subsequent teaching and learning activities (Cerezo, 2011; Oo et al., 2021).

Testing can also be seen as either objective or subjective. Objective tests are scored objectively and employ selected-response questions such as multiple-choice or true-false. Subjective tests, like writing or speaking assessments, require human judgment for scoring. However, the terms "objective" and "subjective" can be misleading. Even in so-called objective tests, subjective decisions are involved in the test's planning, creation, and administration.

Subjective tests can be scored more consistently using scoring rubrics and rater training (Carr, 2011; Cerezo, 2011).

The direct and indirect testing distinction is also commonly discussed. Direct tests require test takers to demonstrate the specific ability being assessed, such as speaking or writing. Indirect tests assess related abilities through tasks that do not directly require the skill in question. While direct tests aim to measure specific abilities more authentically, indirect tests may provide easier scoring but may not yield as accurate information about the targeted skill (Cerezo, 2011; Oo et al., 2021).

Discrete-point and integrated tests represent another perspective. Discrete-point tests assess individual language components or skills separately, often employing multiple-choice questions. They allow for precise measurement but may lack authenticity. Integrative tests evaluate multiple language abilities simultaneously through tasks resembling real-life language use. Although more authentic, integrative tests may be more challenging to score and interpret (Hidri, 2018).

It is essential to note that these distinctions are not strict categories but rather points along a continuum. For instance, tests can combine features of norm-referenced and criterion-referenced approaches or include both integrated and discrete-point tasks. Therefore, it is crucial to consider the specific purpose, context, and desired outcomes when selecting an appropriate test design (Carr, 2011).

### ***1.5.3 Common Task Types***

It is essential to define the terminology associated with task types in language testing. "Task format" refers to how a task is presented, stating the input that needs to be processed and details about the expected response (e.g., selecting an option, writing a word, or performing a

role-play). Different researchers have used various terms to describe this concept, such as "task type," and "test method." However, "task format" is the most appropriate since it suggests that tests can include elements beyond traditional items or questions. Besides, it emphasizes the shape of a task rather than the ability being assessed, and it highlights the content of the response, not just the manner of response (Carr, 2011).

When classifying task formats, three key distinctions can provide a good understanding of a particular task's format. The first classification pertains to the type of response a task demands. Selected-response tasks require examinees to select the correct answer from a given set of options. In contrast, constructed-response tasks require the formulation a written or spoken response. Constructed-response tasks can be subdivided into limited production tasks, requiring brief answers such as single words or short sentences, and extended production tasks, which entail more extensive language samples (Carr, 2011; Purpura, 2004).

Secondly, items are distinguished from prompts in the sense that they refer to questions targeting selected or short answers, constituting selected response or limited production tasks. Prompts, on the other hand, are commonly used in speaking and writing assessments to ask examinees to provide extended responses. Finally, a task can also be passage-based or independent. Passage-based tasks require examinees to read or listen to material before engaging the task. It is crucial that these tasks rely on comprehending the passage, as they cannot be answered without understanding it. In contrast, independent tasks do not require processing additional material; the response is based on the item or the prompt itself, such in the case of independent grammar items. Ensuring passage dependence is essential in passage-based tasks.

According to Boubris and Bouabdallah (2023), test development can be construct-based, relying on theoretical definitions of SLA and language proficiency, such as the MOM of L2 proficiency (Purpura & Dakin, 2020), to drive assessment and develop performance scoring methods. It can also rely on a task-centered approach through the integration of both tasks and skills. In construct-based test development, the focus lies in identifying the construct and

creating tasks that effectively measure it. These tasks aim to elicit evidence regarding each examinee's level of proficiency in the construct under consideration. On the other hand, task-centered test development involves selecting tasks from the TLU (Target Language Use) domain, typically real-world language use tasks, that are of particular interest and adapting them for use in the test (Carr, 2011). This sampling and adaptation are necessary due to limitations in test length and the practicality of observing students in real-world contexts.

It is crucial to consider both the constructs being assessed and the tasks used to evaluate them. The ultimate rationale behind test administration is to draw reasonable inferences about the examinees' abilities, as with placement or achievement tests, based on test scores. These scores are directly influenced by students' performance, indicating their proficiency level. To enhance the authenticity of the test and its strong sense, tasks should closely resemble real-world language use situations, within resource constraints. Failing to achieve this authenticity can compromise the generalizability of the test results to real-world language use.

Neglecting either the construct or task aspect can lead to challenges in extrapolating test performance to other tasks not included in the assessment. Hence, Carr (2011) argues for a comprehensive approach that incorporates both task- and construct-based perspectives in test design. While starting the test planning from either perspective is acceptable, addressing both aspects is crucial. Regardless of their origins, test tasks often fall into recognizable formats. The next step involves exploring several common formats, their typical applications in assessing specific skills, and considerations when evaluating particular constructs. Evaluators must be mindful of the standards and considerations associated with the utilization of a particular measurement instrument.

It is essential to emphasize that the selection of a particular format should not be based solely on its ease of implementation. While novelty and convenience are relevant considerations, they should not overshadow the overall goal of the test's validity and effectiveness. Test developers should make informed and principled decisions. Ultimately, the format must align

with the test's purpose, and should not be haphazardly picked out from list of options. It is crucial to select the task format that best suits the intended objectives.

Language teachers are familiar with a variety of task types commonly used in teaching and assessment settings. It is now recognized that these holistic task types represent combinations of task characteristics designed to elicit specific language performance, and they can vary along multiple dimensions. It is essential to ensure that the test tasks align with real-life language use or instructional domains. Tasks could include, but are not limited to, selected-response, limited productions, and extended productions. Table 1.3 presents a list of common testing activities, along with description, classified according to the type of expected response.

**Table 1. 3** *Common Task Formats (Carr, 2011; Purpura, 2004)*

<b>Task Type</b>	<b>Description</b>	<b>Task Format</b>
Selected Response Tasks	<ul style="list-style-type: none"> <li>- Item-based formats like multiple-choice, true-false, matching, and ordering tasks.</li> <li>- Examinee selects the correct option from given choices.</li> <li>- Popular due to ease of scoring.</li> <li>- Some studies show differences in test-taker behavior compared to non-selected response tests.</li> </ul>	<ul style="list-style-type: none"> <li>- Multiple-choice activities</li> <li>- True/false activities</li> <li>- Matching activities</li> <li>- Discrimination activities</li> <li>- Lexical list activities</li> <li>- Grammaticality-judgment activities</li> <li>- Noticing activities • activities</li> </ul>
Limited Production Tasks	<ul style="list-style-type: none"> <li>- Item-based formats like multiple-choice, true-false, matching, and ordering tasks.</li> <li>- Examinee selects the correct option from given choices.</li> <li>- Popular due to ease of scoring.</li> <li>- Some studies show differences in test-taker behavior compared to non-selected response</li> </ul>	<ul style="list-style-type: none"> <li>- Gap-filling activities</li> <li>- Cloze activities</li> <li>- Short-answer activities</li> <li>- Dictation activities</li> <li>- Information-Transfer activities</li> <li>- Some informartion-gap</li> </ul>

	tests.	activities  - Dialogue (or discourse) comprehension activities
Extended Production Tasks	- Item-based tasks requiring written responses, including short-answer, fill-in-the-blank, sentence writing, etc.  - Authentic assessment, but grading challenges and multiple acceptable answers may be present.	- Summaries, essays - Dialogue, interviews - Role-plays, simulations - Stories, reports. - Some information-gap activities - Problem-solving activities - Decision-making activities

**1.5.3.1 Selected-response Task Formats**

Selected-response tasks encompass various formats where test-takers are presented with an item and required to choose the appropriate response. The form and length of the input can vary, and these tasks typically assess the recognition or recall of grammatical form and/or meaning. They are often scored based on a right/wrong criterion, although partial-credit scoring may be applicable in specific cases based on the construct being measured. Selected-response tasks can vary in terms of reactivity, scope, and directness (Schedl & Malloy, 2013).

One of the common selected-response tasks is the multiple-choice (MC) task. It involves input with gaps or underlined words, and test-takers must select the correct answer from given response options. The key represents the most suitable choice, while other options act as distractors. MC tasks are well-suited for testing discrete grammatical features and are relatively easy to administer and score. They can be pre-tested to determine their psychometric characteristics before operational testing. However, developing MC items can be time-consuming, and the format may encourage guessing and test-wiseness, potentially affecting score validity. Critics argue that MC tasks lack authenticity in real-world language use (Carr, 2011).

Another selected-response task is the matching task, which presents two lists of words, phrases, or sentences that test-takers must match. To avoid guessing, one list includes extra distractors. Matching tasks test multiple discrete grammatical features and encourage test-takers to establish associations between the two lists. They are straightforward to score. The discrimination task presents test-takers with language or non-language input and two response choices that contrast in some way. For example, it could be true–false or right–wrong. Discrimination tasks aim to measure differences between similar areas of grammatical knowledge (Purpura, 2004).

The noticing task involves presenting learners with language and/or non-language input and asking them to indicate their identification of specific language features. It helps learners build a conscious representation of the grammatical feature, making it effective for promoting grammar acquisition. The grammaticality-judgment tasks expose learners to sentences that may be well or ill-formed, and they must determine their acceptability. While these selected-response tasks serve different purposes in language testing, their appropriateness depends on the testing context and objectives, and researchers should make informed decisions when selecting the task format (Schedl & Malloy, 2013).

### ***1.5.3.2 Limited-production Tasks***

Limited-production tasks involve providing test-takers with an item containing language and/or non-language information of varying length and topic. Unlike selected-response tasks, limited-production tasks require test-takers to produce a response with a limited amount of language production, ranging from a word to a sentence. These tasks are intended to assess specific areas of grammatical knowledge, and the range of possible answers can be extensive, even for single-word responses (Neff & Rucinsky Jr, 2013). Scoring for limited-production tasks can take several approaches. Items with a single criterion for correctness can be marked right/wrong, while those with multiple criteria can be scored right/wrong for each criterion,

resulting in separate composite scores. Alternatively, the scores for each item can be aggregated, allowing for full, partial, or no credit. Holistic or analytic rating scales may also be used for scoring, especially when distinct aspects of grammatical ability need to be judged at different levels of mastery (Purpura, 2004).

One common type of limited-production task is the gap-filling task, where test-takers are presented with a sentence, passage, or dialogue with certain words deleted. They must fill the gaps with suitable responses based on the context. Another variation is the cued gap-filling task, where gaps are preceded by lexical cues to guide the appropriate response. Additionally, the cloze task involves mechanically deleting every fifth, sixth, or seventh word in a passage, and test-takers must fill the gaps with the best word for the context. Cloze tasks not only assess grammatical form and meaning but may also involve pragmatic knowledge depending on the passage's scope and relationship with the response. The short-answer task requires test-takers to respond to questions, incomplete sentences, or visual stimuli with answers ranging from a word to a sentence or two. Acceptable responses can vary widely, and these tasks are usually scored as right or wrong with criteria for correctness or partial credit (Carr, 2011). They can also be scored using rating scales.

Another type, the dialogue completion task (DCT), presents a short exchange or dialogue with a part of a turn deleted. Test-takers are expected to complete the exchange with a grammatically accurate and meaningful response. DCTs measure students' ability to use grammatical forms for various literal or grammatical meanings, and in some cases, they can assess pragmatic knowledge, including sociolinguistic or sociocultural appropriateness. DCTs have been widely used in applied linguistics research to investigate semantic formulas and linguistic devices for expressing various meanings. Despite some concerns about their reliability for sociolinguistic performance, DCTs have proven valuable for measuring grammatical forms and meanings (Purpura, 2004). They have been employed in both instruction and language testing, providing meaningful samples of grammatical performance.



**1.5.3.3 Extended-production Tasks**

Extended-production tasks differ from selected-response tasks in that they present a prompt instead of an item, and the input can be language or non-language information of varying lengths. The aim of extended-production tasks is to elicit substantial data from test-takers, with the quality and quantity varying for each individual. Some of these tasks are believed to measure implicit grammatical knowledge, while others, with planning time, can also measure explicit knowledge (Purpura, 2004). They are well-suited for assessing grammatical ability in speaking and writing contexts. When evaluating speaking ability, it is beneficial to record the interaction through audiotape or videotape for more reliable scoring and providing diagnostic feedback to students. These recordings can also aid in self and peer assessments and serve instructional purposes (Carr, 2011).

Scoring extended-production task responses relies on the rating-scale method, which requires defining scales based on the components of grammatical ability being measured and determining different levels of mastery for each scale along with corresponding observable evidence. Once the rating scales are devised, responses are rated according to the rubric and established scoring procedures (Neff & Rucinsky Jr, 2013).

The information-gap task (info-gap) involves presenting test-takers with multiple sets of partially complete information, and they must ask each other questions to obtain a complete set of information. This task aims to measure the ability to use grammatical forms to convey literal functional meanings and, depending on the setup, may also assess pragmatic knowledge. Scoring info-gap tasks utilizes the rating-scale method and may include evaluating task fulfillment, i.e., whether the students exchanged information reciprocally (Purpura, 2004).

Story-telling and reporting tasks prompt test-takers to use their own experiences or imagination to tell a story or report information. These tasks assess the ability to use grammatical forms to convey various literal and implied meanings, and the relationship between the input and

response may require special topical knowledge. Real-time tasks like these can measure implicit grammatical knowledge and can be recorded for more reliable scoring using rating scales derived from the test construct(s) (Carr, 2011).

Role-play and simulation tasks involve test-takers assuming roles to collaboratively solve problems, make decisions, or perform transactions based on input that can be language or non-language information of varying lengths. These tasks elicit large amounts of language and assess grammatical and pragmatic knowledge, topical knowledge, strategic competence, and affective schemata (Purpura, 2004). The scoring method depends on the test's purpose and the construct definition, with the relationship between input and response being reciprocal and indirect. These tasks are also evaluated using the rating-scale method, considering the constructs being measured.

#### ***1.5.4 Task Specifications***

The necessary components for developing the comprehensive specifications of a given test include the context and purpose of the test, in addition to its overall structure. These initial steps of specifications outline the decisions to be made regarding the test's objectives and its structure (type of sections, task formats, questions, prompts, and passages) (Carr, 2011). The significance of test planning prior to test writing is emphasized through research, both implicitly and explicitly. The planning process also involves writing specifications for each individual task format within each section. Once this stage is completed, along with the creation of a few sample tasks, the test writing process can begin (Davidson & Lynch, 2008).

When it comes to specifying individual tasks, certain general considerations must be addressed. Task specifications should go beyond identifying the task format used, already covered in the test structure specifications, and consider each type of question or prompt. For example, separate specifications are required for reading and listening items, as well as for different types of reading questions like "reading for the main idea" and inference questions (Carr, 2011). Similarly, MCQs and short-answer questions based on the same passage also

necessitate separate specifications. Although there may be some overlap between the specifications of different tasks, each task should have its own distinct description (Zandi et al., 2014).

The specifications for individual tasks should include various points. The purpose and constructs of each task should be derived from the context and purpose specifications of the test. Additionally, test tasks must be relevant to the assessed constructs; otherwise, they should not be included (Davidson & Lynch, 2008). Although there may be a temptation to integrate a task to assess multiple constructs simultaneously, such as using an MCQ item to evaluate both grammar and vocabulary, this often results in a poor assessment of both constructs (Carr, 2011). Identical considerations arise when the specifications outline particular segments of a construct. Consequently, when employing a selected-response or limited production task format, each question should focus exclusively on a specific portion of the construct. It is not necessary to have separate test sections for different parts of a multi-part construct.

Moreover, several aspects need to be covered when specifying scoring methods (Brookhart, 2009). In the case of question-based task formats, such as short-answer and MCQs, it must be determined whether the scoring will be dichotomous (right or wrong) or polytomous (including partial-credit scoring). Even for extended production tasks, the scoring rubrics do not have to be developed right away, but the aspects they address must be identified at this stage. These considerations, such as vocabulary use, grammatical accuracy, task performance, penmanship, spelling, punctuation, etc., are crucial (Zandi et al., 2014).

Providing samples and examples of input can make the specifications more tangible. This is particularly important when multiple individuals are involved in creating the test, as it helps ensure a clear understanding of the expected output (Zandi et al., 2014). Including both good and bad examples can be beneficial, as it provides guidance on what to strive for and what to avoid. In some cases, the sample input may be compiled in a separate document

Carr (2011) defends that specifications development is an iterative process and that previous decisions often need to be revisited. Changes may become evident while writing specifications or during the actual test writing phase. In such cases, all related sections, including questions, prompts, and passages, need to be reviewed and adjusted to align with the revised specifications. Further revisions may be undertaken after piloting and operationalizing the test as new problems may arise. For instance, specifications for listening passages might seem clear initially, but practical usage can reveal critical omissions. Incorporating such changes into the specifications makes them a permanent part of the "institutional memory," ensuring that new team members are aware of important decisions without relying on someone's memory.

When tasked with developing a new version similar to an existing test, section, or task reverse engineering specifications becomes imperative for teachers. This process involves deducing the implicit specifications from the test by examining its context, purpose, overall structure, and individual tasks (Walters, 2010). Although this may reveal inconsistencies or ignored aspects in the previous version, teachers should decide whether to adjust the specifications or seek guidance and validation from supervisors or other team members. This approach will result in an improved and better-planned test that aligns with the institution's preferences (Carr, 2011; Walters, 2010).

#### ***1.5.4.1 Specifications for Item-based Tasks***

The majority of language tests heavily rely on items, which include selected response questions (e.g., multiple-choice, true-false) and limited production questions (e.g., short-answer). When creating specifications for these items, it is essential to provide clear details about each item's structure (Carr, 2011; Spaan, 2006). The term "item stem" refers to the part of the item that presents the actual question. In multiple-choice questions, the stem precedes the choices, whereas in short-answer and true-false questions, the stem itself is the question.

A summary of essential components for selected-response and limited production tasks is shown in Table 1.4. For matching tasks, it is effective to include more options than questions to

prevent easy elimination of choices. In vocabulary tests with one-word options, including different parts of speech (nouns, verbs, adjectives) can discourage guessing solely through elimination, promoting knowledge-based responses (Carr, 2011). Short-answer questions should require at least one correct answer, but some may demand multiple answers for full credit (Spaan, 2006). For instance, a listening question might ask students to identify two different information; in such cases, determining the number of required information bits beforehand is essential, as well as whether items will be scored dichotomously (correct or incorrect) or polytomously (allowing for partial credit).

**Table 1. 4** *Specifications for Item-based Task Formats (Carr; 2011, p. 67)*

<b>For multiple-choice questions</b>	<b>For short-answer questions</b>
<ul style="list-style-type: none"> <li>• Whether items are presented in the LI or L2</li> <li>• How long the item stem and options are</li> <li>• Which vocabulary and grammatical structures are to be used (or avoided)</li> <li>• How many options there are</li> </ul>	<ul style="list-style-type: none"> <li>• Whether items are presented in the LI or L2</li> <li>• Whether responses are accepted in the L 1, L2, or both</li> <li>• Whether linguistic accuracy is included in scoring criteria (for answers in the L2)</li> <li>• Which vocabulary and grammatical structures are to be used (or avoided)</li> <li>• How long item stems are</li> <li>• How long responses should be</li> <li>• What is the maximum number of pieces of information to ask for</li> <li>• Whether polytomous or dichotomous scoring is to be used; if polytomous, how many points per item, or per piece of information</li> </ul>
<p><b>For true-false questions</b></p> <ul style="list-style-type: none"> <li>• Whether items are presented in the LI or L2</li> <li>• Whether “false” items need to be corrected</li> <li>• How long the item stems and options are</li> <li>• Which vocabulary and grammatical structures are to be used (or avoided)</li> </ul>	

<b>For matching questions</b>	<b>For deletion-based tasks</b>
<ul style="list-style-type: none"> <li>• Whether items are presented in the LI or L2</li> <li>• How many distractors there are</li> <li>• How long the item stems and options should be</li> <li>• Which vocabulary and grammatical structures are to be used (or avoided)</li> <li>• Which part(s) of speech are to be used (for one-word answers)</li> </ul>	<ul style="list-style-type: none"> <li>• What type of deletions is to be used:                             <ul style="list-style-type: none"> <li>- For fixed (random, nth-word) deletions: how many words between deletions</li> <li>- For rational deletions: what the criteria are for selecting deletions, and what the minimum number of words is between deletions</li> </ul> </li> <li>• How many deletions there should be</li> <li>• How long the intact text at the beginning of the passage should be</li> <li>• Whether multiple passages, or only one passage, can be used to reach the number of deletions</li> <li>• Passages: same as for reading comprehension passages</li> </ul>

**1.5.4.2 Specifications for Reading and Listening Tasks**

One crucial aspect of comprehension tasks is the language used to present the questions. While questions are commonly in the target language, this is not a strict requirement, especially when test takers share a given native language (L1) (Carr, 2011). For listening or reading comprehension assessments, this decision needs careful consideration. Presenting questions in the LI appears to lower the difficulty of the task, particularly for lower language proficiency examinees. Using L1 questions to evaluate L2 comprehension can also help avoid items with more difficult language than the passages themselves, focusing on the construct of interest: comprehending the content in the passage.

Additionally, in the case of short-answer questions, specifying the language of the output (or the response) is important. Responding in the target language, especially at lower proficiency levels, might hinder students from demonstrating their true comprehension (Carr, 2011). However, instances where an L2 response is specifically relevant, such as in short-answer grammar or vocabulary items, are more common. This doesn't imply that short-answer questions should always receive LI responses. Whether to use the LI in such tasks depends on the construct definition(s) guiding the test design. The language used of comprehension questions (and any expected student responses) is supposed to be less difficult than the language in the passages being assessed. Utilizing LI in such tasks is one approach to address the issue of lower language ability levels.

Regarding short-answer responses in the target language, test designers need to decide on whether, or not, linguistic accuracy should be considered when scoring, beyond cases where it affects clarity or factual accuracy. The decision will depend on the specified construct definition instead of intuition. For instance, if the construct is identified as "the ability to comprehend written text and answer questions using accurate language," including accuracy in scoring guidelines aligns with the definition. However, test designers may find it necessary to review both the construct definition and task format if it resembles what is expected in an integrative test. Furthermore, inferences about specific reading and listening "subskills" are only valid when the test genuinely requires students to apply those subskills to answer questions. Vocabulary-in-context items, in particular, pose challenges in this regard as they require students to infer word meanings from context rather than rely on prior knowledge.

In the case of listening comprehension, there is a consideration regarding whether examinees should be allowed to preview questions before the listening act. Allowing previewing aligns with real-life situations, where people often know why they are actually listening (Green, 2017). Decisions about the number of listening times should also consider authenticity of the task, the difficulty of the passage, and the students' proficiency level (Taylor, 2014). More

challenging passages may require multiple listens, while relatively easier ones might be presented only once.

In addition to specifying comprehension items, careful planning of the passages themselves is essential. For tests related to a specific course, this process may be more straightforward. Several key considerations must be taken into account when designing passages. The passages should align with the specified comprehension questions. For instance, if the test requires questions on specific details, inferences, main ideas, and scanning, the passages must be selected, adapted, or written to suit these purposes. For scanning, passages like schedules or catalogues should be chosen. Moreover, considering the length and familiarity of the topic is important, as more extended passages and familiar topics tend to be easier (Taylor, 2014).

Vocabulary plays a crucial role, and passage specifications should consider the level of vocabulary complexity, including lexical variation, sophistication, and density. Syntactic complexity should also be considered, either by counting T-units or by limiting the use of certain structures according to the curriculum (Carr, 2011). It is essential to incorporate information about the genre and rhetorical mode when considering specifications of reading comprehension passages. Genre refers to the type of passage, such as advertisements, letters, articles, or academic journals. Readability can be checked using readability formulas and the difficulty of the passage determined using software and websites.

Listening comprehension passages require considering the type of speech act, such as monologs or dialogs, transactional or interactional communication, and the genre of the speech act (Green, 2017). Authenticity is crucial, and reading aloud written passages should be avoided. The register, communicative functions, and aspects of pronunciation and delivery should also be specified (Davidson & Lynch, 2008). Furthermore, authentic discourse elements, like hesitations or interruptions, must be addressed. Additionally, passages can be scripted, unscripted, or semi-scripted. Semi-scripted passages offer a compromise between control and authenticity, with varying degrees of detail in the specifications.



***1.5.4.3 Specifications for Speaking and Writing Tasks***

It is crucial to create clear and effective prompts to elicit specific responses from test takers. Task specifications for prompts should outline the desired responses and provide directions that will be given to the students during the test (Davidson & Lynch, 2008). The teacher writing the prompt should ensure that the information give students a fair opportunity to perform well on the task. For writing prompts, the specifications should indicate the desired length of the response, the genre, and the rhetorical mode (Carr, 2011). In speaking prompts, it should cover the type of speech act, register, communicative functions, and other discourse aspects. The desired vocabulary, grammatical patterns, and topic specificity should also be specified. These specifications align with the elements needed for reading and listening passages.

In interview-based speaking tests, there are two main approaches: fixed sets of questions or unguided interviews. The former involves asking the same questions to all students, while the latter allows interviewers to improvise the prompts. Most real-world interviews strike a balance between the two extremes, using a mix of fixed questions and follow-up questions to ensure clarification and authenticity. Interviewers may sometimes be given a list of sample questions, grouped by topics, functions, or grammatical structures. This approach allows for more flexibility while ensuring that relevant topics are covered (Carr, 2011).

However, it is important to address any potential discrepancies in interlocutor interventions and the difficulty of questions to maintain fair scoring. Group interviews or discussion tasks can also vary in structure, from specific directed questions to more general topics for discussion. Providing students with questions or topics to discuss keeps the task focused and avoids confusion. Regardless of the approach chosen, the task specifications should clearly outline the format of the interview or discussion, the types of questions or topics to be used, and any guidelines for scoring or providing support to test takers. This information is vital for consistency and fairness in the speaking test.

**1.5.4 Specifications for Grammar Tasks**

Three major considerations, going beyond those applicable to the other types of tests, arise when designing vocabulary and grammar tests. Firstly, it is necessary to distinguish between the ability to recognize a grammatical structure, comprehend it, and the capacity to use it accurately. Secondly, it might be necessary to assess both the range of acquired vocabulary or grammatical structures and the level of proficiency achieved in using them. This consideration also applies when incorporating vocabulary or grammar assessment into scoring rubrics for speaking or writing tasks.

Lastly, in some cases, employing polytomous scoring may be appropriate to evaluate control over both form and meaning (Purpura, 2004). This approach can be implemented in multiple-choice questions and limited production tasks, where two points are assigned - one for form accuracy and another for meaning accuracy.

**1.6 Test Development**

Language assessment plays a pivotal role in the field of language education, serving both institutional and social purposes. Effective language assessment tasks form the foundation of assessment design, wherein learners' performance is judged and communicated, leading to correction of errors and skill development. A well-designed assessment or test must consider various factors, including the purposes of assessment, alignment with teaching and learning outcomes, provision for iterative practice and feedback, and support for accurate judgments of learners' performance. Moreover, assessments need to be sustainable from both the learners' capacity to make evaluative judgments about their future work and the teachers' workload perspective. The introduction of innovative tests and tasks can challenge conventional practices, necessitating collaboration among stakeholders, educators and students.

In language education, language teachers are expected to be adept at designing high-quality language assessments for the skills and constructs they evaluate. Test development is

particularly crucial as poor assessment design can have detrimental effects on language teaching and learning. However, current literature on the topic often presents test development process in isolation, overlooking the impact of the institutional milieu on assessment (Giraldo, 2019). Before delving into the stages of the test development process, it is essential to understand the fundamental qualities of language assessments (Bachman & Damböck, 2018; Bachman & Palmer, 2010; Giraldo, 2019). These qualities are relative rather than absolute, requiring careful consideration of the assessment's purpose to ensure their effective implementation.

In the process of test production, relevant documentation is utilized, encompassing the test framework, specifications, and item-writer guidelines. Test development is a comprehensive and time-consuming endeavor, taking several months or even years. It commences by commissioning test items from suitably qualified item writers, followed by multiple rounds of review and revision. The revised items undergo pre-testing (trailing), and subsequent analyses of item responses lead to further revisions if necessary (Purpura, 2004; Rossi, 2021).

It is essential to acknowledge that such extensive procedures are not always feasible in classroom testing scenarios. Nonetheless, classroom teachers should strive to enhance the quality of their tests. Often, language teachers lack adequate training in test development despite being expected to create tests regularly (Koh et al., 2018). Taylor (2013) emphasized the significance of language assessment literacy (LAL) for test writers, encompassing language testing theory, principles, concepts, and technical item-writing skills.

Language teachers in educational institutions can form working groups to collaboratively develop tests. The involvement of colleagues for idea exchange and peer review is crucial, as creating a high-quality test requires collective effort (Rossi, 2021). These groups should begin by producing test specifications, ensuring the intended construct is targeted and allowing for comparability between test versions developed in different years or by different teachers. Purpura (2004) presented the three stages of grammar test development: design,

operationalization, and administration, highlighting the importance of adapting the outlined test-development process based on the specific situation rather than treating it as a rigid formula.

### *1.6.1 Test Design*

In the realm of test development, the design stage stands as a pivotal phase that involves the careful amalgamation of crucial information and the initial formation of decisions that govern the entire testing process. While smaller-scale tests may undergo a more informal design process, when dealing with larger audiences such as joint final exams or placement tests, test developers find themselves in the midst of intricate discussions and negotiations with multiple stakeholders. At the heart of this stage lies the indispensable design statement, a comprehensive document that comprises essential elements outlined by Bachman and Palmer (1996, 2010).

This statement encompasses a description of the test's purpose, the TLU domains and task types it covers, the characteristics of the test-takers, the construct(s) to be measured, a well-thought-out plan for evaluating the test's usefulness, and a strategy to handle resources effectively.

The initial phase of test development involves establishing the purpose of the grammar assessment. This purpose defines the specific inferences to be made about grammatical knowledge or its practical application. For example, the test might aim to measure grammatical ability concerning comparative forms and meanings and inform decisions about student progress, placement, or language proficiency. The purpose statement also considers the impact of test results and how they will be used, such as promoting further learning or providing instructional feedback.

After defining the purpose, the Test Language Use (TLU) domain is identified, which can be real-life or language-instructional. TLU task types are then selected based on the needs analysis, which involves gathering information about the target-language needs of the test-takers. The choice of TLU tasks can vary based on the context, with some tests focused on

communicative language teaching, while others concentrate on discrete language features. The TLU domain influences the selection of tasks to ensure their relevance to the test-takers' abilities.

The design statement provides a detailed description of the test-takers' characteristics to clarify the population for whom the test is intended and the generalizability of test scores. Test-taker attributes, such as age, native language, gender, and language proficiency level, are considered as potential influences on test results. The design statement also defines the construct(s) to be measured in the test. Construct definition can be based on instructional objectives, standards, theoretical definitions, or a combination of these factors. For grammar tests, construct definitions may focus on specific grammar points taught during a period or be derived from a theoretical model of grammatical knowledge. This definition forms the foundation for test construction, score interpretation, and test validation.

The test designer determines the role of topical knowledge in the construct definition of grammar tests. Three options are provided: excluding topical knowledge, including it in the construct, or defining it separately from the language construct. The choice depends on whether topics are integral to the curriculum, contextualize language learning, or are the main focus, as in content-based language programs. Strategic competence, involving metacognitive, cognitive, social, or affective strategies in test performance, is generally assumed to be invoked in grammar tests. However, in certain cases, separate inferences to strategic competence may be necessary and should be specified in the construct definition if required.

The test design statement outlines a plan for assessing the test's usefulness, considering six qualities: reliability, validity, authenticity, practicality, impact, and fairness. Throughout the development process, developers continuously evaluate these qualities and gather empirical evidence to support test usefulness. The test design statement also addresses the human, material, and time resources needed for test development. Priorities are set based on test usefulness and considering any limitations in resources.

In the pursuit of crafting an effective and meaningful test, the design stage plays a paramount role, channeling the vision and purpose of the assessment. Whether the test is tailored for a specific class or targets a broader audience, the decisions made during this stage have a profound impact on the overall testing process (Bachman & Adrian, 2022). Test developers must navigate the delicate intricacies when collaborating with various stakeholders to ensure the test's alignment with its intended goals (Kane & Wools, 2019; Rossi, 2021). This collaborative effort culminates in the formulation of the design statement – a blueprint that encapsulates the test's essence, leaving no aspect unaddressed.

Bachman and Palmer's comprehensive framework for the design statement offers invaluable guidance for test developers seeking to create an assessment that leaves no room for ambiguity. It encompasses critical elements such as a clear articulation of the test's purpose, enabling developers to set a distinct direction for the evaluation process. Additionally, the inclusion of the target language use (TLUs) domains and task types fosters a comprehensive approach to assessment, catering to a diverse range of skills and knowledge areas. The design statement also takes into account the test-takers themselves, recognizing the importance of understanding their characteristics and tailoring the assessment to suit their unique needs. Moreover, by defining the construct(s) to be measured, the design statement ensures a focused and coherent evaluation, aligning the test's content with the intended learning outcomes. Furthermore, with a plan for evaluating test usefulness and a strategy for resource management, the design statement exhibits a well-rounded approach that fosters continuous improvement and efficient utilization of available assets.

In summary, the design stage of test development emerges as a crucial phase where informed decisions and comprehensive planning lay the foundation for a successful assessment. By systematically addressing the test's purpose, the TLU domains and task types, the characteristics of the test-takers, the construct(s) to be measured, test's usefulness, and resources in the test design statement, test developers can create effective assessments that align with their

intended purpose and accurately measure the knowledge, skills, and abilities of test-takers. By following the guidelines set forth by Bachman and Palmer's design statement components, test developers can navigate the complexities of the process and create tests that are purposeful, fair, and yield meaningful results for all stakeholders involved (Armstrong et al., 2004).

### ***1.6.2 Test Operationalization***

In the operationalization stage of test development, the focus is on assembling the entire test, including multiple tasks, and detailing the specifications, writing, and scoring of each task (Zandi & Kaivanpanah, 2014). The outcome of this stage is a blueprint for the test, encompassing its structure and scoring materials, along with a draft version of the test itself. The blueprint consists of two parts, as described by Bachman and Palmer (1996, 2010): an overview of the test structure and a set of test-task specifications for individual tasks, which serve as a foundation for item writing and scoring.

The first part of the blueprint provides an overview of the test structure, indicating the number of test parts or tasks used to measure knowledge, ability, and skills, their significance, sequence, and the number of tasks per part (Taylor, 2014; Zandi & Kaivanpanah, 2014). Additional information can also be included. The descriptions of tasks from the Target Language Use (TLU) domain serve as a starting point for constructing test-task specifications, which outline detailed task characteristics needed for writing the actual test. These specifications are essential in creating parallel test forms and evaluating the congruence between intended and actual test content.

Test specifications facilitate constructive discussions and consensus among test developers regarding the final shape of the test. Bachman and Palmer (1996, 2010) specify that test-task specifications include the task's purpose, construct definition, setting characteristics, time allotment, instructions, input and expected response details, the relationship between input and response (reciprocal, non-reciprocal, or adaptive), and the scoring method (Zandi

&Kaivanpanah, 2014). Many of these specifications might already be outlined in the design statement and should be reiterated in the blueprint to guide the test-writing process.

With the blueprint prepared, test writing commences, and instructions for each section are developed. The operationalization phase concludes with a draft of the test, ready to be administered to test-takers.

### ***1.6.3 Test Administration***

The final phase of test development involves test administration, initially to individual students or small groups and subsequently to a larger group of examinees on a trial basis. Piloting the entire test or individual test tasks allows for the collection of response data and other relevant information to enhance the test's usefulness. This data is then analyzed, and the test is revised before its operational use. Before a test, especially a high-stakes one, can be employed with larger groups of test-takers, it must undergo thorough analysis and revision.

In classroom settings, extensive piloting might not always be feasible. In such cases, careful consideration should be given to interpreting the results and making decisions about the test-takers. Nevertheless, if the same tests are used in subsequent administrations, the tasks should be reviewed and the test revised beforehand.

During the actual test administration, it is essential to create a physically comfortable and distraction-free environment that supports the test-takers. Clear instructions should be provided, and the administration process should be organized. This phase also provides an opportunity to gather valuable information about the test-takers' initial reactions to the test tasks and specific test procedures, such as time allocation. Empirical data, obtained through questionnaires or interviews, can be collected from examinees after the test to gain insights into issues like instruction clarity, item quality, and time management, which often evoke strong initial responses (Taylor, 2014).

Test analyses offer valuable information to evaluate the test's characteristics and usefulness, serving as a foundation for further test revisions before its operational



implementation. During the operational phase, additional data is gathered, and iterative and recursive analyses are conducted (Purpura, 2004). Eventually, the accumulated bank of test tasks is archived, providing a valuable resource for future test administrations.

In recent years, the integration of modern computer technology in education has revolutionized the way learning outcomes are assessed. Educational institutions around the world have tried to adopt, or adopted, e-assessment moving towards online or computerized modes of test administration. This shift offers various advantages, including reduced reporting lag time for scores, increased assessment efficiency, flexibility in terms of time and location, immediate feedback to students, and the ability to analyze individualized student performance (Hüseyin & Özturan, 2018). The online administration of assessments and tests helps minimize paper consumption and related costs, making them environmentally friendly (Noyes & Garland, 2008). Computer-based assessments can serve as effective alternatives to traditional testing methods without compromising essential principles such as validity, reliability, and practicality (Hüseyin & Özturan, 2018). Computers offer rich test content, immediate feedback to test-takers, and eliminate subjectivity in scoring, benefiting both students and teachers.

### ***1.7 Integrated Content and Language Assessment***

Traditional assessment procedures have long been criticized, being decontextualized, unauthentic, and negatively impactful. One recurring washback of traditional assessment is that students end up mastering the format of tests, considering their importance, and that of the scores obtained. Embedding instruction into a content-based, learning-oriented, activity serves as an assessment of, for, and as learning. It can shift learners' focus from scores to personal development, fostering motivation and self-esteem, as highlighted by Boubris and Bouabdallah (2023). Rethinking assessment procedures can bring about considerable positive change. In this sense, it is crucial for the whole educational system to work towards the promotion of recent assessment theory and best practices.

Implementing ICL methodologies in ESP contexts offers a robust framework for learner-centered and domain-specific assessment practices that support the development of learners' linguistic and non-linguistic KSAs, culminating in complex ESP competencies. ICL fosters a holistic learning environment where language acquisition is seamlessly integrated into the teaching and learning of specialized content. Linguistically, learners gain proficiency in the specialized vocabulary, grammar, and discourse structures of their chosen field. They develop the ability to comprehend and produce texts, communicate effectively in oral and written forms, and critically analyze specialized information. Non-linguistically, learners acquire a deeper understanding of their chosen field, developing critical thinking, problem-solving, and research skills. They learn to apply their linguistic and non-linguistic KSAs to real-world scenarios, preparing them for future professional endeavors.

Through ICL methodologies, learners cultivate the core competencies essential for success in ESP contexts and domain-related professions. These competencies encompass (1) disciplinary expertise with learners gaining a comprehensive understanding of the concepts, theories, and methodologies of their chosen field; (2) communicative competence with learners developing the ability to communicate effectively in both written and oral forms, using appropriate language for their specific field; (3) critical thinking, enhancing learners' ability to analyze information, evaluate arguments, and form their own conclusions; (4) problem-solving, developing learners' skills to identify and solve problems effectively, applying their knowledge to real-world scenarios; and (5) study skills with learners acquiring the ability to conduct research, evaluate sources, and present their findings in a clear and concise manner (Alaye & Tegegne, 2019; Кошарна, 2020; Pardede, 2020, Snow, 2005).

ICL methodologies emphasize learner-centered and domain-specific assessment practices that align with the learning objectives and the specific needs of ESP learners. These assessment practices go beyond traditional language testing, incorporating a variety of methods to evaluate learners' linguistic and non-linguistic competencies, including authentic, performance-based, and

learning-oriented assessment designs that take into consideration learners' socio-cognition and affect, among many other performance moderators, to provide the best conditions for learning-through-evaluation to occur.

### ***1.7.1 Learning-oriented Assessment***

Studies within the field of LA assessment have predominantly focused on the evaluative processes, encompassing assessment methods and assessment strategies. The Learning-oriented Assessment (LOA) framework, as proposed by Turner and Purpura (2016), offers ICL practitioners an authentic structure to adjust the array of factors influencing or indicating learners' performance. It shifts focus from the commonly known assessment of learning and assessment for learning, to a whole new perspective that puts emphasis on assessment as an agent at the service of learning. Within the LOA framework, two dimensions, among eight, focus on learners' performance, precisely outlining the targeted linguistic proficiency and disciplinary competence. Additionally, the framework specifies the procedures employed to elicit and record this performance. Furthermore, this framework proves valuable in aligning instruction with assessment considerations.

The remaining dimensions in the framework serve as performance moderators, directly influencing learners' performance. These encompass critical elements of the assessment context, specifying key aspects such as the audience, settings, and topic, in addition to the communicative event, target language use-domain, enabling skills, and the culminating competency, among others. Social-cognition is identified as another performance moderator, acknowledging mental processes, cognitive load, and distributed cognition essential for co-constructing knowledge. The framework also takes into account the affective dispositions of learners in relation to the context and the assessment itself, encompassing positive and negative psychological, behavioral, and social aspects.

The instructional dimension focuses on how learners assimilate and process feedback and support to reassess their understanding and enhance their performance. This entails the transmission of new information, such as feedback or assistance, and the method of delivery. At the core of learning-oriented assessments, this dimension incorporates input, support, and feedback as instructional components. It may also include explicit instructional episodes, when necessary, all aimed at fostering learning. The social-interactional dimension examines the sequential organization and exchange patterns within assessments or instructional activities involving interaction. Lastly, the technological dimension takes into account users' characteristics when technology is involved in assessment or instruction, including learners' computer skills and digital literacy.

### ***1.7.2 Meaning-oriented Model of L2 Proficiency***

The ability to acquire and utilize a second or foreign language (S/FL) stands as a testament to the innate capacity for linguistic adaptation. To unravel the complexities of Second Language Acquisition (SLA), scholars have developed various models, each offering unique perspectives on the underlying processes and factors that shape S/FL proficiency. One of the earliest and most influential models is Krashen's Input Hypothesis. This theory posits that comprehensible input or language that is slightly beyond the learner's current level but still understandable, is the primary driving force behind S/FL acquisition (Lichtman & VanPatten, 2021). Another prominent model is The Communicative Language Proficiency (CLP) framework, introduced by the Council of Europe (1996), emphasizing the ability to use language effectively for communicative purposes (Kamiya, 2006).

These models, while diverse in their approaches, provide valuable insights into the intricate processes and factors that influence S/FL proficiency. They highlight the importance of comprehensible input, the interplay between implicit and explicit knowledge, the dynamic nature of language, and the role of individual and contextual factors. Their significance extends beyond

academic understanding as they have practical applications in language teaching and assessment. By understanding the underlying principles of S/FL acquisition, educators can design more effective teaching methods that cater to the diverse needs of learners. Additionally, these models inform the development of assessment tools that accurately measure learners' communicative abilities and progress.

The Meaning-Oriented Model of /FL Proficiency, proposed by Purpura and Dakin (2020), offers a unique perspective on language acquisition, emphasizing the role of meaning and context in shaping learners' proficiency. The model posits that S/FL2 proficiency is not merely about achieving accuracy and fluency; it is about developing the ability to convey and understand meanings effectively in a variety of contexts. This emphasis on meaning aligns with the notion that language is a tool for communication, and that its primary function is to convey ideas, thoughts, and emotions. The model identifies two key components of meaning-oriented proficiency: (1) The Semantico-Grammatical Knowledge that refers to the learner's knowledge of word meanings, grammatical forms, and how they can be applied to create meaningful expressions; (2) The Pragmatic Knowledge which encompasses the learner's understanding of how language is used in different contexts, including the social, cultural, and situational factors that influence communication.

The Meaning-Oriented Model (MOM) of S/FL proficiency highlights the dynamic relationship between these two components, suggesting that they interact and influence each other throughout the language acquisition process. Semantico-grammatical knowledge provides the foundation for understanding and producing language, while pragmatic knowledge guides the appropriate use of language in context. The model also emphasizes the role of learners' individual experiences, beliefs, and attitudes in shaping their meaning-oriented proficiency. Learners bring their unique backgrounds and perspectives to the language learning process, and these factors influence how they interpret and use language.

The Meaning-Oriented Model has significant implications for language teaching and assessment. It suggests that instructors should focus on creating meaningful learning experiences that promote authentic communication and the development of both semantico-grammatical and pragmatic knowledge. Utilizing the Meaning Oriented Model (MOM) of L2 Proficiency offers a basis for establishing specifications within the proficiency dimension of the LOA framework. This model informs curriculum, instruction, and assessment by highlighting the significance of effectively encoding and decoding meanings through the systematic use of language to demonstrate real-world competencies. Emphasizing the use of linguistic resources, MOM underscores the ability to convey both literal and implied propositional meanings in real-world communication. S/FL proficiency, according to this model, encompasses semantico-grammatical knowledge (comprehending the literal meaning of linguistic forms) and pragmatic knowledge (proficiency in using language to express and interpret functional and implicational meanings). The MOM of L2 proficiency becomes a valuable guide for curriculum design, instruction, and assessment, recognizing the pivotal role of language in collaboratively constructing meanings in real-life situations.

### **1.8 Language Assessment Literacy**

Language assessment was set for a major expansion ever since it had made its step into educational research. Assessment has witnessed several theoretical, analytical, and methodological developments that have encouraged educational reforms, particularly in the way learning is monitored and achievements measured. The shift of focus from teacher- to learner-centered approaches has also influenced assessment development trajectory. Research has shown that teachers devote a considerable amount of time on evaluative activities because of the importance attached to the assessment of learners' progression and performance.

Henceforth, there has been a surge of interest in teachers' knowledge and understanding of assessment due their central role in this process – design, implement, and interpret assessment,

besides giving and acting upon feedback (Levi & Inbar-Lourie, 2020; Nimehchisalema & Bhattib, 2019). Mertler and Campbell (2005) have defined language assessment literacy by proposing that teachers who do not “need remediation or assistance in applying assessment concepts and techniques, as well as making assessment-related decisions” are said to possess basic assessment literacy (p. 2). Hence, these teachers are considered as agents of quality teaching. The data they gather among learners is supposedly valid and reliable, providing meaningful and functional feedback.

However, the consensus view seems to be that there is a distinct lack of language assessment literacy among EFL teachers (Davidson & Coombe, 2019; Kavakli & Arslan, 2019; Nimehchisalema & Bhattib, 2019; Şişman & Büyükkarçi, 2019; Sultana, 2019; Ashraf & Zolfaghari, 2018; Jannati, 2015; Volante & Fazio, 2007; Volante & Fazio, 2007; Mertler & Campbell, 2005). Literature has often pointed out this lack as “one of the worst problems in the profession of education today” (Nimehchisalema & Bhattib, 2019, p. 45). Popham’s (2004) original work, entitled *Why Assessment Illiteracy is Professional Suicide*, defends that knowledge about assessment design, implementation, and its further implications, in addition to awareness about key concepts such as assessment validity and washback can determine the quality of education, most often impairing it (Popham, 2004).

Language assessment literacy has been defined in almost every research that explored the concept or its relation to another, such as quality teaching and washback. Broadly speaking, language assessment literacy can be considered as teachers’ evaluative competence, upon which they plan, administer, and apply assessments as well as their outcomes efficiently. Assessment literate teachers will neutralize washback, raise learners’ motivation, and most importantly produce valid and reliable appreciations of learners’ performance (Coombe et al., 2020; Giraldo, 2018; Deuce et al., 2016; Popham, 2013). Teacher competency in educational assessment has henceforth become a professional requirement across the world (Deuce et al., 2016). Furthermore, researchers have argued that language assessment literacy is necessarily teacher-

specific, and to achieve quality teaching policy makers, examination boards, as well as concerned social actors should possess an understanding of language assessment functions and their outcomes across society (Krimmer & Harding, 2020; Diggers & Malone, 2019; Nimehchisalema & Bhattib, 2019; Red, 2019).

In a nutshell, assessment literate language teachers should have the “capacity of asking and answering critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, and about what is going to happen on the basis of the results” (Nimehchisalema & Bhattib, 2019, p. 46). Language assessment illiteracy is dysfunctional in the sense that it hinders the design of reliable assessments, which in turn produces ambiguous feedback upon which invalid instructional decisions would be made. Teachers’ competence in educational assessment, in addition to the translation of this competence into evaluative practices, are major keys to quality teaching.

### ***1.8.1 Language Assessment Literacy Standards***

Teachers’, test developers’, and stakeholders’ knowledge and theoretical beliefs about the measurement of students’ learning is broadly referred to as assessment literacy. The various definitions that have been provided attempted to conceptualize assessment literacy both from a broad and narrow perspective. Language assessment literacy, as described by Davies (2008), has three dimensions: knowledge, principles, and skills. The first refers to a growing awareness about measurement and language. Principles are about the proper use of language assessments and tests, fostering ethical practices, validity, and reliability. The last dimension, skills, refers to the training in necessary and appropriate methodology. Additionally, Filcher (2012, p. 131-132) described assessment literacy as:

The knowledge, skills and abilities required to design, develop, maintain or evaluate large-scale standardized and/or classroom-based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin



practice, including ethics and codes of practice and the ability to evaluate the role and impact of testing on society, institutions, and individuals.

Fulcher (2012) expanded definition of assessment literacy encompasses the contexts, principles, and practices of language assessment. The context includes the historical, social, and political aspects of assessment, with a particular focus on its origin, purpose, and impacts. The principles refer to the theory that guides the practice, including the concepts, methods, and related knowledge, such as advantages and drawbacks. Practices are more concerned with the manifestation of knowledge, skills, and abilities related to the practice of language assessment and testing.

In his review of the international standards and measure of teacher assessment literacy, DeLuca et al. (2016a) have analyzed fifteen assessment literacy governmental standards, in Australia, Canada, New Zealand, UK, and US, in addition to Europe. They have identified, through a thematic analysis, eight prominent research-based assessment measures, which are:

(1) Assessment Purposes refers to choosing the appropriate form of assessment based on clearly stated instructional goals. (2) Assessment Processes encompasses constructing, administering, and scoring assessment and interpreting assessment results to facilitate instructional decision-making. (3) Communication of Assessment Results entails communicating assessment purposes, processes, and results to stakeholders. (4) Assessment Fairness involves cultivating fair assessment conditions for all learners with sensitivity to student diversity and exceptional learners. (5) Assessment Ethics means disclosing accurate information about assessments and protecting the rights and privacy of students that are assessed. (6) Measurement Theory focuses on understanding psychometric properties of assessments (e.g., reliability and validity). (7) Assessment for Learning describes the use of formative assessment during instruction to guide teacher practice and student learning. (8) Assessment Education and Support for Teachers represents supporting teachers' assessment competency through explicit education opportunities or resources (p. 13).

DeLuca et al., (2016a) have highlighted the evolution of assessment literacy measures during the last three decades. From 1990 to 1999, assessment purposes, processes, fairness, and communication of assessment results were considered as crucial competencies for teachers. Governmental documents concentrated on teachers' abilities to select and use summative and standardized assessments, to produce fair educational decisions. The next decade, from 2000 to

2009, was marked by the emergence of novel measures, linked to assessment for learning and learner-centered approaches to teaching. The necessity for teachers to display new competencies and update their assessment theory has led to the emergence of *Assessment Education and Support for Teachers*, as a measure. Nowadays, assessment standards express a major emphasis, not only on the early measures of assessment competency, but *Assessment for Learning* and *Assessment Ethics* as well.

### **1.8.2 Impact of Training and Education programs on Teachers' Assessment**

#### ***Literacy***

DeLuca et al. (2016b, p 2) have reported that “recently published assessment standards articulate contemporary approaches to classroom assessment,” which underscores the pivotal role of assessment literate teachers in the success of the teaching-learning experience, and the decision making process. Research on EFL teachers' assessment competency has nevertheless shown that most teachers failed to articulate a research-based, objective, knowledge about assessment (Davidson & Coombe, 2019; Kavakli & Arslan, 2019; Nimehchisalema & Bhattib, 2019; Şişman & Büyükkarci, 2019; Sultana, 2019; Ashraf & Zolfaghari, 2018).

The majority of studies into language assessment literacy have shown that teachers' knowledge about assessment is narrowly linked to their beliefs about assessment (Davidson & Coombe, 2019; Nimehchisalema & Bhattib, 2019; Şişman & Büyükkarci, 2019; DeLuca et al., 2016a). Boubris and Haddam (2020) define these beliefs as “teachers' subjective and experience-based knowledge, forming their views and arguments concerning teaching and learning” (p. 238). However, since they lack objectivity, teachers' beliefs are often dysfunctional, and tend to impede good teaching, learning, and assessment.

Researchers and scholars are giving significant attention to assessment literacy training programs and professional development in the field of education. These programs are recognized for their substantial impact on the success of both teachers and students. Consequently, there is a

growing emphasis on designing effective professional development and training programs for teachers to yield valid outcomes and enhance the quality of learning and teaching. These initiatives are also aimed at addressing the challenges posed by global competition in educational performance.

Furthermore, it is acknowledged that training and development initiatives can provide teachers with improved opportunities to acquire knowledge, information, skills, and competences necessary for effective performance. These enable teachers to practice more effectively, leading to enhanced productivity, fostering positive attitudes among (Giraldo, 2019). Training programs can influence teachers' performance and professional development by improving their understanding of educational contexts and their ability to achieve desired outcomes. Furthermore, well-trained assessors play a functional role in achieving the targeted outcomes.

Research on language assessment can play a vital role in promoting assessment literacy. By providing teachers with a deeper understanding of the different types of assessments, their purposes, and how to select and use them appropriately, trainers can participate actively in the readjustment of teachers' dysfunctional beliefs that can lead to negative washback (Boubris & Haddam, 2020). This knowledge is essential for teachers to be able to design and implement useful and effective assessments that support learning. Teacher development programs can also help teachers to develop a more critical understanding of the assessment system in which they are working. This includes understanding the strengths and weaknesses of the system, as well as the potential impact of assessments on students' motivation and learning outcomes. This knowledge can help teachers to identify areas where the assessment system needs to be improved and to advocate for change. Overall, research on language assessment and teacher development programs can play a vital role in promoting assessment literacy and guiding teachers' beliefs and practices.

## **1.9 Conclusion**

Assessment is an integral component of the Algerian educational system, deeply intertwined with teaching and learning. Its implications are multifaceted, extending beyond the mere measurement of student achievement. Algerian teachers are increasingly recognizing the importance of aligning teaching methods with learning outcomes, learners' needs, and the learning environment. Assessment plays a vital role in this process, providing valuable insights into students' progress and enabling teachers to make informed decisions about instruction.

In the context of Foreign Language Teaching (FLT), assessment literacy is particularly significant, given the theoretical and methodological developments in this field. Algerian FLT teachers need to be equipped with the knowledge and skills to design, implement, and interpret assessments effectively. Unfortunately, there is a lack of assessment literacy among Algerian teachers, attributed to inadequate training, limited resources, and heavy workloads. This deficiency leads to an overreliance on traditional assessment methods, even when they do not represent serve the learning objectives.

Research on the impact of assessment in Algeria is limited, particularly regarding its effects on student motivation and learning outcomes. However, some studies suggest that the current assessment system, characterized by high-stakes exams, can be demotivating and hinder learning. There are several gaps in research on language assessment in Algeria, including insufficient research on assessment literacy, assessment impact, the design and delivery of language assessments, and the assessment of multilingual and multicultural classrooms. Addressing these gaps is essential for developing more effective and equitable assessment practices that support the diverse needs of Algerian students.

While Communicative Language Teaching (CLT) is the official approach to FLT in Algeria, many teachers still favor a "teach-to-the-test" approach. This may be attributed to the high stakes associated with standardized exams, such as the Baccalaureate examination (Benmoussat & Benmoussat, 2018). However, the "teach-to-the-test" approach may not align

with the principles of CLT, which emphasizes the development of communicative abilities and language proficiency.

Research studies on the washback effects of standardized exams in Algeria reveal both positive and negative influences on teaching practices. Some teachers adapt by incorporating communicative-oriented tasks to enhance students' communicative competence, while others resort to test-like activities to boost exam scores. Current assessment practices in Algeria have several limitations. Achievement tests, for example, often fail to measure students' higher-order thinking skills and communicative abilities. Additionally, there is a need for a more systematic approach to assessment that provides students with regular feedback and opportunities for improvement. The combination of construct-based and task-based assessments can play a crucial role in addressing these limitations.

Assessment practices in Algeria play a pivotal role in shaping education, influencing teaching methods, curriculum design, student motivation, and overall learning experiences. The multifaceted impact of assessments underscores the importance of conducting further research to address existing gaps and enhance assessment literacy among teachers. Additionally, aligning assessment practices with the principles of CLT and recognizing the broader implications of standardized exams are essential steps towards improving the quality of education in Algeria. Ultimately, understanding and addressing the impact of assessments in the Algerian context is imperative for the advancement of education and the optimization of learning outcomes.

Some of the key challenges facing Algeria in terms of assessment practices include the scarcity of assessment literacy among teachers, the overreliance on traditional assessment methods, and the high-stakes nature of standardized exams, in addition to the gaps in research on language assessment in the Algerian context, and the mismatch between the official approach to FLT (CLT) and the actual assessment practices used by many teachers. To address the challenges facing Algeria in terms of assessment practices, the following recommendations can be made, based on the literature: (1) provide teachers with training on assessment literacy and

## ***CHAPTER ONE***      ***Higher Education Language Assessment***

---

ICLT principles; (2) design assessments according to standards that measure students' higher-order thinking skills, TLU domain communicative abilities, and integrated skills; (3) reduce the emphasis on high-stakes exams and shift towards a more formative and/ or learning-oriented approach to assessment; (4) conduct further research on assessment in the Algerian context.

## **CHAPTER TWO: Situation Analysis and Research Design**

### 2.1 Introduction

### 2.2 Situation Analysis

#### 2.2.1 *ESP in Algeria*

#### 2.2.2 *ESP assessment in the Algerian HEIs*

### 2.3 Theoretical Framework of the Research

#### 2.3.1 *Research Design*

#### 2.3.2 *Research Objectives*

### 2.4 Research Instruments

#### 2.4.1 *An Overview*

#### 2.4.2 *Document Analysis*

##### 2.4.2.1 *Purpose*

##### 2.4.2.2 *Procedure*

##### 2.4.2.3 *Conclusion*

#### 2.4.3 *Classroom Observation*

##### 2.4.3.1 *Purpose*

##### 2.4.3.2 *Procedure*

##### 2.4.3.3 *Observational Guide*

##### 2.4.3.4 *Conclusion*

#### 2.4.4 *Learners' Survey*

##### 2.4.4.1 *Purpose*

##### 2.4.4.2 *Description*

##### 2.4.4.3 *Administration*

##### 2.4.4.4 *Conclusion*

#### 2.4.5 *The Test*

##### 2.4.5.1 *Purpose*

##### 2.4.5.2 *ICL Test Design*

##### 2.4.5.3 *Description*

##### 2.4.5.4 *Administration*

##### 2.4.5.5 *Conclusion*

#### 2.4.6 *Conclusion*

### 2.5 Sampling Procedures

#### 2.5.1 *Defining the Target Population*

#### 2.5.2 *Defining the Source Population*

#### 2.5.3 *The Sampling Frame*

#### 2.5.4 *Sampling Method*

#### 2.5.5 *Sample Size*

#### 2.5.6 *Conclusion*

### 2.6 Data Analysis

#### 2.6.1 *Qualitative Data Analysis*

#### 2.6.2 *Quantitative Data Analysis*

### 2.7 Ethical Considerations

### 2.8 Conclusion

## **2.1 Introduction**

In pursuit of contributing to academic discourse on assessment in the Algerian context and abroad, this dissertation undertakes an exploration of teachers' evaluative practices role in the development of learners' language skills. The researcher has to explore the context according to which his strategy will be deployed. This research focuses on the Algerian Higher Education, with a particular interest in ESP classrooms within the Faculty of Sciences, Tlemcen University. Moreover, the foundation of any research effort lies in its methodology, influencing the study's direction and the credibility of its findings. The review of the literature has shown that researchers tend to favor mixed-methods design to explore the dimensions of impact and washback in LA.

In this chapter, we will first explore the ESP situation in the Algerian HE context, with a particular focus on language assessment to provide an ontological perspective of the problem showing that assessment and language skills development do not exist in isolation. The second part of this chapter explores the framework underpinning the study, focusing on research design, data collection and analysis methods, as well as ethical considerations guiding the research process. It ensures transparency and rigor in the research process.

## **2.2 Situation Analysis**

### ***2.2.1 ESP in Algeria***

English is becoming increasingly important in the Algerian education system, particularly in the technical, scientific, and medical fields (Ouahmiche et al., 2017). This is evident in the shift towards English for Specific Purposes (ESP) courses in these areas (Maarouf & Lamouri, 2022). ESP is a type of language teaching that is tailored to the specific needs of learners in a particular field. In the context of the Algerian education system, ESP courses are intended to



help students develop the English language skills they need to succeed in their academic studies and future careers.

There are a number of reasons why ESP is becoming increasingly important in Algeria. First, English is the global language of science and technology. This means that students who have a strong command of English have access to a vast body of knowledge and resources that are not available in other languages. Second, Algeria is increasingly integrated into the global economy. As a result, there is a growing demand for Algerian professionals who are able to mobilize their linguistic and non-linguistic skills simultaneously to achieve both communicative and professional purposes.

The Algerian higher education is responding to this demand by offering more ESP courses at almost all levels of university (Ouahmiche et al., 2017). ESP courses are now available in a wide range of fields, including engineering, medicine, business, and law. These courses are supposedly taught by qualified ESP teachers who have expertise in both the English language and the relevant field of study. However, as pointed out by Assassi (2020), English for Specific Purposes (ESP) courses at the University of Mohamed Khider in Biskra, whether offered by the CEIL (Centre for English for Instrumental Learning) or different departments, do not align with the essential criteria of ESP classes. Instead, they predominantly function as English for General Purposes (EGP) courses.

The lack of alignment between language instruction and learners' professional needs may not provide significant benefits for their current or future careers. Throughout his work, the researcher pointed out a critical issue: “The four teachers believe that assigning newly graduated EFL students as part-time teachers in different departments without providing any clear syllabus or training is not pedagogical” (Assassi, 2020, p. 445). He recommends afterwards to “Avoid issues such as hiring part-time teachers who mostly do not show commitment to their duties

giving the very low salaries” (Ibid, p. 453). His study reveals “one of the weighty weaknesses in these courses, as they are not tailored according to learners’ needs as no needs analysis process was carried, showing that “no teacher among the ones observed has been trained to be an ESP teacher or has executed any procedure related to ESP course design” (Ibid, p. 439).

Akkar and Idri (2021) among others have reported on the situation of ESP in the Algerian context very similarly. According to Fehaima (2022), ESP teachers at the Faculty of Sciences, Tlemcen University are: “language instructors who have been assigned to teach ESP courses despite having no professional training” (p.250). Some even see that: “ESP, in particular, still has a long journey to be well-established and well-implemented in different specialties” (Hadj Djelloul & Melouk, 2022, p. 157). From Khadam's (2023) analysis of ESP courses at Bechar University, Algeria, various issues have appeared that call for consideration.

The requirement for competent instructors well-versed in ESP teaching methodologies and subject-specific expertise posed a hurdle to the administrative staff, given the potential difficulty in locating suitable candidates willing to do serve as part-time teachers regardless of the financial problematic (Akkar & Idri, 2021). In her paper, Khadam (2023) reported that:

ESP courses, not only in the university under study but in most of the Algerian universities are not taught in the universal norms. Furthermore, the obtained results showed that the teaching of ESP in the Algerian universities failed to satisfy the requirements of its learners because of a number of factors such as the administrative constraints, the lack of the suitable teaching materials, the lack of cooperation between the ESP tutors and the subject matter specialists, the timing allocated to the ESP modules and the poor language knowledge of the students (p. 482).

High quality ESP instruction and assessment can help students to develop the English language skills they need to succeed in their academic studies and future careers. This will make them more competitive in the global job market and contribute to the economic development of Algeria. There is an imperative need to establish a pedagogical framework that aligns with the evolving English language needs within the context of globalization. As researchers have pointed

out, this transformation requires innovative strategies and a shift in mindsets. Universities, as influential entities in the socio-economic development of the nation, should spearhead this change. The language policy should focus on generating opportunities for field development. The Ministry of Higher Education and Scientific Research should play a pivotal role in enhancing this fundamental aspect of the ESP teaching policy due to its significant contribution to socio-economic development.

### ***2.2.2 ESP assessment in the Algerian HEIs***

One of the biggest challenges to ESP assessment in Algeria is the lack of qualified ESP teachers. ESP teachers need to have expertise in both the English language and the relevant field of study. However, many ESP teachers in Algeria are newly graduated EFL students who have neither received any training in ESP teaching methodologies nor in assessment theory. This reveals another challenge to ESP assessment in Algeria, which stems from the lack of alignment between ESP courses and learners' professional TLU domains needs, which is expressed here in the interrogative: what are part-time English language instructors in ESP contexts assessing?

Many ESP courses in Algeria are not tailored to the specific needs of the learners. This is because no needs analysis is typically conducted before ESP courses are designed (Faiza, 2010). In this sense, whether or not assessment is aligned with instruction is irrelevant, since instruction itself is not aligned with real-life professional needs. In addition, there is a lack of cooperation between ESP tutors and subject matter specialists in Algeria. This lack of cooperation can make it difficult to ensure that ESP assessments are aligned with the needs of the stakeholders and the needs of the international job industry. The limited amount of time allocated to ESP modules is another challenge to ESP assessment in Algeria. ESP modules are often only one or two hours per week, which makes it difficult for ESP teachers to cover all of the necessary material and to assess learners' progress effectively. Finally, the poor language knowledge of some students is

also a challenge to ESP assessment in Algeria. Many students who take ESP courses have a limited command of the English language. This can make it difficult for them to complete ESP assessments successfully.

The lack of research and guidance on ESP assessment in Algeria is a problem, as it can make it difficult for ESP teachers to design and implement useful assessments. Without clear assessment procedures, it is difficult to measure students' learning and ensure that they are developing the English language skills they need to succeed in their academic studies and future careers. As Saraa (2023) pointed out:

The specific English language needs of the Algerian students at higher education have received considerable attention among Algerian scholars and teachers (Mebtil,2012; Boudresa, 2018; Bouroumi, 2017). However, only few research works investigate the issue of ESP syllabuses in Algerian higher education (Bouزيد, 2012; Saraa, 2020). Besides, the review of the Algerian biology ESP syllabus at higher education shows that this document commonly referred to as Canvas is very poor in terms of content specification, methodological guidance, teaching materials, and assessment procedures (p. 511).

Abbassi's (2022) research offers crucial insights into the challenges surrounding (ESP) assessment in the Algerian context. Her study featured six ESP teachers from the Chemistry department, Tlemcen University. The research uncovered a significant challenge in the preparedness of ESP teachers. Most instructors, originally trained for general English, found themselves teaching ESP without the necessary training. This lack of formal preparation hampers their ability to integrate language and subject-specific content effectively. According to Abbassi's (2022) work, classroom assessment methods in ESP vary, with teachers employing a mix of traditional and advanced techniques. The study revealed reluctance among teachers to adopt alternative assessment models like self/peer assessment and portfolios potentially due to a lack of qualifications, creativity, or resources. The absence of a unified syllabus further highlights the need for teacher training to be operational in such a context.

Abbassi's (2022) research underscores the pressing challenges posed by the readiness of ESP teachers in Algeria, who often lack the necessary training for effective ESP instruction and assessment. This gap between teacher qualifications and the demands of ESP education calls for proactive measures, such as comprehensive training programs and support for educators. Additionally, the study highlights the need to diversify assessment methodologies and promote innovative approaches to align with learners' needs effectively.

According to Benabdi (2022), most ESP instructors in Algeria typically design their own assessment materials, and the Algerian higher education system does not equip them with the necessary skills for ESP assessment design. This poses a substantial problem because useful ESP assessment should ideally be tailored to the precise needs of the learners they serve. Furthermore, the prevailing assessment approach in Algerian ESP courses predominantly leans towards summative evaluation, emphasizing end-of-semester examinations. Continuous assessment, which is better suited for ESP courses, as it enables ongoing monitoring of students' progress and the provision of timely feedback, is given limited emphasis. As all of the fundamental aspects of any ESP syllabus, assessment methods and procedures must be considered in advance (Saraa, 2020).

To address the multifaceted challenges associated with ESP assessment in Algeria, potential research avenues emerge. These include investigations into the effectiveness of diverse ESP assessment methodologies within the Algerian context, the formulation of guidelines for ESP assessment tailored to the Algerian higher education landscape, and initiatives focused on training ESP educators in assessment methodologies. Furthermore, there is room for researchers to delve into the development and evaluation of ESP course materials that precisely cater to the requirements of Algerian learners. By embarking on these research initiatives, scholars can play a pivotal role in enhancing the quality of ESP assessment, thus empowering ESP learners to

cultivate the English language proficiency requisite for their academic pursuits and future professional endeavors.

In summary, the increasing importance of English within the Algerian higher education system, particularly in technical, scientific, and medical domains, is indisputable. This paradigm shift towards English for Specific Purposes (ESP) courses signifies a crucial response to the growing global demand for fluent professionals. However, while ESP courses are proliferating across various fields in Algerian universities, it is evident that numerous challenges hinder their alignment with the fundamental criteria of ESP classes. This disconnection raises concerns about the extent to which learners can truly benefit from these courses, particularly when it comes to addressing their professional needs effectively.

A significant impediment to ESP success in Algeria lies in the shortage of qualified ESP teachers. The intricacies of ESP require instructors well-versed in both English language proficiency and the intricacies of the relevant field of study. However, the prevalent practice of employing newly graduated EFL students as part-time teachers without comprehensive training or clear syllabi compromises the pedagogical effectiveness of these courses. Furthermore, a lack of needs analysis in ESP course design contributes to the misalignment of instruction with real-world professional requirements, rendering assessment concerns more than critical.

ESP assessments in Algeria encounter a host of challenges. Many ESP courses are not tailored to learners' specific needs due to the absence of needs analysis, and limited cooperation between ESP tutors and subject matter specialists impedes alignment with international job industry requirements. The constrained time allocation for ESP modules further complicates effective teaching and assessment. Additionally, some students' limited English language proficiency presents hurdles in successfully completing ESP assessments. Despite these challenges, there remains a dearth of research and guidance on ESP assessment within Algeria.

Furthermore, it is important to underscore the notable scarcity of research dedicated to ESP assessment within the Algerian context. While ESP course design, instruction, and needs analysis have gained prominence in the country's higher education landscape, there remains a noticeable gap in the literature regarding assessment design, scoring, and overall usefulness. This lack of research not only limits the availability of well-informed guidance for ESP educators but also hinders the establishment of clear assessment procedures. Without a robust body of research dedicated to ESP assessment within Algeria, educators may find it challenging to design and implement assessments that effectively measure students' learning and foster the development of the English language skills necessary for their academic pursuits and future professional endeavors. Addressing this research gap is crucial to ensure the continued improvement of ESP evaluative practices in Algeria and the overall success of ESP learners.

Addressing these multifaceted challenges necessitates a concerted effort to enhance the quality of ESP assessment. Research avenues must explore the effectiveness of diverse ESP assessment methodologies within the Algerian context, formulate tailored assessment guidelines, and provide training to ESP educators in assessment methodologies. Additionally, there is a pressing need to develop ESP course materials that precisely cater to Algerian learners' unique requirements. By undertaking these research initiatives and proactively addressing these challenges, scholars can empower ESP learners to cultivate the English language proficiency essential for their academic and future professional pursuits.

## **2.3 Theoretical Framework of the Research**

### ***2.3.1 Research Design***

Research design functions as the roadmap that ensures a research journey unfolds with clarity, rigor, and purpose, yielding meaningful insights and that can significantly contribute to the broader academic discourse. It forms the fundamental framework upon which the entire

research endeavor is constructed, serving as the essential blueprint guiding the systematic exploration of a research question or issue (Leavy, 2022). Its paramount importance lies in its profound influence on the quality, validity, and reliability of research findings. Within the realm of research design, critical components such as the selection of data collection methods, definition of variables and constructs, delineation of the study's scope and limitations, and the establishment of ethical protocols play integral roles. Moreover, it is imperative to recognize the intricate interplay between research objectives and research design, as the design choices must align closely with the specific aims and goals of the study (Cohen et al., 2002).

The research design serves as the conceptual framework shaping data collection and analysis that has to align with the research questions, objectives, and context (Cohen et al., 2002). The present research work is based on a sequential exploratory case study. The significance of case studies in research lies in their deployment when issues with regard to education are explored As Zainal (2007). The sequential mixed-methods approach was chosen, combining qualitative and quantitative instruments to offer a comprehensive understanding of evaluative practices in language skills development (Alam & Aktar, 2019; Gogo & Musonda, 2022; Schoonenboom & Johnson, 2017). Qualitative content analysis and quantitative surveys provide diverse perspectives, enabling a thorough exploration of the correlation between the evaluative practices and language skills development.

Based on Algozzine and Hancock (2017), this research is a case study because it involves an in-depth and holistic investigation of assessment impact on language skills development within its real-life context. It centers on the specific case of the Faculty of Sciences at Tlemcen University, Algeria, and its population. This research draws attention to the context in which the case unfolds. Case studies can rely on either quantitative data, qualitative, or both through mixed-methods (Starman, 2013).



The exploratory sequential technique is a progressive strategy employed when quantitative results are complemented by qualitative data. Consequently, quantitative data is utilized to analyze and elucidate the results in a sequential manner (Gogo & Musonda, 2022). This approach stands apart in that it prioritizes the exploration of a problem before its validation, thereby enhancing flexibility in uncovering innovative ideas within the qualitative data (Almeida, 2018). Moreover, the quantitative research instruments used in the present study were informed by the previously collected qualitative data, which is also a characteristic of sequential mixed-methods design (Almeida, 2018). As stated by Berman (2017) "The exploratory sequential mixed methods design is characterized by an initial qualitative phase of data collection and analysis, followed by a phase of quantitative data collection and analysis, with a final phase of integration or linking of data from the two separate strands of data" (p. 1).

The choice to employ a sequential mixed-methods approach case study stems from the desire to achieve a more holistic and nuanced understanding of the complex phenomenon that is assessment and its impact on learners' language skills development. In the context of this research, the choice of a mixed-methods approach acknowledges that the phenomenon of impact is multifaceted and dynamic. From an ontological standpoint, it is recognized that there are multiple layers of reality and various facets to the issue. Language skills development and assessment do not exist in isolation but are shaped and linked by a multitude of factors, including assessment standards, individual experiences, educational contexts, and political frameworks. By incorporating both qualitative and quantitative methods, this research recognizes the complexity of these interwoven realities.

Moreover, the choice of a mixed-methods approach reflects an epistemological stance that acknowledges the value of multiple ways of knowing and understanding a phenomenon. Assessment influence on language skills development can be understood through both qualitative exploration and quantitative analysis. From a qualitative perspective, artifact analysis and

classroom observation provide a solid ground on which logical inferences about the expected impact can be drawn. Thus, direct correlations can be established. This qualitative dimension allows us to uncover the context-dependent aspects of knowledge, then matching them with theory. It recognizes that the understanding of language skills development and assessment is influenced by factors which cannot always be captured quantitatively.

On the other hand, from a quantitative standpoint, learners' survey generates empirical data that can be analyzed statistically. This approach provides insights into broader patterns, trends, and correlations related to assessment impact. It contributes to a more generalizable understanding of the phenomenon, confirming correlations and highlighting aspects that might not be readily apparent through qualitative means alone. Ultimately, in order to effectively address the research problem, confirm the research hypotheses, and consider recommendations through quantitative data, a test was administered to a group of learners. Tests are also purposeful instruments for information gathering in research endeavors. Tests can facilitate balanced judgments in knowledge dissemination and application (Olajide, 2018).

By combining both qualitative and quantitative methods, the research acknowledges that knowledge about evaluative practices in language skills development can be enriched when these different ways of knowing are integrated. It recognizes that the convergence and divergence between qualitative and quantitative findings can offer a more robust and comprehensive understanding of the research questions (Almeida, 2018; Gogo & Musonda, 2022; Taylor, 2005). In summary, the choice of a mixed-methods approach in this research aligns with both ontological and epistemological considerations. It acknowledges the complexity of the phenomenon under investigation, the multiple facets of reality shaping it, and the value of diverse ways of knowing to gain a more holistic and nuanced understanding of the relation between assessment and language skills development.

**2.3.2 Research Objectives**

The assessment of students' performance in educational settings plays a pivotal role in shaping the overall teaching-learning experience and determining the quality of educational outcomes. In this context, the impact of assessment on language skills development has garnered significant attention within the broader field of educational research. Numerous studies conducted in various educational contexts worldwide have delved into the nature and intensity of impact, providing valuable insights into how assessment practices influence learners and other stakeholders involved in the educational process. However, when we turn our focus to the Algerian higher education context, we find a noticeable gap in the existing body of research. In particular, research related to assessment, whether in general or with a specific emphasis on its impact, remains scarce in this context.

This lack of research is particularly striking given the vital role that language proficiency plays in the preparation of Algeria's future international workforce. In institutions such as the University of Tlemcen, where students embark on their higher education journeys, the development of strong language skills is of paramount importance. It is essential for both academic success and future employability, as proficiency in languages opens doors to a wide range of career opportunities in an increasingly interconnected world.

The need for research in this domain becomes even more pressing when we consider the ever-evolving landscape of education, both globally and locally. The way assessments are designed, administered, and scored has a profound impact on students' experiences and learning outcomes. Therefore, it is crucial to understand how assessment practices are influencing language skills development in the Algerian higher education system. Furthermore, this research gap calls for action research specifically, where research is not only conducted to gain insights but also to directly inform and improve evaluative practices. By actively engaging with the

assessment processes within the Faculty of Sciences at the University of Tlemcen, we can ensure that our findings translate into tangible improvements that benefit both educators and learners.

Turning to the objectives of our research, our primary aim is to investigate the impact of assessment on first-year students at the Faculty of Sciences, Tlemcen University. To achieve this, we intend to explore two main aspects of assessment: formative and summative. Classroom formative assessments, intended to provide continuous feedback and support learning, will be analyzed to understand how they contribute to language skills development, including listening, speaking, reading, and writing. On the other hand, we will evaluate the influence of formal summative assessments, which measure overall learning outcomes, on these same language skills.

In the pursuit of understanding the impact of assessment, we will not only analyze existing evaluative practices but also actively engage with students. Their feedback, experiences, and perceptions regarding the impact of assessments on their language skills development are invaluable. By incorporating them into our research, we aim to provide a more comprehensive and balanced perspective.

In conclusion, this research endeavors to address the current research gap related to ESP assessment situation within the Algerian higher education context, with a particular focus on how it influences language skills development among 1<sup>st</sup> year Computer Science (I and MI), Science and Technology (ST), and Matter Sciences (SM), at Tlemcen University. Through an investigation within the Faculty of Sciences, we aspire to contribute to the enhancement of higher education practices in Algeria and offer further insight into the Algerian higher education ESP current situation. The findings have the potential to inform pedagogical strategies, promote quality education, enhance student learning experiences, and support further research in this vital area. The core motivation behind this research lies in the belief that by examining how

assessment affects the development of language skills, we can actively contribute to the advancement of higher education objectives in Algeria and ensure that students are well-prepared for the challenges and opportunities of the modern world.

## **2.4 Research Instruments**

### ***2.4.1 An Overview***

In educational research, research instruments are fundamental tools utilized to systematically and rigorously collect data on a diverse array of topics. These instruments play a pivotal role in facilitating the collection of data from a large set of participants in a consistent and reliable manner. This data serves as the foundation for addressing research questions and producing evidence-based recommendations to enhance educational practices.

There exists a variety of research instruments, offering unique strengths and limitations. Commonly employed instruments include surveys, which entail questionnaires administered to a substantial number of participants through various means, including in-person, over the phone, or online (Hinds, 2002). Interviews facilitate one-on-one conversations between researchers and participants, delving deep into their experiences, beliefs, and attitudes. Observations involve the systematic observation and recording of participants' behavior in natural settings, shedding light on interactions between participants and their environment. Additionally, tests are employed to measure participants' knowledge, skills, or abilities, and these tests can be administered individually or to groups (Cohen et al., 2017b).

The choice of the most appropriate research instrument narrowly linked to the specific research question at hand. For instance, surveys might be the tool of choice when seeking insights into students' attitudes toward a given curriculum, instruction, or assessment, while observations are more suitable for understanding how teachers implement specific procedures and plans. It is imperative to acknowledge that all research instruments come with a degree of

error, stemming from various factors such as design, administration, and participant characteristics (Hinds, 2002). Researchers must remain vigilant about these potential sources of error and take measures to minimize their impact on study outcomes.

This case study is designed to investigate the influence of teachers' evaluative practices on the development of learners' language skills. To address the initial research question concerning the nature of these evaluative practices, we gathered data by examining final exam artifacts and conducting classroom observations. Building upon the insights gained from these initial research instruments, we used a learners' survey to focus on learners' perspectives about specific aspects of language skills development and their correlation with evaluative practices, considering the core of our second research question. The third research question, which primarily pertains to enhancing ESP assessment methods, was addressed through the administration of a test. This test was also constructed based on the findings from the analysis of artifacts and classroom observations. Notably, this study follows a sequential exploratory case study approach, where qualitative instruments inform the development and implementation of quantitative ones, facilitating a comprehensive investigation. The deployment of quantitative and qualitative research tools such as surveys and interviews is what has allowed for what we now know about washback, or impact, in both high-stakes exams and low-stakes classroom assessment (Green, 2013; Rahman 2017).

#### ***2.4.2 Document Analysis***

Document analysis is a fundamental qualitative research instrument defined as the "systematic procedure for reviewing or evaluating documents—both printed and electronic (computer-based and Internet-transmitted) material" (Bowen, 2009, p. 27). It offers valuable insights into the experiences of individuals who have had direct involvement with a particular subject. In the process of developing their studies, researchers rely on established sources to

gather ideas and evidence that substantiate their claims. This approach enables researchers to evaluate the content, quality, and purpose of the documents they select, ensuring that the information contained therein is pertinent to their research objectives.

Researchers typically engage with three primary types of documents in their research endeavors: public records, personal documents, and physical evidence. Public records encompass official records of organizational activities. Personal documents comprise firsthand accounts of individual experiences, such as diaries, emails, and social media posts. Physical evidence encompasses objects found within the research context, including flyers, posters, and instructional materials.

As stated by Morgan (2022): "Document analysis has been an underused approach to qualitative research. This approach can be valuable for various reasons. When used to analyze pre-existing texts, this method allows researchers to conduct studies they might otherwise not be able to complete." (p. 64). This research tool plays a pivotal role in researchers' endeavors as it allows them to discern the most relevant and applicable resources for their studies. These documents offer qualitative data, which encompasses non-numerical information categorized into various themes or concepts. Researchers can seamlessly integrate document analysis with other research instruments to gain a more comprehensive understanding of their subject matter (Morgan, 2022).

Furthermore, they leverage documentary material to provide context, generate pertinent research hypotheses, suggest observations, offer additional insights, and corroborate evidence from other research sources. Researchers often employ data obtained through document analysis in triangulation, comparing multiple data sources to establish credibility. This technique also finds its place in mixed-method studies, combining quantitative and qualitative data to create a synergistic relationship that enhances data utilization. It is very common that "documentary

evidence is combined with data from interviews and observation to minimize bias and establish credibility" (Bowen, 2009, p. 38).

The process of document analysis involves several key steps to ensure a systematic and comprehensive evaluation of the chosen documents. Researchers initiate the process by listing their selected resources, taking into consideration inclusion criteria, which comprise the medium, the genre, and other parameters. Next, they determine how to organize the information, identifying units of meaning and sets of categories that will facilitate a structured analysis. Making copies of the documents allows researchers to annotate and highlight key information without compromising the integrity of the original sources. Considering factors such as authenticity, representativeness, credibility, and meaning, when selecting documents, is crucial to ensure that reliable sources are used in the research (Morgan, 2022).

Researchers should also remain vigilant for biases present in documents, including confirmation bias, culture bias, question-order bias, halo effect, and wording bias (Bowen, 2009). Asking pertinent questions about the document's background, content, origin, and relevance aids in its effective analysis. Finally, researchers evaluate the document's data, seeking patterns, valuable information, and interpretations that align with their research questions and objectives. In summary, document analysis is a versatile and essential research tool that empowers researchers to extract meaningful insights from diverse sources, contributing to the advancement of knowledge across various disciplines.

#### ***2.4.2.1 Purpose***

The primary aim of this research is to investigate the impact of assessments on the development of learners' language skills within the specific TLU domains. To achieve this, a mixed-methods research approach is employed, combining document analysis with various other



research instruments. This mix of research methods is designed to enhance the credibility of the findings, ensuring a robust and thorough exploration of the research topic.

Document analysis plays a pivotal role in this research, serving as a qualitative tool to unravel the complex influence of summative assessments on language skill development. By scrutinizing assessment artifacts, we gain valuable insights into the strengths, weaknesses, and areas for improvement within this particular context. These artifacts, whether physical or digital, represent tangible evidence of targeted/tested proficiency and competency. Through artifact analysis, we delve deeper into how educators employ assessment tools and how these evaluative practices may promote, or hinder, the development of language skills.

The specific focus of our exploration centers on 1st-year Computer Science (I and MI) students at the Faculty of Sciences. Through comprehensive assessment artifact analysis, the researcher aims to discern correlations between learners' language skills development and evaluative procedures. This systematic examination seeks to address our first research question: "How do ESP teachers at the Faculty of Sciences assess 1<sup>st</sup> Year Computer Science students?" Through evidence analysis and other research methods, our research endeavors to provide substantive answers and contribute valuable insights to the field of language assessment in the Algerian context. Overall, assessment artifacts analysis was carried out to provide direct information about summative assessments design, delivery, scoring, and usefulness, to allow for an estimation of the impact they may have on learners' language skills development, at the Faculty of Science, Tlemcen University.

#### ***2.4.2.2 Procedure***

In the context of this study, physical evidence comprises various objects obtained from the research setting, encompassing materials such as flyers, posters, and instructional resources. Specifically, for this investigation into ESP teachers' evaluative practices, the process of

evidence collection involved accessing and analyzing examination archives maintained by different departments, including the Departments of Computer Science, Mathematics, Physics, and Chemistry.

The inclusion criteria applied during the evidence selection process were systematically defined. These criteria considered factors such as academic audience, examination type, medium, and the academic year. The research focused on gathering assessment artifacts that pertained to final semester exams administered to first-year students with designations *Informatique* (I) and *Mathematiques Informatiques* (MI). This deliberate exclusion of other academic levels was motivated by the research's specific objectives, which centered on first-year Computer Science students. Importantly, this approach minimized potential bias in the analysis, given the substantial volume of artifacts and the differing evaluative practices that might be encountered across various audiences and academic levels.

Even in the context of the Covid-19 pandemic, when educational instruction transitioned to online delivery, it is noteworthy that final examinations continued to be conducted in traditional classroom settings. Consequently, all collected evidence for analysis remains in a paper-based format. The documents sourced for analysis span from the year 2010, providing a diverse selection of artifacts primarily authored by part-time instructors who had previously served at the faculty, along with those currently engaged in the research's timeframe. Table 2.1 enumerates the chosen assessment artifacts, adhering to the aforementioned inclusion criteria.

Table 2. 1 Selected Assessment Artifacts (see Appendix A)

<b>Audience</b>	<b>Function</b>	<b>Academic Year</b>
1 <sup>st</sup> year Computer Science (I & MI)	Final Exam (FE)	2010-2011
	Makeup Exam (ME)	2010-2011
	Final Exam (FE)	2011-2012
	Makeup Exam (ME)	2011-2012
	Final Exam (FE)	2012-2013
	Makeup Exam (ME)	2012-2013
	Final Exam (FE)	2013-2014
	Final Exam (FE)	2014-2015
	Final Exam (FE)	2017-2018
	Final Exam (FE)	2018-2019
	Final Exam (FE)	2019-2020
	Final Exam (FE)	2022-2023

In adherence to Morgan's (2022) inclusion criteria, it is crucial to emphasize the authenticity and unaltered nature of the selected documents. These documents are characterized by their reliability, as they originate from reputable sources and have not undergone any modifications. Their significance lies in their comprehensive representation of ESP teachers' evaluative practices, particularly concerning final examinations and summative assessment. Furthermore, these documents offer invaluable insights into the historical evolution of first-year I and MI ESP examinations, at the Faculty of Sciences, dating back to the year 2010.

As Bowen (2009) stated, "Document analysis involves skimming (superficial examination), reading (thorough examination), and interpretation. This iterative process combines elements of content analysis and thematic analysis. Content analysis is the process of organizing information into categories related to the central questions of the research." (p. 32). Table 2.2 shows the categories upon which the exploration of artifacts was ground. The categories were identified based on the review of the literature.

Table 2. 2Artifact Analysis Categories

Categories	Specifications
Test Planning	<ul style="list-style-type: none"> <li>- Context and purpose of the test</li> <li>- Overall test structure</li> <li>- Sections, task formats, and number of items</li> <li>- Individual task format specifications</li> <li>- Purpose and construct definition</li> <li>- Relevance to constructs being assessed</li> <li>- Scoring method and criteria</li> <li>- Sample items, prompts, and passages</li> <li>- Iterative process for specification development</li> <li>- Reverse engineering specifications when modifying existing tests</li> </ul>
Item-based Tasks	<ul style="list-style-type: none"> <li>- Item stem structure (for multiple-choice, true-false, short-answer)</li> <li>- Vocabulary and grammatical structures</li> <li>- Number of options (for matching tasks)</li> <li>- Requirements for short-answer questions</li> <li>- Polytomous or dichotomous scoring</li> <li>- Scoring criteria for linguistic accuracy</li> <li>- Sample questions, prompts, passages</li> </ul>
Grammar Tasks	<ul style="list-style-type: none"> <li>- Distinguishing recognition from usage</li> <li>- Assessing vocabulary and grammar range and proficiency</li> <li>- Scoring (polytomous form and meaning)</li> </ul>
Reading and Listening	<ul style="list-style-type: none"> <li>- Language for presenting questions</li> <li>- Language of response (for short-answer)</li> <li>- Comprehension question relevance</li> <li>- Short-answer linguistic accuracy</li> <li>- Previewing questions for listening tests</li> <li>- Number of listening repetitions</li> <li>- Passage alignment with question types</li> <li>- Vocabulary and syntactic complexity</li> <li>- Genre and rhetorical mode (for reading)</li> <li>- Speech act type (for listening)</li> <li>- Authenticity in listening passages</li> <li>- Passage scripting level</li> </ul>
Speaking and Writing	<ul style="list-style-type: none"> <li>- Effective prompt design</li> <li>- Desired response length, genre, and rhetorical mode</li> <li>- Vocabulary, grammar, and topic specifications</li> <li>- Interview format (fixed, unguided)</li> <li>- Balance in interlocutor support</li> <li>- Discussion task structure</li> <li>- Task format (writing or speaking) specifications</li> <li>- Prompt type, register, communicative functions</li> <li>- Vocabulary, grammatical patterns, topic specificity</li> <li>- Guidelines for scoring or providing support.</li> </ul>

***2.4.2.3 Summary***

In conclusion, document analysis emerges as a pivotal research instrument that holds great significance across educational research. This qualitative method allows researchers to systematically review and evaluate an array of documents, both in print and digital form, offering a rich source of primary data. Through this process, researchers can gain deeper insights into the experiences and perspectives of individuals directly involved in a particular subject.

As highlighted in our discussion, document analysis plays a crucial role in providing context, generating research hypotheses, suggesting observations, offering additional insights, and corroborating evidence from various research sources. It serves as a valuable tool for researchers to enhance the credibility of their findings through techniques like triangulation, which compares data from multiple sources. This approach not only minimizes bias but also allows for a comprehensive understanding of complex phenomena.

Our research specifically leverages document analysis in conjunction with other research methods to investigate the impact of summative assessments on the language skill development of first-year Computer Science students. By systematically analyzing twelve (12) assessment artifacts, we aim to uncover evaluative practices and correlations with language skill development. This study's significance lies in its potential to shed light on the intricate relationship between assessments and language skills, particularly within the Algerian Higher Education context.

Furthermore, the careful selection of evidence and adherence to inclusion criteria ensure the reliability and authenticity of the documents under examination. These documents, dating back to 2010, not only provide valuable insights into ESP teachers' evaluative practices but also offer a historical perspective on the evolution of first-year ESP examinations at the Faculty of Sciences.

**2.4.3 Classroom Observation**

Effective research instruments must be objective, reliable, and valid to rigorously test research hypotheses. Questionnaires are a common research instrument that emphasizes objectivity by advocating for researchers to remain detached from respondents during data collection. While questionnaires excel at providing valuable quantitative data, they often fail to provide a complete picture of the classroom environment.

Observation methods, on the other hand, are a distinct category of research instruments that allow researchers to collect data by unobtrusively watching, listening, and recording (Cohen et al., 2017a). Unlike surveys, observation allows for a more immersive exploration of the intricacies of teacher-student interactions, encompassing various aspects of the teaching-learning process, such as instruction, assessment, and feedback. This research tool, primarily valuable in qualitative research, offers insights into activities and behaviors that questionnaires may not fully capture.

In instances where written data collection procedures are inadequate, classroom observation emerges as a valuable resource, providing researchers with the qualitative data necessary to address research questions and substantiate hypotheses. The effectiveness of this tool is particularly evident when phenomena can be systematically observed. Classroom observation encompasses distinct types, each suited to different research contexts and objectives.

Participant and non-participant observation represent distinct approaches to data collection within educational research. Participant observation immerses the researcher within the classroom or group under scrutiny, allowing for active interaction and engagement with the subjects of study. While this approach provides an insider's perspective, it raises concerns about potential bias due to the observer's involvement. Conversely, non-participant observation adopts a passive stance, permitting the observer to discreetly witness behaviors, events, and interactions

in educational settings without intervention or influence. This method is esteemed for its objectivity, as it encourages participants to behave naturally and facilitates examination of observed phenomena.

Further distinctions emerge in controlled and uncontrolled observation. Controlled observation is commonly associated with scientific experiments, combining elements of participant observation by involving the observer in the observed activities. This approach enables the researcher to set and conduct experiments during observation. Conversely, uncontrolled observation aligns with non-participant observation, maintaining an observer's passive role. It is characterized by observation and recording, without interaction or experimentation. This approach preserves the natural context of the observed phenomena, refraining from any interference.

The dichotomy extends to structured and unstructured observation methodologies. Structured observation adheres to a systematic approach, demanding the researcher's non-participation while collecting data. It hinges on predetermined procedural tools for data collection, often yielding quantitative data. Nevertheless, it is also capable of producing reliable qualitative insights when guided by a predefined focus. Researchers must design record-keeping mechanisms, such as ethnographic narratives, transcriptions, or checklists, to facilitate data capture during structured observation. Unstructured observation, in contrast, closely resembles structured observation in terms of non-participation. Unstructured methods diverge from the specific and targeted nature of structured techniques, requiring keen observational skills to extract relevant data. While it might appear unfocused, unstructured observation can uncover unexpected findings, although it carries the risk of drifting away from the original research objectives.

Observational checklists are indispensable tools in structured observation. These forms guide observers in systematically recording the occurrences and frequencies of targeted elements, preventing reliance on memory alone. The design of observational checklists may vary based on research objectives and the observer's preferences. While they often generate quantitative data using rating scales and coding systems, these checklists can accommodate qualitative comments, enhancing the depth of data collection. Each approach has its own strengths and weaknesses, and the most appropriate approach will depend on the specific research questions being asked (Cohen et al., 2017a).

### ***2.4.3.1 Purpose***

The primary objective of incorporating classroom observation as a research instrument in this study is to investigate how ESP teachers' evaluative practices influence the development of language skills. To accomplish this, a mixed-methods research approach has been employed, aiming to enhance the credibility and comprehensiveness of the research findings. Classroom observation assumes a central role in this research, serving as a qualitative tool to uncover the impact of formative assessments on language skill development. By closely observing teachers in action, this method provides valuable insights into the strengths, weaknesses, and potential areas for improvement within the classroom context.

Through the process of observation, we aim to collect data by immersing ourselves in the nuances of classroom assessments, including the targeted language and topical KSAs, in addition to the feedback provided. This research instrument enables us to gain deeper insights into the activities and behaviors that may not be fully captured by surveys or other research methods. Our goal is to delve into how educators employ formative assessment tools and how these evaluative practices may either facilitate or impede the development of language skills, within the classroom setting.



Given our specific focus on 1st-year students at the Faculty of Sciences, our observation efforts are concentrated on 1st-year ESP classes. Through systematic observation, we seek to identify correlations between the development of learners' language skills and the evaluative procedures applied within these classrooms. Furthermore, our observation-based discussion will endeavor to address our initial research question: "How do ESP teachers at the Faculty of Sciences assess 1<sup>st</sup> Year Computer Science students?" By combining classroom observation with other research methods, our study aims to provide substantial answers and contribute valuable insights to the field of language assessment in the context of Algerian education. Classroom observation has been employed to directly gather information about formative assessments, facilitating discussions about their potential influence on the development of learners' language skills, within the Faculty of Sciences at Tlemcen University.

#### ***2.4.3.2 Procedure***

In this research, an undisguised, non-participatory, and structured approach to classroom observation was adopted. The choice of a non-participant role was deliberate, aiming to maintain a passive stance during the observation process. This approach was selected to minimize potential bias that might arise when observing classroom behaviors, events, and interactions. By assuming a non-participatory role, the researcher sought to enhance objectivity, encouraging participants to act naturally in their classroom settings. It's important to note that the intention was not to introduce experimental elements during the observations; therefore, uncontrolled classroom observation was deemed most appropriate, aligning with the researcher's passive observer stance.

Furthermore, the researcher opted for a structured observation methodology, following a predefined approach. This choice was made to ensure the collection of reliable qualitative insights, guided by research objectives. To facilitate data collection during the observation, a

checklist was developed, serving as a data-keeping tool. This structured approach helped maintain focus and prevented deviations from the original objectives, a potential concern with unstructured observation methods.

Prior to commencing the observation process, the researcher carefully selected a sample of participating teachers to be observed. The selection of these teachers was based on their availability, accessibility, and willingness to take part in the research. Informed consent was obtained from each teacher, granting permission for the observer to enter their classrooms, witness their activities, and record relevant data.

Each of the sampled teachers was observed on two separate occasions. The primary aim of these observations was to gather data related to teachers' evaluative practices in the context of assessing language skills in ESP classes. Consistency was maintained by using the same data-keeping forms during both observation sessions, which included an observational guide (see Appendix B) and note-taking. While the guide predominantly focused on teachers' classroom evaluative practices, note-taking served as a supplementary means to capture information relevant to the research problem that may not have been covered.

The classroom observations were conducted during two distinct periods: the second semester of the academic year 2018-2019 and the first semester of 2019-2020. This time frame was chosen to ensure a comprehensive examination of teachers' practices over different academic terms. The selection of both the first semester and the second semester for classroom observations was guided by the assumption that different teachers might emphasize specific language skills during these distinct periods of the academic year. It was hypothesized that some teachers may prioritize the development of particular language skills in the first semester, while others might focus on different skills in the second semester. By conducting observations in both semesters, the research aimed to capture potential variations in teachers' evaluative practices and

instructional emphasis, thus providing a more comprehensive understanding of language skill development across different timeframes within the academic year.

### **2.4.3.3 Observational Guide**

Observational guides and checklists are indispensable tools in structured observation to guide the observers in the systematic process of recording the nature, occurrence and frequency of the targeted element. As the design of observational checklists may vary based on research objectives and the observer's preferences, it may also produce both qualitative and quantitative data (Olajide, 2018). The design of the observational guide for the present research (see Appendix B) was based on the literature review, the situation analysis, the research objectives, and the research questions. The guide focuses on many aspects of formative assessment that have a direct impact on learners' language skill development (Table 2.3).

**Table 2.3** *Observation Guide*

<b>Aspect</b>	<b>Explanation</b>
Assessment Type	Whether various tools and strategies for assessment are employed (informal question, a practice quiz, a one-minute paper, a clearest/muddiest point exercise, or a K-W-L chart).
Input Language	Whether the formative assessments are delivered entirely in English or if they incorporate the use of other languages.
Targeted Language KSA	Whether it targets learners' linguistic knowledge, or their ability to use language in real-world contexts.

Targeted Topical KSA	Whether it targets learners' topical knowledge, or their ability to use domain-specific skills in real-world contexts
Individual Needs	Whether the teacher identifies and highlights the unique needs of each student during the class.
Timely Feedback	Whether the teacher provides immediate and constructive feedback to students during the class activities.
Integration into Lesson Plan	Whether formative assessment is intentionally embedded into the lesson plan and learning activities.
Emphasis on Progress	Evidence of the teacher concentrating on students' progress and growth rather than numerical scores for instance.
Student Goal Participation	Whether students are actively engaged in setting and discussing their learning goals related to the curriculum.
Teacher Inquiry	Whether the teacher uses formative assessment to answer questions about students' learning during the class.
Alignment with Goals	Whether formative assessment aligns with both teacher's objectives and students' goals for the learning process.
Teacher Expertise	Whether the teacher's expertise is evident in

	interpreting formative assessment data and adapting instruction accordingly.
Seamless Integration	Whether formative assessment fits into the classroom environment without disrupting the flow of the lesson.
Focus on Student Welfare	Whether the teacher prioritizes students' well-being and creates a supportive and caring learning atmosphere.
Next Steps Guidance	Whether formative assessment guides immediate next steps for students and influences curriculum decisions.
Enhanced Understanding	Whether formative assessment helps students and teachers gain a better understanding of the learning process in general and for individual students.
Student Responsibility	Whether students are encouraged to take responsibility for monitoring and supporting their own learning.
Multiple Data Sources	Whether various kinds and sources of information are employed.

The aspects included in the observation guide are intended to offer insights on how formative assessment is put into practice within the classroom and its influence on the growth of students' language skills. This will be accomplished by examining how classroom assessment methods align with established literature, shedding light on their potential effects on language skill development.

***2.4.3.4 Summary***

The use of classroom observation in this study serves the primary objective of delving into ESP teachers' evaluative practices and their impact on the development of language skills. To ensure the comprehensiveness and credibility of our research findings, we have adopted a mixed-methods research approach, strategically integrating classroom observation as a qualitative tool to collect evidence about the intricate influence of formative assessments on language skill development.

The role of classroom observation within our research framework is to provide an immersive and firsthand perspective on the dynamic of formative assessment. This approach facilitates a holistic understanding of the strengths, weaknesses, and areas for potential improvement within the classroom context.

Through the process of observation, our aim is to gather context-specific data by closely examining the qualities of formative assessment that may not be fully captured through surveys or other research methods. Classroom observation helps in building understanding of how educators employ formative assessment tools and whether these act as catalysts or hindrances to the development of language skills. Our research focuses on 1st-year Computer Science students at the Faculty of Sciences, and systematic observation is employed to establish correlations between the development of learners' language skills and the evaluative procedures enacted within their ESP classes.

Furthermore, our observation-based data is crucial to address our core research question: “What impact does ESP assessment have on 1st-year Computer Science students at the Faculty of Sciences?” The structured approach allows the researcher to maintain objectivity while capturing the authentic dynamics of the classroom assessment. Moreover, our selection of both

the first and second semesters for observation is rooted in the assumption that different teachers may emphasize distinct language skills during these academic periods.

The observational guide was designed to serve as an indispensable tool in our structured observation approach, systematically guiding the data collection process and facilitating the recording of various aspects related to formative assessments and their potential effects on language skill development. Each element in the checklist aligns with our research objectives and questions, allowing us to glean rich qualitative insights. By considering the diverse aspects included in the observation guide, the aim is to shed light on the practical implementation of formative assessments and their implications for language skill growth, bridging the gap between theory and classroom practice. Through the combination of classroom observation and other research methods, we endeavor to contribute significantly to the field of language assessment, ultimately enhancing the educational landscape at Tlemcen University.

#### ***2.4.4 Learners' Survey***

Surveys serve as invaluable tools in research, allowing for the systematic collection of data from large and diverse groups of individuals (Dörnyei, 2007). These surveys, deployable in various formats such as online, in-person, or over the phone, are powerful instruments for gathering information on a wide range of topics, including personal beliefs, behaviors, and experiences (Dillman et al., 2014). Two primary types of surveys exist: cross-sectional and longitudinal. Cross-sectional surveys capture data from a sample of individuals at a single point in time, while longitudinal surveys track responses from the same group over an extended period. These distinctions enable researchers to tailor their data collection methods to their research questions effectively (Fowler, 2013).

Surveys offer several advantages, including cost-effectiveness, the ability to reach dispersed populations, and the capacity to explore sensitive topics. Moreover, surveys can collect

both quantitative and, to some extent, qualitative data, providing researchers with comprehensive insights. However, they are not without limitations, as survey design can introduce biases, response rates may be low for lengthy surveys, and challenges may arise when surveying individuals with low literacy skills or investigating complex topics (Dörnyei, 2007; Fowler, 2013).

Effective survey design is essential to obtain reliable data. Researchers should begin by formulating clear research questions and selecting appropriate question types (open-ended or close-ended). Keeping surveys concise and understandable, using clear language, and pretesting on a small group before deploying them to the target population can help mitigate issues related to unclear questions or confusing instructions.

Surveys find applications across various research settings, from gauging public opinion to evaluating educational programs, assessing satisfaction, identifying the needs of specific population groups, and tracking changes in behavior or attitudes over time. Through meticulous design and adherence to best practices, researchers can maximize the potential of surveys to provide valuable insights in their respective fields.

In addition to these considerations, researchers should keep several additional points in mind when using surveys in research. First, ensure a sufficiently large sample size to enhance the representativeness of results. Second, carefully choose the sampling method, whether random, stratified, or cluster sampling, as it determines who is included in the sample. Lastly, strive for a high response rate, as this improves the representativeness of findings. Once data is collected, researchers should employ appropriate statistical software for data analysis to effectively address their research questions and contribute meaningful insights to their field of study (Dillman et al., 2014; Dörnyei, 2007; Fowler, 2013).



***2.4.4.1 Purpose***

The decision to employ a learners' survey in the present research is rooted in its numerous advantages and its alignment with the research objectives. Generally speaking, surveys allow for efficient data collection, including both quantitative and qualitative data. However, the present survey is used as a quantitative instrument. This decision is also driven by its intrinsic value in gathering quantitative insights from the learners' perspective. This survey is a vital component of our research, aligning with the objective of exploring the relationship between ESP teachers' evaluative practices and the development of language skills among learners.

By administering the survey (see Appendix C), we aim to systematically collect quantitative data regarding learners' perceptions of ESP final exams and their experienced impact on the growth of language and topical knowledge and skills, which help answering the second research question: "How do undergraduate Computer Science students experience ESP assessment?" Importantly, learners' viewpoints serve as a critical lens through which we can establish correlations between teachers' evaluative practices and the students' genuine appreciation of their own progress.

Furthermore, the survey produces valuable insights into the specific needs and expectations of learners concerning language assessment within their specific fields. Surveys can be used to explore sensitive topics, as they provide respondents with a degree of anonymity. In our research, we aim to understand students' perceptions of ESP evaluative practices, which may involve sensitive opinions and experiences.

To ensure the comprehensiveness and representativeness of our data, the survey encompasses students across various academic levels, specifically 1st, 2nd, and 3rd year LMD students at the Faculty of Sciences. This strategic inclusivity enhances the reliability and generalizability of our research findings, ultimately contributing to a more holistic understanding

of language assessment dynamics within the academic realm. Besides, learners' surveys, including a diversified audience, is ideal for tracking changes in learners' development over time, aligning with our research underlying objective of understanding how students' perceptions may evolve over their academic journey.

#### ***2.4.4.2 Description***

The learners' survey (see Appendix C) serves as a crucial instrument for obtaining quantitative data that complements the qualitative data generated through assessment artifacts analysis and classroom observation. Its design is aligned with the research questions, objectives, and the existing literature, ensuring that it effectively captures the insights we seek to gain. The survey can be categorized as a cross-sectional form collecting data at a single point in time. Moreover, considering the sequential nature of the present research, the design of the survey was informed by the qualitative data collected previously. Details on how the qualitative data from the earlier phase have shaped the design of the learners' survey will be presented in the next chapter.

The survey itself is structured to delve into learners' experiences with ESP assessment and how these experiences impact the development of their topical and language knowledge and skills. It comprises six (07) sections, each containing a set of items designed to collect specific types of data. The sections are organized to ensure a comprehensive exploration of learners' perspectives. These sections evolve around: (1) respondents' background information, (2) exam procedures (3) reading skill, (4) writing skill, (5) speaking skill, (6) listening skill, (7) and language skills in specific domain.

The use of close-ended scales in our survey design is strategic. These scales feature statements and questions that prompt learners to express their level of agreement or disagreement. This format is chosen for its ease of implementation across various modes of

communication, making it accessible to our diverse group of participants. Additionally, close-ended scales simplify the data analysis process, enabling the researcher to quantitatively measure a range of opinions, behaviors, and attitudes related to ESP assessment and language skills development. This approach ensures that we can capture subtle differences in learners' perspectives, enriching our understanding of their experiences.

The learners' survey is a systematically designed research tool that aligns with the objectives of our study. Its incorporation into our research methodology facilitates the systematic collection of quantitative data, allowing us to triangulate and enrich our findings with insights from multiple sources, ultimately contributing to a more comprehensive understanding of the complex interplay between ESP assessment and language skill development.

#### ***2.4.4.3 Administration***

The administration process of a survey in any research plays a pivotal role in shaping the quality and reliability of the data collected. It is not merely a logistical step but a crucial aspect that ensures the research aligns with its objectives and context (Dörnyei, 2007). In the specific context of our study on the impact of teachers' evaluative practices on language skills development among university students, the administration process carries particular significance.

This process enables us to gather valuable insights from a diverse group of learners, shedding light on their perspectives and experiences within the academic environment. To provide a comprehensive understanding of this critical phase, we will explore the administration procedures within the specific context of our research, underscoring the considerations and choices made to ensure the survey's effectiveness and relevance.

The survey was created using Google Forms, an online survey administration tool available as part of the free Google Docs Editors suite. This software offered several advantages for both the survey designer and the respondents. Its user-friendly interface ensured a flawless experience for both parties. Furthermore, Google Forms allowed for an unlimited number of submissions, making it feasible to include a large number of students in the survey. Importantly, this software is accessible to all types of users at no cost.

To administer the survey, we targeted 1st, 2nd, and 3rd-year Computer Science students at the Faculty of Sciences, as these cohorts were most likely to provide insights into the impact of teachers' evaluative practices on language skills development. The survey link was distributed via email, with students' electronic addresses collected through the institution's administrative channels.

Our administration process took contextual factors into account; specifically, it was conducted during the second (2nd) semester of the academic year 2022-2023. This timing allowed 1st-year students to have experienced their first (1st) semester final exams, which was critical to our research focus. Lastly, the survey was available online for a duration of three weeks, ensuring that students had sufficient time to participate while also considering their commitments related to preparing for makeup exams.

#### ***2.4.4.4 Summary***

The learners' survey employed in this research serves as a fundamental tool for collecting quantitative data, complementing the qualitative insights obtained through assessment artifacts analysis and classroom observation. This survey was systematically designed to align with our research objectives, questions, and the existing literature, making it an integral component of our investigation into the relationship between ESP teachers' evaluative practices and language skills development among learners at the Faculty of Sciences. By administering this survey,

quantitative data on students' perceptions of ESP final exams and their perceived impact on language and topical knowledge and skills development were collected, addressing our second research question effectively.

Furthermore, the survey facilitated a deeper understanding of students' needs and expectations regarding language assessment within their specific academic fields. Its inclusion of students across various academic levels ensured comprehensive and representative data, enhancing the reliability and generalizability of our findings. The survey's strategic use of close-ended scales enabled participants to express their opinions and attitudes conveniently while simplifying the subsequent data analysis process.

Ultimately, the learners' survey stands as a well-structured research tool that enriches the present study by providing insights from multiple sources, contributing to a comprehensive comprehension of the interplay between ESP assessment and language skill development. The administration process, carried out using Google Forms, was carefully planned to ensure accessibility, inclusivity, and relevance within the context of our research. Through this planned process, a robust dataset that will aid in addressing our research questions effectively was collected.

#### ***2.4.5 The Test***

Tests are powerful tools for data collection in educational research. They can be used to measure a wide range of variables, including knowledge, skills, attitudes, and behaviors. Tests can be used in a variety of research designs, including experimental studies, quasi-experimental studies, explanatory, and exploratory studies (Cohen et al., 2017b). As Olajide (2018) explained,

a test is an instrument designed, produced and implemented to elicit information about an individual respondent in respect of his knowledge, attitude, skills, assumed values and preference. In the humanities (which include the arts, education, social sciences, and law - at times the social sciences are made to stand

alone), test is used to evaluate teaching and learning, and obtain data for research. (p. 403)

The process of designing a test for research purposes begins with a clear understanding of the study's objectives and research questions. It is essential to define the purpose of the test: What specific aspect or variable is a researcher aiming to measure? The choice of test type is another critical decision. Researchers must select the test type that aligns best with their research question, whether it is a knowledge-based test, a performance-based assessment, or another suitable format for the research.

The construction of the test involves determining how it will be designed, including the format and structure. Researchers must decide on the test's format, whether it will be multiple-choice, essay-based, or another appropriate style. The construction process also includes crafting test items and ensuring they are aligned with the research objectives and strategy.

Setting clear objectives for the test is equally vital. These objectives should outline the specific goals the test aims to achieve within the research context. Moreover, careful consideration must be given to the content the test should cover. The content should directly relate to the research question and objectives, ensuring that the test provides meaningful data.

Tests can serve various purposes in educational research. One common use is to collect data at different points in time to track student progress over the course of a study. For example, researchers may administer a pretest at the study's outset and a posttest at the study's conclusion.

Tests are also valuable for comparing the performance of different groups of students. Researchers may use tests to evaluate the effectiveness of a new instructional program, or assessment approach, by comparing the performance of participants who received the new approach with those who received the traditional one (Cohen et al., 2017b).

Tests are essential tools for collecting data in educational research. When designing a research test, it is critical to consider factors such as the test's purpose, type, objectives, content, construction, and format. Usefulness, particularly validity and reliability, must be ensured to guarantee the quality of the collected data. Tests can be used to track students' progress and compare the performance of different groups. Researchers should also pay attention to sample size, test bias, consistent administration, and thoughtful data interpretation to obtain meaningful insights from their research.

#### ***2.4.5.1 Purpose***

The primary purpose of incorporating a test as a research instrument in this study is to measure the effectiveness and impact of two distinct assessment approaches by comparing the performance of separate student groups. To accomplish this, it is imperative to establish precise and well-defined objectives for the test, outlining its specific goals within the research context.

This test is designed to generate quantitative data essential for addressing the core research question: "What impact does ESP assessment have on 1st-year students at the Faculty of Sciences?" It serves as a crucial tool for exploring the influence of evaluative practices on the development, mobilization, and manifestation of language skills within domain-specific contexts, with a particular emphasis on 1st-year Computer Sciences (I and MI) students.

The test is employed to collect and compare data representing students' performance in two fundamentally distinct assessment approaches. This dataset, in combination with data obtained through other research methods, will serve as the foundation for discussing the correlation between evaluative practices and learners' language development within the ESP context, which constitutes our central and concluding research question.

Furthermore, this research instrument serves the purpose of introducing and evaluating a novel assessment approach known as Learning-oriented Assessment (Turner & Purpura, 2016) while providing a blueprint and template, grounded in assessment standards. By doing so, this research also seeks to assess the effectiveness of this modern approach within the Algerian Higher Educational context, aiming to contribute valuable insights for educators and researchers alike.

#### **2.4.5.2 ICL Test Design**

To create the ICL 1st-year Computer Science Test, based on the LOA framework, a collaborative effort involving subject matter instructors and English language instructors was essential to bridge the gap between theory and practice. The initial step in the design process was to establish the contextual dimension, as it serves as the foundation for all other dimensions. This particular test was tailored for first-year Computer Science undergraduate students at Tlemcen University, within an ESP class. The test centered on requirements elicitation, a fundamental component of software development.

In their paper, entitled “*ICLHE Task Design: Case of LI Computer Science, Tlemcen University*,” Boubris and Bouabdallah (2023) provide crucial information about the design of the present test. Within this design framework, the target language use domain was identified as professional, aligning with the task of writing a coherent and accurate series of questions (an interview) to extract information from an end-user (a client). A form-function analysis played a crucial role in identifying the linguistic resources required for this task, enabling the creation of assistance and a scoring rubric (see Appendix D).

The test itself consisted of four tasks, emphasizing integrated skills development while acknowledging the significance of socio-cognition and social interaction in the co-construction of knowledge. The final task, representing the culmination of the activity, required students to



utilize various forms of input and assistance. During this phase, learners engaged in a limited-production task, employing the provided resources to formulate their requirements interview questions.

The proficiency dimension of the test focused on students' ability to craft questions using interrogative pronouns, make requests using auxiliaries (morphosyntactic forms), and interpret meanings related to the end-user's needs. Additionally, the test assessed students' competence in employing greetings, expressing gratitude (interactional forms), and employing specialized vocabulary for precise communication. The aspect of formality and politeness was also taken into account. Proficiency in this context necessitated the ability to infer the client's needs accurately from the provided input, demonstrating topical control and linguistic accuracy.

Although the test did not incorporate social-interactional elements, it did consider human-computer interaction, recognizing its potential impact on the test-taker experience. In terms of technology, the activity required basic computer literacy due to its online settings. The test offered four types of assistance to learners: input as instruction (text and interview sample), instructional assistance (algorithm), embedded explicit instruction (operational definitions and terminology), and feedback as instruction (pre-programmed key answers).

The socio-cognitive dimension of the test relied on higher-order thinking processes, necessitating learners to analyze the interview sample, algorithm, and sample answers provided. This analytical process enabled them to evaluate both topical and linguistic resources while creating propositional and functional meanings. The culminating task demanded increased reasoning and concentration compared to the enabling tasks.

Finally, the psychological dispositions of test-takers were considered, recognizing that confidence, motivation, confusion, and engagement levels may vary based on language

proficiency and psychological readiness. Expected behavioral dispositions included persistence and tolerance for ambiguity, while social dispositions were not a primary focus of the test.

The test is characterized by its objectivity, incorporating Multiple-Choice Questions (MCQs) to evaluate semantico-grammatical proficiency, topical knowledge, and reading comprehension. To minimize subjectivity in evaluation, the culminating task is designed as a limited production task, with a predetermined scoring rubric (see Appendix D). It adheres to a criterion-referenced approach, emphasizing performance assessment and encompassing both direct and indirect tasks. This comprehensive design enables the test to serve a dual purpose by providing both summative and formative information. Table 2.4 provides more details about each of the dimensions of the LOA framework when applied to the ICL Test.

Table 2. 4LOA Dimensions of the ICL Test (Boubris & Bouabdallah, 2023, p 33-34)

<b>Performance moderators</b>	
Proficiency dimension	<p><b>SL/FL KSAs:</b></p> <p>To use interrogative pronouns, auxiliaries (make requests), simple tenses, and topical content to accurately form questions.</p> <p>To use interactional forms to increase coherence and politeness</p> <p>Direct functional meanings include direct questions with targeted needs.</p> <p>Implied functional meanings include inferences based on client’s answers.</p> <p><b>Topical KSAs:</b></p> <p>To develop and display a full understanding of the concept of user’s needs.</p> <p>To understand the purpose of a given algorithm.</p> <p>To read an algorithm.</p> <p>To inquire about general and specific user’s needs.</p>

Elicitation dimension	<p><b>Task:</b>limited production (dialogue completion)</p> <p><b>Scoring:</b>scored</p> <p><b>Timed:</b>90 minutes</p> <p><b>TLU:</b>professional</p> <p><b>Input:</b>text / interview / algorithm</p> <p><b>Instruction (culminating task):</b> Complete the dialogue between the developer and the client, focusing on asking relevant questions and actively listening to the client's responses.</p> <p><b>Prompt:</b> assume the role of the developer.</p> <p><b>Expected response:</b> based on the given input students are expected to write meaningful, coherent, correct, and targeted series of questions (an interview) to evaluate the needs of an end-user (a client) based on his feedback.</p> <p><b>Process:</b>students use the assistance to evaluate previously activated topical and linguistic resources then form questions.</p>
<b>Performance indicators</b>	
Contextual dimension	<p><b>Disciplinary domain / Course:</b> Computer science / Technical English1</p> <p><b>Topic / Theme:</b> Software development / Requirements elicitation</p> <p><b>Audience / Institution:</b> first year undergraduate students in an ESP class, Tlemcen University</p> <p><b>Language use domain:</b> Professional</p> <p><b>Setting:</b>online</p> <p><b>Purpose:</b> curriculum-based achievement test.</p> <p><b>Enabling skills:</b> to activate learners' schemata and reinforce topical and linguistic understandings through assistance</p> <p><b>Envisioned language resources:</b> lexical, morphosyntactic, and interactional forms in addition to disciplinary resources</p>

	<b>Culminating competency:</b> the ability to ask relevant questions and actively listen to the client's responses.
Instructional dimension	<b>Input:</b> text and interview sample <b>Assistance as instructional:</b> algorithm. <b>Embedded explicit instruction:</b> operational definitions <b>Feedback:</b> pre-programmed key answers
Affective dimension	<b>Positive psychological dispositions:</b> confidence and motivation <b>Negative Positive psychological dispositions:</b> boredom and disengagement <b>Positive behavioral dispositions:</b> persistence, effort and tolerance for ambiguity <b>Negative behavioral dispositions:</b> lack of initiative <b>Social dispositions:</b> NA
Social-cognition dimension	Higher-order thinking processes: analysis, application, and creation Involves STM and LTM, information processing, and reasoning.
Technological dimension	Basic computer literacy. Requires an email and internet connection
Social-interactional dimension	Human-computer interaction.

The task development process was guided by a combination of approaches, incorporating a construct-based perspective, notably the Meaning-oriented Model (MOM) of L2 proficiency, and a task-centered approach characterized by complexity and integration. For detailed specifications about the ICL test, please refer to the blueprint below (Table 2.5). This blueprint differs slightly from the original one, presented by Boubris and Bouabdallah (2023, p. 35-36), to fit the context of the context of research and align with research questions.

**Table 2. 5ICL Test Blueprint**

Test component	Task types	Time	Length	Scoring
<b>Reading ability (enabling task)</b>				
-Theme: requirements elicitation  -Proficiency: negotiating meaning from a text	-Input: adapted text  -Task type: selected-response MCQ  -Task title: maximizing client's Engagement	-Input: 10 mins  -Items: 10 mins  -Feedback: 5 mins	5 items: propositional and functional meanings	Right/Wrong Scoring  0/1  5 pts.
<b>S-G Knowledge (enabling task)</b>				
-Theme: requirements elicitation  -Proficiency: Morphosyntactic forms: simple tenses, auxiliaries, interrogative pronouns  Interactional forms: greeting, expressing gratitude.	- Input: requirements elicitation interview transcript  -Task type: selected-response MCQ  -Task title: asking meaningful questions	-Input & Task: 15 mins  -Feedback: 5 mins	10 items	Right/Wrong Scoring  0/1  10 pts.
<b>Topical knowledge through Reading (enabling task)</b>				
-Theme: requirements elicitation  -Proficiency: negotiating meaning from an algorithm	-Input: discount scheme algorithm  -Task type: selected-response MCQ  -Task title: coding client's needs	-Input: 7 mins  -Task: 8 mins  -Feedback: 5 mins	5 items	Right/Wrong Scoring 0/1 5 pts.
<b>ICL writing ability (culminating task)</b>				
-Theme: requirements elicitation	- Input: requirements	- Input: 20 mins	-Interview transcript	-Analytic scoring rubric(see

<p>- Proficiency: writing a requirements elicitation questionnaire</p>	<p>elicitation interview transcript</p> <p>- Task type: limited production (dialogue completion)</p> <p>- Task title: collecting relevant information</p>	<p>- Task: 30 mins</p> <p>-Feedback: 5 mins</p>	<p>including 11 items (questions).</p>	<p>Appendix D)</p> <p>-3 criteria for correctness (topical control, language accuracy, rhetorical control)</p> <p>- 0 to 5 pts for each criterion (max. 15/rater)</p> <p>-1 rater (T1 + T2)</p>
--	---	---	--	---

The test's primary objective is to be highly practical and effective. To ensure its utility, the test design meticulously considers critical factors such as construct validity, as per the Model of Measurement (MOM), as well as reliability, both internal and external. Additionally, the design accounts for its impact on learning, with a focus on fostering positive washback effects. This positive washback is anticipated because the test adheres to the LOA framework, which guides performance moderators. The test's practicality is further enhanced by its online delivery format (static), making it accessible and convenient for implementation. Google Forms, equipped with pre-programmed feedback, including key answers and task scores, facilitates the online administration of the test.

The test serves the purpose of assessing learners' proficiency in crafting a fundamental requirements interview based on their inferences drawn from authentic input. This assessment necessitates students to deploy their questioning abilities and actively engage with a client to extract information regarding their needs and requirements. Proficiency in this task is demonstrated through effective utilization of open-ended questions, active listening to the client's responses, and seeking clarification, when necessary, all performed in a professional manner. Achieving this requires students to mobilize both their linguistic and non-linguistic knowledge

and skills effectively. To support learners in this endeavor, the test offers assistance through the provision of input and feedback.

### ***2.4.5.3 Description***

In our case study, the development and administration of the tests played a pivotal role in minimizing bias and extracting meaningful insights. To ensure the integrity of our experiment, we carefully designed two distinct tests that were administered to 1st-year Computer Science (I) students. The design of these tests was rooted in the preceding qualitative data collected during the initial phase of our research, aligning with the principles of sequential mixed-methods design (Almeida, 2018). An explanation of how the qualitative data influenced the test design will be presented in the next chapter.

These two tests differed in their approach to assessing language skills, with the first test, or *Generic Test*, (see appendix E) adhering to a conventional methodology, grounded in the findings derived from our assessment artifacts analysis, and designed to mirror the characteristics of the 1<sup>st</sup> Year CS English Exam. In contrast, the second test, or the *ICL Test*, (see appendix F) adopted a Learning-oriented approach based on the principles outlined by Turner and Purpura (2016). Each test comprised a set of tasks, carefully designed to extract relevant data while minimizing biases.

Significant attention was devoted to ensuring that the test content was suitable to our research questions, objectives, and the characteristics of the test-takers within their specific context. It's important to note that, given the primary research aim of establishing direct correlations between assessment and its impact, the tests were not curriculum-based i.e., the content of these tests deliberately deviated from the instructional materials provided to the students. This strategic decision was made to isolate the assessment component from instruction,

thereby obtaining less biased data. The focus was primarily on evaluating students' prior and test-related (input) language and topical KSAs.

Furthermore, to prevent potential bias and maintain fairness in the testing process, the topics and culminating tasks featured in the tests were deliberately omitted from classroom instruction. These culminating tasks, marked by their complexity and real-world relevance, encompassed both language KSAs, learners are assumed to have developed in High School, and topical KSAs they are assumed to have developed through their academic year at university. However, to support the integrity of the research, students were not informed of their scores, and these scores exclusively served the research purpose.

In order to encourage an enthusiastic participation in the tests, all students were awarded an additional two points in their final exam grades. To address concerns related to students merely taking the test without real engagement, specific conditions were established. Students were explicitly informed that they needed to complete all tasks and achieve at least the average score to qualify for the reward. From an ethical standpoint, all students were rewarded as promised, further ensuring a balanced and equitable research environment.

#### ***2.4.5.4 Administration***

The administration of these two tests in our experiment was carefully carried out to ensure the data's reliability and validity; several key considerations were taken into account throughout the process. Formatting the sample and organizing the groups was the first priority. To address this concern, an invitation was addressed to all 1st-year Computer Science (I) students to participate in the experiment. The autonomy of those who chose not to participate was respected. Lists of students were created in advance, and potential absences were anticipated. The participating students were randomly divided into two groups, each consisting of 38 students. For those students who were not initially on the lists but expressed their desire to



participate either just before or shortly after the test began, a random assignment process can be employed to allocate them to one of the two groups.

To prevent potential biases, ensuring that students had no prior knowledge of the test content or the grouping was necessary. Their only awareness pertained to the experimental nature of the event and the conditions required to earn the reward. The tests were administered on-site, a deliberate choice aimed at preventing any issues related to internet use during the exam, thus bolstering the authenticity and reliability of the results. The present test is mainly different from the one presented in Boubris and Bouabdallah (2023) in its administration procedures. From an LOA perspective, this directly affects the technological, social-interactive, and instructional dimensions. Learners do not have to employ computer literacy to interact with the computer and take the test, as opposed to an online administration. Furthermore, due to the inability to pre-program feedback as with online tools, key answers were furnished to students once the allotted time for each task had elapsed.

To ensure an effective testing process, an assistant/ invigilator was present in each of the four classrooms where the tests were conducted, alongside the researcher. Students were allocated 90 minutes to complete the Generic Test (the standard duration of any Algerian Higher Education Final Exam) and 120 minutes to complete the ICL Test, based on the blueprint (Table 9). Additionally, students were provided with the option to leave the exam rooms after an hour, depending on them meeting the first condition: completing all four tasks.

It's worth noting that the experiment took place at the conclusion of the second semester of the 2022-2023 academic year, a strategic timing choice that allowed students to acquire the necessary prior topical knowledge, which was crucial for their performance on the second test. This ensured that students were adequately prepared and equipped to participate in the experiment, enhancing the validity of our data.

***2.4.5.5 Summary***

In this comprehensive examination of the use of tests as research instruments, several critical aspects emerge that highlight their pivotal role in educational research. The primary purpose of incorporating tests into this research study is to assess the effectiveness and impact of two distinct assessment approaches by comparing the performance of different groups of students. The objectives are clearly defined to measure the influence of evaluative practices on language skill development, with a specific focus on 1st-year Computer Sciences (I) students. Additionally, the research seeks to introduce and evaluate a novel assessment approach called Learning-oriented Assessment, contributing to educational insights.

The ICL test design process is systematically outlined, emphasizing the collaborative effort between subject matter instructors and English language instructors. The test is grounded in the contextual dimension, tailored to the specific student group and aligned with professional language use. It comprises four tasks that assess integrated skills development, proficiency in language usage, and socio-cognitive dimensions. The test is designed to be objective, incorporating multiple-choice questions and a criterion-referenced approach. It takes the form of an achievement test based on content and language Integration and the LOA framework.

The development and administration of the tests are crucial in minimizing bias and extracting meaningful insights. Two distinct tests were carefully designed based on prior qualitative data. These tests differ in their assessment approaches and content, strategically isolating assessment from instruction to obtain less biased data. The tests are administered on-site, and specific conditions are established to encourage student engagement. Ethical considerations are paramount, and all students were eventually rewarded.

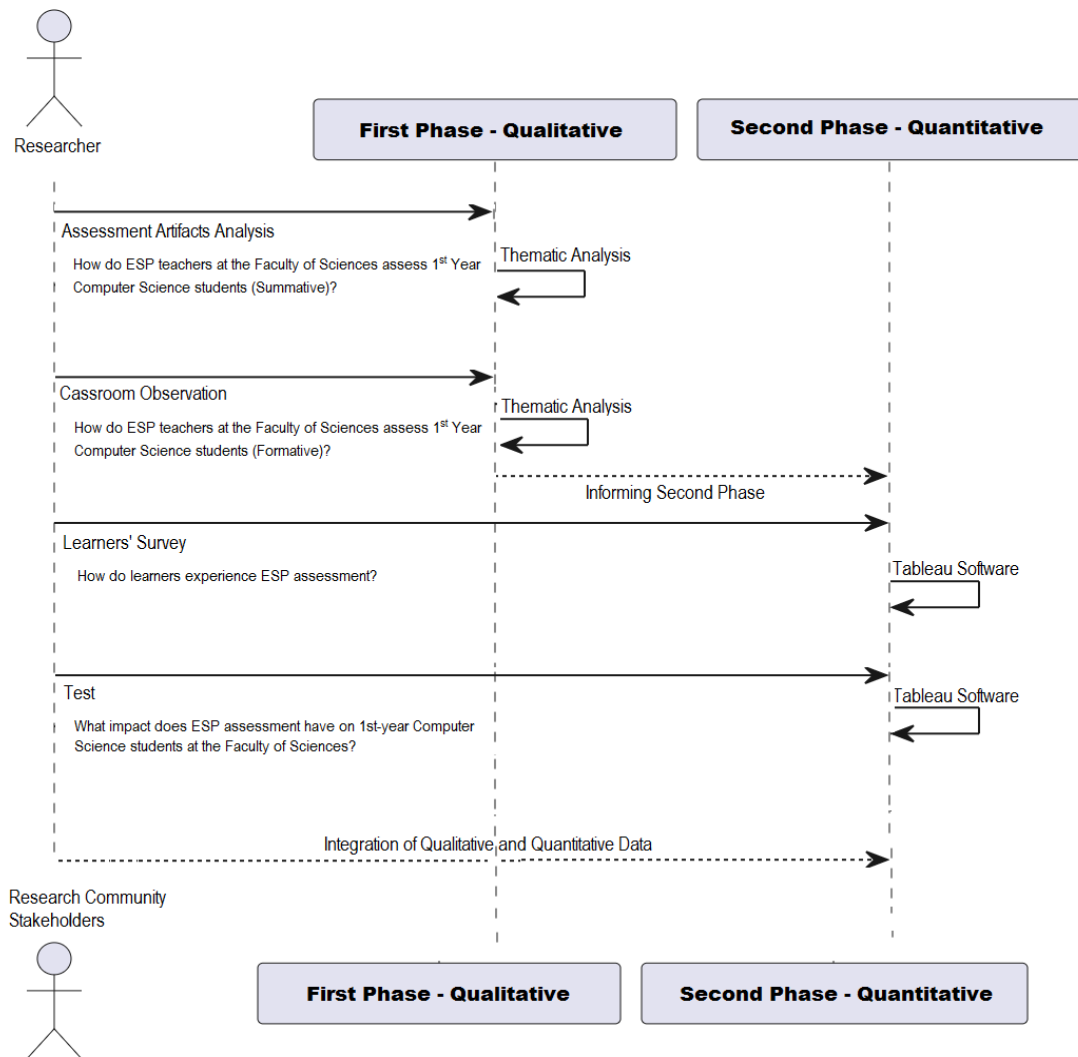
The administration process is characterized by careful sample formatting and group organization. All 1st-year Computer Science (I) students were invited to participate, with random

group assignment to reduce biases. Test content and grouping information are kept from students, and the tests are administered on-site to ensure data authenticity. An assistant/invigilator supervises the process, and students are given a standard amount of time to complete the exam. The strategic timing of the experiment enhances data validity.

#### **2.4.6 Conclusion**

This case study is designed to explore the impact of teachers' evaluative practices on the development of learners' language skills. To address the core research question regarding the nature of these evaluative practices and their potential impact, a combination of research instruments was used. This involved the examination of final exam artifacts and conducting classroom observations to gain critical insights. Building upon these qualitative research instruments, learners' survey was developed to delve into specific facets of language skills development and their correlation with evaluative practices, forming the basis of the second research question. Furthermore, the third research question, aimed at enhancing ESP assessment methods, was tackled through the administration of a test, designed based on insights drawn from the analysis of artifacts and classroom observations. This research adopted a sequential exploratory case study approach, where qualitative research instruments inform the development and implementation of quantitative ones, ensuring a holistic investigation into the research questions at hand. Figure 2.1, below, provides a comprehensive view about the research design.

Figure 2. 1Mixed-methods Design



### 2.5 Sampling Procedures

Sampling is a fundamental aspect of research, serving as the initial gateway to gather data and derive meaningful insights. It sets the stage for the entire research process, influencing the reliability and generalizability of the findings. Researchers must make critical decisions when selecting and implementing a sampling strategy (Haque, 2010). At its essence, sampling is the art and science of selecting a subset of individuals or elements from a larger population. This selected subset, known as the sample, serves as a microcosm that researchers scrutinize in pursuit of broader conclusions applicable to the entire population from whence it was derived.

In essence, it entails extracting a subset of individuals, termed a sample, for the purpose of gathering information. The data collected from this sample can then be extrapolated to draw broader conclusions that encompass the original population from which the sample was drawn. To achieve meaningful results, it is crucial to opt for a representative sample that mirrors the characteristics of the source population (Acharya et al., 2013).

Acharya et al. (2013) characterize the sample as a miniature reflection of the population, upon which the validity and generalizability of research findings are founded. Sampling, as a process, comprises a series of methodical steps aimed at selecting research units from the source population. This source population consists of individuals from which the study population, i.e., the sample, is derived. Both the study population and source population are subsets of the target population, the group to which the researcher intends to extend the research findings, based on the principle of representativeness.

In the realm of sampling methods, a pivotal distinction arises: probability sampling versus non-probability sampling. Probability sampling, exemplified by techniques such as simple random sampling, systematic random sampling, and stratified random sampling, offers a rigorous and objective approach. In probability sampling, every member of the source population has an equal chance of inclusion, ensuring a truly representative sample (Berndt, 2020). In contrast, non-probability sampling methods, including judgment sampling and convenience sampling, rely on the researcher's judgment or convenience, raising questions about subjectivity and diminishing sample representativeness.

In summary, sampling is a cornerstone of research methodology, allowing researchers to draw meaningful conclusions from a subset of a population. The choice between probability and non-probability sampling methods carries significant implications for the quality and generalizability of research findings. By meticulously defining target and source populations,

constructing a sampling frame, selecting appropriate sampling methods, and determining sample size, researchers can navigate the complexities of sampling and enhance the validity of their research endeavors.

### ***2.5.1 Defining the Target Population***

It is essential to identify and establish the target population, which encompasses the entire group that a researcher intends to analyze. Subsequently, a source population and a sampling frame are derived. This case study aims to explore the influence of assessment on learners' language skills development, within the Faculty of Sciences at Tlemcen University. Therefore, it is appropriate to define the entire group of ESP teachers within this context as the target population.

### ***2.5.2 Defining the Source Population***

The source population is a subset of the broader target population, specifically denoting the initial population from which the researcher draws a sample. This ensures that every member of this population has an equitable chance of being selected for participation in the research study. In this particular research context, the investigator has chosen to define the source population as encompassing 1st-year ESP teachers in Computer Science (I and MI), Science and Technology (ST), and Matter Sciences (SM). These teachers have served as the primary source for selecting samples to undergo observation and analysis of assessment artifacts.

In parallel, a survey was administered to students in the 1st, 2nd, and 3rd years at the Computer Science Department, Faculty of Sciences. This comprehensive approach allows the researcher to generalize findings from the delimited source population to the entire target population. Due to a number of limitations, such as human resources, time constraints, and

financial resources the final research instrument (the test) involved 1<sup>st</sup> year Computer Science (I) only.

### 2.5.3 The Sampling Frame

The sampling frame, a crucial component in the research methodology, serves as the comprehensive list from which potential individuals are drawn to constitute the research sample. As Table 2.6 shows, the source population consists of a diverse group of 1<sup>st</sup> year ESP teachers, within the Faculty of Sciences. To ensure privacy and confidentiality, this list of potential participants will remain anonymous and randomly coded throughout the study with the teachers being referred to as:  $T(n^{\circ})$ . This selection process and categorization of individuals within the source population lay the foundation for the subsequent sampling and data collection phases of the research.

**Table 2.6** Observation Sample Frame

Level	Teacher(s)	Position	Education
1 <sup>st</sup> year I and MI	01	full-time	PhD student
1 <sup>st</sup> year ST	06	part-time	Master 2
1 <sup>st</sup> year SM	02	part-time	Master 2

### 2.5.4 Sampling Method

To establish the sample of teachers to be observed in this research, an initial consideration was given to employing a probabilistic approach, which is generally regarded as the optimal method for attaining an objectively selected sample representative of both the source and target populations. This approach would have provided the highest level of objectivity and reliability in extrapolating research data. However, practical constraints and difficulties were encountered during the planning phase, primarily stemming from issues related to convenience.

In essence, the probability sampling approach, although theoretically desirable, became unfeasible due to these encountered complications. Consequently, an alternative sampling method was adopted—one that, while not entirely ideal, would still preserve a reasonable degree of representativeness in the sample and overcome the practical hindrances. This selected method is known as convenience sampling, a non-probabilistic technique where the sample is chosen based on accessibility and availability. While this approach may introduce some degree of subjectivity and potentially limit the generalizability of the findings, it was a pragmatic choice given the specific circumstances, ensuring that the research could proceed effectively despite the encountered inconveniences.

### ***2.5.5 Sample Size***

The convenience sampling method offers a straightforward and practical approach to sample selection. In this case, the source population consists of seven (10) ESP 1st year teachers, although not all of them were readily accessible or available for observation. Following an assessment of the teachers' availability and accessibility, only three (03) teachers remained eligible to constitute the sample, representing approximately 33% of the source population.

To maintain a reasonable level of representativeness within the sample and ensure that the findings remain meaningful, all three (03) of the available teachers were included in the study population designated for observation. These three (03) participants, who were willing and able to participate, are referred to as: T1, T2, and T3. Furthermore, it is anticipated that the learners' survey will involve a substantial participation of about four-hundred (400) respondents, while the test will be taken by approximately one-hundred (100) participants.



### **2.5.6 Conclusion**

Sampling, as a practice, involves the process of selecting a subset (sample) from a larger population, with this sample serving as a microcosm through which researchers aim to draw broader conclusions about the entire population. The choice between probability and non-probability sampling methods carries significant implications for research outcomes. Defining the target population is a crucial step, as it encompasses the entire group under scrutiny, while the source population represents a subset from which the sample is drawn. In this study, ESP teachers within the Faculty of Sciences at Tlemcen University constitute the target population, with 1st-year ESP teachers in various disciplines forming the source population. The sampling frame, a comprehensive list of potential participants, is a key component, ensuring privacy and confidentiality while facilitating the subsequent sampling and data collection phases.

Considering practical constraints, the study was based on convenience sampling, where the sample is primarily chosen based on accessibility and availability. While this method introduces some subjectivity, it remains a pragmatic choice, enabling the research to proceed despite encountered inconveniences. The sample size for observed teachers is included four (04) participants. Moreover, the learners' survey anticipates a substantial participation of over four-hundred (400) respondents, while the test is expected to involve approximately one-hundred (100) participants.

## **2.6 Data Analysis**

### **2.6.1 Qualitative Data Analysis**

The analysis of qualitative data is a critical component, primarily focusing on understanding abstractions and exploring non-numerical aspects of the study. Qualitative data analysis is inherently different from its quantitative counterpart, relying on an interpretive and constructive philosophy. This approach is particularly valuable when dealing with research

domains where quantitative aspects may not adequately capture the depth and complexity of the phenomena under investigation.

The research employs two qualitative research instruments: assessment artifact analysis and classroom observations. Both of these instruments yield rich qualitative data, which are subject to analysis. When analyzing data derived from assessment artifacts, a rubric serves as a valuable tool for grading or scoring documents. However, the analysis process goes beyond assigning scores. The analysis of documents is akin to the methodology applied to focus group transcripts or interview transcripts, known as content analysis and thematic analysis (Bowen, 2009).

Coding is an initial step in organizing the information within documents. This process involves categorizing data into meaningful clusters or categories that are directly related to the central research questions. This categorization aids in structuring and understanding the content within the documents. Building on coding, thematic analysis is a versatile method for further dissecting the data. Thematic analysis is especially suited to uncovering the nuances within the qualitative data. It enables researchers to identify recurring patterns and themes that may not be immediately apparent through traditional content analysis. As Morgan (2022, p. 73) reported: “After a researcher has selected a sample of texts, the next step in documentary research is to conduct an analysis of those texts. Because of its versatility, thematic analysis is an ideal method for this process.”

The coding process is an essential component of thematic analysis. Coding involves assigning meaningful labels or codes to segments of the data, highlighting key elements, phrases, or concepts within the documents. This process serves to structure and categorize the data in a systematic manner.

Qualitative data analysis extends beyond the mechanical aspects of coding and categorization. It involves interpretation, where the researcher applies their own assumptions, commitments, and scholarly knowledge to the data (Morgan, 2022). This interpretive approach allows for a deeper understanding of the underlying meanings and implications within the data. One of the strengths of qualitative data analysis, especially thematic analysis, is its ability to uncover unexpected meanings and patterns. Rather than merely summarizing the data, this approach seeks to reveal emergent themes that reflect shared patterns of meaning among the data.

Similar to the analysis of assessment artifacts, the qualitative data obtained from classroom observations also undergoes a rigorous analysis process. Researchers use coding techniques to identify meaningful segments within the observations. The data are then subjected to a thematic approach, where patterns and themes in the classroom behaviors and interactions are identified and analyzed. This approach is grounded in interpretive and constructive philosophies, making it particularly suitable for the purpose of the present study.

Through thematic analysis and coding processes, the qualitative data are transformed into structured insights that contribute to a more comprehensive understanding of the impact assessment may have on the development of learners' language skills. This holistic approach to qualitative data analysis enables researchers to not only identify patterns but also to interpret and make meaning from the data, ultimately enriching the research findings.

### ***2.6.2 Quantitative Data Analysis***

Data analysis plays a pivotal role in testing hypotheses and deriving meaningful insights from the collected data. Quantitative data, in contrast to qualitative data, is characterized by its objectivity and suitability for hypothesis testing. To harness the full potential of this data, a structured and technical approach is essential.

Quantitative data analysis begins with the conversion of raw data into numerical forms. This transformation allows for the application of statistical techniques to draw meaningful conclusions and identify patterns within the data. This conversion process serves as the foundation for subsequent analyses (Babbie, 2010).

Once the data is in numerical format, it is subjected to various statistical analyses. These analyses aim to uncover trends, relationships, and significant findings within the dataset. Statistical techniques such as descriptive analysis, inferential analysis, and correlation analysis are employed to explore the quantitative data thoroughly.

To conduct the quantitative data analysis in this research, Tableau Desktop 2020.1 was chosen as the primary tool. Tableau is a versatile and powerful data visualization platform that simplifies the process of transforming raw data into easily understandable visual representations. This platform does not require advanced technical skills or coding knowledge, making it accessible to researchers across various domains (Murray, 2013).

Tableau Desktop offers several advantages that make it a good choice for quantitative data analysis. It is known for its simplicity and user-friendly interface, enabling researchers to work with data effectively without the need for extensive training. This accessibility has contributed to its popularity across industries and among researchers.

Tableau Desktop excels in data visualization, allowing users to create various types of visual representations effortlessly. Researchers can generate charts, graphs, and dashboards to visualize quantitative data, making it easier to communicate findings and insights to a broader audience (Murray, 2013).

Tableau is capable of handling large datasets with ease, making it suitable for research projects that involve extensive data collection. It can process millions of rows of data efficiently,

ensuring that researchers can work with comprehensive datasets without performance limitations.

One notable advantage of Tableau Desktop is its cost-effectiveness compared to other data analysis software options like IBM SPSS. This affordability makes it an attractive choice for researchers with budget constraints, allowing them to access robust analytical capabilities without breaking the bank.

In this research, quantitative data analysis is conducted using Tableau Desktop 2020.1, a powerful data visualization platform. The technical procedures involved in converting raw data into numerical forms and applying statistical analyses are essential to draw meaningful conclusions and test hypotheses. Tableau's user-friendly interface, scalability, and cost-effectiveness make it a valuable tool for researchers seeking to analyze and visualize quantitative data efficiently and effectively.

### **2.7 Ethical Considerations**

Ethical considerations play a pivotal role in research, particularly when human participants are involved. Researchers are responsible for upholding the rights and well-being of those who contribute to the study. Throughout this work, several research instruments were deployed, each requiring careful ethical consideration to ensure that participant rights were respected and confidentiality maintained.

One of the fundamental ethical principles in research involving human participants is obtaining informed consent. Informed consent ensures that individuals willingly participate in the study with full awareness of the research's purpose, procedures, and potential risks. In this research, informed consent was a crucial aspect of participant engagement, and it was obtained from participants across various research instruments.

The initial research instrument involved the analysis of assessment artifacts. To address potential biases in this analysis, a systematic sampling process was employed. This approach helps ensure that a representative sample of artifacts is chosen, minimizing any unwanted influence on the research outcomes. Moreover, ethical considerations extended to the confidentiality of the documents under examination and the integrity of the teachers who had created it. Confidentiality was maintained to protect sensitive information, and the identities of those who created the assessments were safeguarded.

Classroom observation is a valuable tool in educational research, but it also raises ethical concerns related to privacy and participant rights. Prior to conducting classroom observations, informed consent was obtained from the participants. Anonymity was guaranteed to protect the identities of both educators and students. This step was essential to ensure that participants felt comfortable and secure in their learning environment while contributing to the study.

Surveys are commonly used in research to gather insights from participants. Similar to classroom observations, consent was obtained from participants before administering the learners' survey. Additionally, participants were provided with a clear description of the survey's purpose and procedures to ensure they were fully informed. Anonymity was guaranteed to protect the privacy of respondents, allowing them to express their opinions freely.

The use of tests as research instruments requires careful ethical consideration as well. In this research, obtaining consent from participants was a prerequisite before their participation in the test. Anonymity was guaranteed to safeguard the identities of those taking the test. Importantly, the scores obtained from the test were not communicated to the students and not used to make any student-related decision. This was made to prevent any undue stress or pressure related to test performance. Furthermore, students were rewarded for their participation in the test, ensuring that their contributions were recognized and valued.

In conclusion, ethical considerations are central to the successful conduct of research involving human participants. This research exemplifies a commitment to upholding participant rights and confidentiality across various research instruments. Informed consent, anonymity, and clear communication about the research procedures were key elements in addressing ethical concerns. By implementing these ethical safeguards, the research ensured that participants were treated with respect and that the integrity of the study was maintained. Such ethical practices are essential in promoting trust, credibility, and the ethical conduct of educational research.

### **2.8 Conclusion**

Several critical challenges and opportunities have been identified concerning the ESP situation in the Algerian context. The importance of English proficiency in technical, scientific, and medical fields is undeniable, given its global relevance. However, the proliferation of ESP courses across various disciplines in Algeria comes with significant hurdles.

One of the primary concerns is the shortage of qualified ESP practitioners. ESP practitioners are language instructors who put ESP theory into practice, whether when designing syllabi, providing instruction, or administering tests. The complexity of ESP requires educators with expertise in both the English language and the specific field of study. The practice of employing newly graduated EFL students as part-time teachers without proper training or clear syllabi compromises the effectiveness of these courses, emphasizing the need for comprehensive training programs and support for educators.

Furthermore, ESP courses in Algeria often lack alignment with learners' professional needs due to the absence of needs analysis in course design, and other socio-economic factors that would need further exploration. Limited cooperation between ESP tutors and subject matter specialists further hinders alignment with international job industry requirements. Time

constraints for ESP modules and the challenge posed by students with limited English proficiency add layers of complexity to the issue.

A significant gap in the literature pertains to ESP assessment within the Algerian context. While course design and instruction have received attention, assessment procedures remain underexplored. This gap limits the availability of informed guidance for ESP educators and impedes the establishment of clear assessment procedures. Addressing this gap requires research into diverse ESP assessment methodologies, tailored assessment guidelines, and training for ESP educators. ESP plays a vital role in preparing Algerian students for success in academic studies and professional careers. To address the several challenges identified in this study, a collaborative effort among educators, administrators, and researchers is imperative to enhance the quality of ESP instruction and assessment.

Through innovative research initiatives and a commitment to aligning ESP courses with learners' evolving needs, Algeria can ensure that its students develop the English language proficiency necessary for global competitiveness and socioeconomic development. Ultimately, investing in quality ESP education is an investment in the nation's future prosperity and global standing.

This research adopts a sequential exploratory case study approach, combining qualitative and quantitative research instruments to provide a comprehensive understanding of the research questions at hand, all while upholding ethical considerations and maintaining participant confidentiality. Data collection employed a two-fold approach: qualitative data were gathered through artifacts analysis and observation, while quantitative data were collected through surveys and by testing a diverse sample of L1 Computer Sciences (I) learners, targeting their experience of two different evaluative approaches and their impact on language development.



## CHAPTER THREE: Data Analysis and Discussions

### 3.1 Introduction

### 3.2 Qualitative Phase

#### 3.2.1 Assessment Artifacts Analysis

##### 3.2.1.2 Findings

##### 3.2.1.2.1 Content Analysis

##### 3.2.1.2.2 Thematic Analysis

##### 3.2.1.3 Discussion of the Findings

##### 3.2.1.4 Conclusion

#### 3.2.2 *Classroom Observation*

##### 3.2.2.1 Findings

##### 3.2.2.2 Discussion of the Findings

##### 3.2.2.3 Conclusion

#### 3.2.3 Conclusion

### 3.3 Planning the Second Phase

#### 3.3.1 *Informing the Design of Learners' Survey*

#### 3.3.2 *Informing the Design of the Tests*

#### 3.3.3 Conclusion

### 3.4 Quantitative Phase

#### 3.4.1 *Learners' Survey*

##### 3.4.1.1 Findings

##### 3.4.1.1.1 Exam Procedures

##### 3.4.1.1.2 Reading Skill Assessment and Development

##### 3.4.1.1.3 Writing Skill Assessment and Development

##### 3.4.1.1.4 Speaking Skill Assessment and Development

##### 3.4.1.1.5 Listening Skill Assessment and Development

##### 3.4.1.1.6 Language Skills in Specific Domain

##### 3.4.1.2 Discussion of the Findings

##### 3.4.1.3 Conclusion

#### 3.4.2 *The Test*

##### 3.4.2.1 Findings

##### 3.4.2.2 Discussion of the Findings

##### 3.4.2.3 Conclusion

#### 3.4.3 Conclusion

### 3.5 Discussion of the Main Findings

### 3.6 Conclusion

### **3.1 Introduction**

The present research follows an exploratory sequential mixed methods design, a methodology characterized by its distinctive two-phase approach (Berman, 2017). This methodology commences with an initial qualitative phase of data collection and analysis, followed by a phase of quantitative data collection and analysis, and ultimately culminating in an integration of data for a comprehensive discussion. In this chapter, we delve into the findings and discussions resulting from the systematic execution of these research phases.

During the initial qualitative phase, the study focused on collecting and analyzing qualitative data. This phase aimed to address the first research question by examining and discussing the qualitative findings derived from the collected data. The subsequent phase of the research involved the design and deployment of instruments to gather quantitative data, aligning with the second research question. The analysis and discussion of these quantitative findings constitute a crucial component of this chapter. Ultimately, the integration of findings from both phases will enable us to address the core and final research question of the study, providing a comprehensive understanding of the research topic.

In this chapter, we will first explore the analysis of qualitative data, emphasizing the interpretive and constructive philosophy that underpins this approach. We employed two qualitative research instruments, namely assessment artifact analysis and classroom observations. The analysis process involved coding and thematic analysis, enabling us to uncover meaningful patterns and themes within the data. This holistic approach to qualitative data analysis allows us to delve deeper into the nuances of ESP practitioners' evaluative practices and enrich our findings.

## ***CHAPTER THREE Data Analysis and Discussions***

---

Subsequently, we will delve into the quantitative data analysis process. Quantitative data, known for its objectivity for hypothesis testing, undergoes a structured and technical analysis. We highlight the transformation of raw data into numerical forms as the foundational step for statistical analyses. Various statistical techniques, including descriptive and inferential analyses, are employed to explore and interpret the quantitative data. Tableau Desktop 2020.1 was used to conduct the quantitative data analysis effectively. It is a versatile data visualization platform known for its simplicity, user-friendliness, and capacity to handle large datasets.

### **3.2 Qualitative Phase**

This initial qualitative phase focused on collecting data through assessment artifacts analysis and ESP Practitioners' observation during instruction. This phase aimed to address the first research question related to the nature and characteristics of assessment in the Algerian HE ESP context: How do ESP teachers at the Faculty of Sciences assess 1st Year Computer Science students?

#### **3.2.1 Assessment Artifacts Analysis**

The initial phase started with a systematic examination of twelve (12) assessment artifacts, representing nearly all the English Final Exams administered to Computer Science students (I and MI) at the Faculty of Sciences since 2010. The primary objective was to discern the fundamental characteristics of these summative exams, facilitating a comprehensive discussion regarding their design and usefulness.

### **3.2.1.2 Findings**

#### **3.2.1.2.1 Content Analysis**

This research collected a corpus of assessment artifacts that genuinely expose their characteristics. Upon a systematic analysis of these artifacts, it becomes evident that while they may diverge in various aspects, they share commonalities in terms of purpose and administration procedures. Notably, the artifacts align with the curriculum, adhering to a paper-based format with a standard duration of 60 minutes for completion. Furthermore, these assessments are characterized as criterion-referenced, focusing on mastery of predetermined learning objectives, without considering relative performance compared to peers. An important observation is the absence of feedback provisions for learners, suggesting that these assessments primarily serve as terminal evaluations without a formative dimension.

The category-based content analysis generated key findings about the planning and organization of these assessments. It is evident that the Department of Mathematics primarily organizes this event, often exerting a significant influence on the test design process. The target audience for these assessments primarily comprises Mathematics and Computer Science (MI) students, who are confronted with a pivotal decision in their academic journey, choosing between majors in mathematics and computer sciences after the first year. The prevailing trend indicates that the majority of students opt for computer sciences, making this a notable and essential consideration in the assessment design process.

Furthermore, a consistent overall structure emerges across these assessments. They typically consist of three to five tasks, each including specific instructions pertaining to the task format. These tasks are generally discrete and objective, with little to no integration. The time allocated for these assessments aligns adequately with the number of tasks and items included. The tasks cover diverse language skills, encompassing vocabulary and grammar, along with

## ***CHAPTER THREE Data Analysis and Discussions***

---

reading and writing components. Notably, the artifacts reveal a conspicuous absence of listening and speaking constructs in the specifications development processes, irrespective of the times it has been iterated.

A particularly interesting finding emerges when exploring the item-based tasks within these assessments. Among the twelve assessments examined, only two incorporate item-based tasks, presenting students with an item and requiring them to select the appropriate response. The 1st year MI Final Exam 2013-2014 (see Appendix A) shows this by including both Multiple Choice (MC) and True-False tasks. The MC task exhibits an appropriate item structure but is hindered by the low number of options per item, which increases the likelihood of guessing and subsequently affects the reliability of the assessment (two options per item). Moreover, the MC task falls short in terms of authenticity and construct validity, as it indirectly assesses the recall of grammatical knowledge, rather than language skills relevant to this ESP context.

The True-False task, from the same exam, aims to evaluate students' ability to write a letter to a stranger. However, this task, too, lacks authenticity, indirectly addressing the writing construct. Additionally, it raises reliability concerns due to the limited two options, which increase the guessing factor. A lack of clarity is observed in the True-False task, as items lack sufficient contextual support (a prompt) to enhance understanding. Both tasks, by their nature, deviate from the context of ESP and domain-specific content, with their content being better suited for an English for General Purposes (EGP) context. The True-False task from ME 2012-2013 is used to assess learners' reading comprehension skills, with items that target direct functional meanings. The number of items is low and lacks representativeness, since items do not test learner's ability to negotiate implied meanings. The tasks also lack to connect with the ESP context and learners' subsequent needs.

## *CHAPTER THREE Data Analysis and Discussions*

---

Through content analysis, the presence of selected-response tasks came emerged, shedding light on a facet of these evaluations that is less common but nonetheless significant. The frequency of selected-response tasks within these assessments is relatively low, however not as low as that of item-based tasks. Selected-response tasks in these artifacts often take the format of noticing tasks, a pedagogical approach aimed at fostering grammar acquisition. These tasks challenge learners to identify specific language features within a given context. Notably, this particular type is prominently featured in the artifacts, as evidenced in the various artifacts in Appendix A (See ME 2010-2011; FE 2011-2012; ME 2012-2013; FE 2013-2014; FE 2017-2018; FE 2018-2019). A notable trend among these noticing tasks is the focus on grammar and language feature identification, particularly regarding the recognition of silent letters in words, as exemplified in the ME 2010-2011, ME 2012-2013, and FE 2013-2014.

These tasks require learners to identify silent letters within a list of words, ostensibly assessing the speaking construct through a grammar-focused task. However, it is crucial to acknowledge that this indirect assessment of the speaking construct raises concerns about the authenticity of the task and its construct validity. This is a prime example of where the alignment of task format with objectives and construct validity should be thoughtfully considered by test designers. Other noticing tasks use a grammaticality-judgment format, compelling learners to identify and rectify grammatical errors within given sentences. The task found in the 2011-2012 Final Exam shares the limitation of lacking authenticity in its design, as it also fails to integrate domain-specific knowledge and skills, which is a fundamental aspect of ESP.

However, not all selected-response tasks exhibit such drawbacks. In the 2017-2018 and 2018-2019 Final Exams, selected-response tasks assumed the form of parsing tasks. In these tasks, learners were tasked with deconstructing sentences to label their grammatical components. This type of task, although considered somewhat unconventional for an ESP context, displayed an intriguing connection to the field of computer science. The analysis of these tasks revealed an

## ***CHAPTER THREE Data Analysis and Discussions***

---

iterative process, as the same tasks were used in two consecutive academic years. This suggests that, likely based on feedback and insights, the test designer made specific improvements to the tasks. Notably, in the second version of the task, each grammatical component of the sentences was underlined, which aimed to prevent confusion and to enable students, ultimately enhancing the clarity and effectiveness of the task. The instructions of selected-response tasks are clear and scoring is objective. Furthermore, the fifth task of the 2011-2012 Final Exam represents a para-jumble, where test-taker are simply asked to re-arrange words in a meaningful manner. The analysis revealed that this was the only time when such type of selected-response tasks was used. The task aimed at testing students' grammatical skills in meaning construction and formulation. The artifact did not provide any information related to the scoring of this task in particular. However, it is assumed to be objective.

The findings of the content analysis highlight that limited-production tasks are a prevalent and recurrent task format within the assessed artifacts. These tasks require test-takers to provide responses with restricted language production, ranging from simple words to complete sentences. For instance, text-based comprehension questions in the FE 2010-2011 and ME 2012-2013 serve as prime examples of limited-production tasks in the form of comprehension questions. These questions aim to assess the learner's capability to extract direct, functional meanings from the provided text and formulate concise responses.

Furthermore, the inclusion of open-ended topical questions in the ME 2011-2012, FE 2012-2013, and FE 2014-2015 (see Appendix A) reflects another category of limited-production tasks. These open-ended questions center on both topical and language KSAs, genuinely representing a form of Integrated Content and Language (ICL) assessment. To answer these questions accurately, students must mobilize both their linguistic and topical competencies. However, the artifacts lack information concerning the scoring process for these questions. It's worth noting that these artifacts predominantly emphasize mathematics, with limited relevance

## ***CHAPTER THREE Data Analysis and Discussions***

---

to the computer science domain. Some questions delve into highly technical concepts, potentially posing challenges for students with weaker mathematical backgrounds. Some questions required learners to provide definitions for words such as: "*Angle*," "*Radius*," and "*Circumference*," while other questions required students to calculate "*ratios*" and "*fractions*". Certain tasks even lack comprehensive instructions, relying on item-specific guidance (e.g., ME 2011-2012). The presence of typos in some items suggests a rushed test design process (e.g., ME 2011-2012 and FE2012-2013). These artifacts also reveal redundancy in the tasks and items, indicating last-minute test design adjustments.

Another common type of limited-production task is vocabulary gap-filling tasks. For example, the ME 2010-2011 and FE 2011-2012 incorporate vocabulary gap-filling tasks that align more closely with English for General Purposes (EGP) than English for Specific Purposes (ESP). While the former is context-dependent, with a passage providing contextual clues, the latter is entirely independent, lacking contextualization. Despite this, it is assumed to be curriculum-related, possibly stemming from an instruction on letter writing as it was also observed in a True-False task from the FE 2013-2014. In contrast, the vocabulary gap-filling tasks in ME 2011-2012, FE 2012-2013, and FE 2014-2015 are firmly linked to mathematics, as previously noted. These tasks offer no additional contextual information beyond the provided instructions and items, requiring students to draw upon both mathematical and linguistic knowledge and skills to obtain full scores. Nevertheless, the absence of information regarding the scoring method for these tasks remains a persistent issue. Information regarding the overall task worth is also conspicuously absent, particularly for ME 2011-2012 and FE 2014-2015. These tasks also represent an instance of Integrated Content and Language syllabus.

The prevalence of vocabulary gap-filling tasks continues in the FE 2017-2018, FE 2018-2019, FE 2019-2020, and FE 2022-2023 (see Appendix A) artifacts, with a distinct focus on the computer science domain. These tasks, based on contextualized passages or sentences, involve



## ***CHAPTER THREE Data Analysis and Discussions***

---

the removal of words based on their grammatical category and conceptual relevance. Learners receive a word list, often with extra words to enhance complexity and minimize guessing. These tasks evaluate learners' prior knowledge and reading comprehension skills to fulfill the objective. They provide greater reliability by employing a monotonous scoring method that exclusively assesses topical vocabulary, omitting writing skills. Furthermore, they indirectly evaluate reading comprehension skills through vocabulary selection, indicating alignment with the construct being tested. Clear and comprehensive instructions are provided, specifying that not all words in the list may be applicable, though an exception exists in the FE 2022-2023 exam.

In contrast to gap-filling tasks, cloze tasks involve the omission of every *n*th word, albeit without a specific rationale. The ME 2012-2013 features a cloze task more aligned with EGP that suffers from incomplete instructions, a lack of information about word usage, and no specifics about the task's overall worth. Nevertheless, the scoring procedure is clearly monotonous. FE 2022-2023 introduces a contextualized labeling task focused on vocabulary, requiring learners to determine both the label (topical) and its written form (linguistic). This type of task also necessitates a dichotomous or polytomous scoring method, although details are missing from the artifact.

Moreover, other gap-filling tasks concentrate on grammatical components, applicable to both ESP and EGP contexts. For instance, the ME 2010-2011 involves a task targeting cohesive forms that aligns better with EGP. Learners are provided with an instruction but lack further guidance. In contrast, FE2017-2018 and FE 2018-2019 contextualize grammar-focused gap-filling tasks within the computer science domain. These tasks provide learners with answer choices and examples for task completion. Clear and comprehensive instructions are offered. The scoring method for these tasks should encompass two linguistic criteria: the choice of the cohesive word and the correct sentence combination. The instructions for these tasks distinctly indicate a dichotomous scoring approach, as in the third task of the FE 2017-2018.

## ***CHAPTER THREE Data Analysis and Discussions***

---

The last form of limited-production tasks is a dialogue completion task from the 2010-2011 Final Exam. It primarily lacks relevance to the specific domains of learners. This task evaluates the ability to decode written functional and implied meanings from provided sentences and encode the meanings to fill in with sentences in the dialogue. However, information about the scoring procedures for this task is conspicuously absent. To ensure the reliability and validity of this task, a dichotomous or polytomous scoring method, or potentially a scoring rubric, should be aligned with the task's objectives and the constructs being assessed.

The examination of the artifacts has shown that the inclusion of extended-production tasks is somewhat limited within the assessment design process. Typically, these tasks offer prompts instead of specific items, with the aim of eliciting substantial responses from test-takers. They are particularly well-suited for assessing grammatical competence in both spoken and written language. However, the analysis did not reveal the presence of performance-focused tasks such as simulations and storytelling, or process-focused tasks like observation and reflective activities.

Two specific artifacts, namely the ME 2010-2011 and ME 2011-2012, featured a product-focused task. In these tasks, learners were required to craft a summary of their respective projects. The 2010-2011 Makeup Exam transitions from an English for General Purposes (EGP) focus in the first three tasks to an English for Specific Purposes (ESP) focus in the final task. Conversely, the 2011-2012 (see Appendix A) Makeup Exam underscores its Integrated Content and Language (ICL) nature from the outset. Despite the contextual differences, the two extended-production tasks are identical in instructing learners to summarize their projects in six (6) lines. Unfortunately, these tasks lacked guidelines, prompts, and scoring rubrics, which would have been beneficial for learners in understanding the task requirements and evaluation criteria. Furthermore, the incorporation of prompts could have clarified the Target Language Use (TLU) domain, enabling learners to determine the form and content of their products. The

## ***CHAPTER THREE Data Analysis and Discussions***

---

absence of these elements notably impacts the task's reliability and authenticity. Additionally, the validity of these tasks is in question, as they might not be aligned with their intended purpose i.e., to write a project summary. In such cases, scoring should have been derived from the construct's definition.

The analysis of the FE 2010-2011, the FE 2011-2012, and the ME 2012-2013 also revealed that certain exams were presented in the form of texts with associated text-based tasks (see Appendix A). These exams typically consisted of two to three sections, with the first section focusing on comprehension, the second on grammatical knowledge, and the third featuring a writing task, either in a limited or extended production format. While the comprehension questions were presented in English, it was unclear whether linguistic accuracy was considered in the scoring process of the responses. The initial sections of the FE 2010-2011 and the ME 2012-2013 contained three to four comprehension questions, with moderate language complexity, targeting learners' understanding of direct functional meanings. The texts in the FE 2010-2011, the FE 2011-2012, and the ME 2012-2013 (see Appendix A) all comprised three paragraphs and revolved around familiar EGP topics, such as travel and natural disasters. Furthermore, the reading comprehension sections in the FE 2011-2012 and ME 2012-2013 also included vocabulary-in-context tasks related to content and functional words. However, in the 2010-2011 Final Exam, the vocabulary-in-context task was included in the grammar section, signaling a potential issue during the assessment design process.

The examination of the 2011-2012 Final Exam revealed a couple of structural issues. This particular test featured only two sections, with the first section titled "reading comprehension" focusing primarily on grammar, rather than comprehension. Learners were tasked with grammar-noticing exercises and para-jumbles. Notably, no information regarding the total score of sections, questions, or items was provided on the artifacts. The grammar-related tasks in the FE 2010-2011, the FE 2011-2012, and the ME 2012-2013 aimed at testing learners'

## ***CHAPTER THREE Data Analysis and Discussions***

---

morphosyntactic knowledge through selected-response tasks concerning passive/active voice and direct/indirect speech.

The final sections of these exams assessed learners' writing skills through tasks involving limited production and selected response formats (see Appendix A). For these tasks, as the dialogue completion task in the FE 2010-2011, a dichotomous or polytomous scoring method would have been appropriate. The reliability and validity of these tasks were called into question due to misalignment between scoring, test format, content, the tasks' actual objectives, and the whole context of the assessment. Notably, the analysis of the artifacts highlighted the absence of assessment components related to speaking and listening skills, except in the case of the silent-letters-noticing tasks in the ME 2010-2011, the ME 2012-2013, and the FE 2013-2014 (see Appendix A), which indirectly assessed the speaking skill through grammar-focused exercises, suggesting issues related to validity, reliability, and authenticity.

### **3.2.1.2.2 Thematic Analysis**

To explore the underlying patterns and themes within these assessment artifacts, our research used thematic analysis in combination with category-based content analysis. This holistic approach allowed us to gain insights that may not have been immediately apparent through traditional content analysis alone. After conducting a thematic analysis of the selected assessment artifacts (see Appendix A), several key aspects emerged, shedding light on various issues and patterns within the assessed artifacts. The analysis has revealed ten (10) themes that can be summarized as follows:

- a) **Dysfunctional Assessment Design:** Recurring issues in the overall design of assessments point to inherent problems that need addressing, such as untrained, constantly replaced, task designers. These issues can be observed in the overall artifacts structure, task specifications, and scoring methods among many other aspects

## *CHAPTER THREE Data Analysis and Discussions*

---

that signal problems related to the assessment design process. The 2011-2012 Makeup Exam (see Appendix A), for instance, includes two gap-filling tasks that assess the same construct. The items of these two tasks, besides their non-alignment with the gap-filling tasks specifications, could have been presented in one single task, instead of two.

- b) **Baccalaureate-oriented Construction:** The corpus highlights concern regarding the repeated application of the Baccalaureate English Exam blueprint and drafts for designing assessments in this ESP context, without adaptation or innovation as in the case of the FE 2010-2011, the FE 2011-2012, and the ME 2012-2013 (see Appendix A). The artifacts based on the Baccalaureate English Exam model showed a reliance on EGP topics, including passage-based tasks that are not necessarily relevant to learners' needs in this ESP context. The standardization of these ESP assessments based on the Baccalaureate English Exam blueprint also influences tasks format and item development, as exemplified by the FE 2011-2012 requiring students to report on the number of interrogative sentences in the text.
- c) **Content Redundancy:** The analysis identified a pattern of recurring tasks, items, and content, suggesting a lack of diversity and freshness in instructional materials and evaluative tools. For instance, the extended-production tasks in ME 2010-2011 and ME 2011-2012 are identical. Moreover, the similarity of items in ME 2011-2012 and FE 2014-2015 is another example of how tasks and items are cloned in the test design process.
- d) **Lack of Item-Based Tasks:** The thematic analysis exposes the absence of item-based tasks in these artifacts, and particularly MC tasks that are used worldwide to test reading comprehension. Mainly, comprehension is tested through open-ended questions and vocabulary-in-context tasks.

## ***CHAPTER THREE Data Analysis and Discussions***

---

- e) **Selected-Response Challenges:** The corpus exposes issues related to guessing and cheating in item-based and selected-response tasks, drawing attention to their potential vulnerabilities and to their systematic design that takes into consideration both the context and the purpose (e.g., FE 2013-2014). The low number of items is not sufficient to provide reliable data on learners' mastery of the construct being assessed. The item construction procedures represent another challenge in this very particular type of tasks, since the test designers have to abide by the task-related specifications. However, as in the case of the first task in FE 2012-2013, items are somewhat unclear and blanks can be filled with one or two words depending on the item. Moreover, the FE 2014-2015, for instance, includes a gap-filling task (*Task N°3*) that assesses multiple constructs (see Appendix A).
- f) **Usefulness Challenges:** Thematic analysis highlights concerns about the authenticity of assessments, particularly with the overwhelming use of indirect, discrete, uncontextualized, weak, and non-integrated task formats. Moreover, the analysis underscores the absence of integrated tasks targeting real-life competencies, which raises questions about their validity in general. The tested constructs are sometimes conflictive as they vary from purely linguistic to purely topical, depending on the test designer. This causes 1<sup>st</sup> Year CS English exams to be biased towards a group of students. Eventually, the absence of scoring rubrics and point allocation problems in assessments emerged as notable challenges in the corpus.
- g) **Lack of Speaking and Listening Tasks:** Thematic analysis reveals the total absence of speaking and listening tasks in these exams, signaling a gap in language skill evaluation. Pronunciation, as shown in the third task of the ME 2010-2011, is sometimes assessed through indirect methods.
- h) **Content and Language Integration Challenges:** The corpus showcases fairness issues related to the integration of content, language, and skills within the assessments. The

## ***CHAPTER THREE Data Analysis and Discussions***

---

polarity between assessments based on Mathematics (ME 2011-2012; FE 2012-2013; FE 2014-2015) and the ones based on Computer Science (FE 2017-2018; FE 2018-2019; FE 2019-2020; FE 2022-2023) is at the core of assessments being unfair to learners. The ME 2011-2012, FE 2012-2013, and FE 2014-2015 have shown more ambition to evaluate learners' topical KSAs than language KSAs. This indicates the need for a more cohesive approach, authentic, and learner-oriented, approach.

- i) **Emphasis on Lower Cognitive Skills:** The analysis highlights an emphasis on lower-order cognitive skills in the examined tasks, potentially limiting the depth of learning and development. The majority of the selected assessment artifacts are based on recalling tasks, as shown in the FE 2013-2014, FE 2019-2020, and the FE 2022-2023). Other tasks required learners to explain ideas, or use information in new situations, as found in the FE 2012-2013 and the FE 2017-2018. However, no task at all has asked students to connect various ideas, justify their viewpoints, or produce an original work.
- j) **Assessment Evolution:** The analysis points to the evolution in the assessment design process, with a great variation in the content. This is probably related to the different backgrounds of each test designer, indicating the need for greater consistency and reliability. Assessments have evolved through three main stages, the first being the EGP-based ESP assessment (FE 2010-2011; ME 2010-2011; FE 2011-2012; FE 2012-2013). The second stage is characterized by an over reliance on mathematical knowledge and skills, as shown in the ME 2011-2012, FE 2012-2013, and FE 2014-2015. The third and final stage is characterized by its convergence towards computer science topics and knowledge, not necessarily involving topical skills, illustrated in the FE 2017-2018, FE 2018-2019, FE 2019-2020, and FE 2022-2023 (see Appendix A).

### **3.2.1.3 Discussion of the Findings**

The evaluation of students' language proficiency in English for Specific Purposes (ESP) is a critical component of higher education, particularly in fields like computer science, where effective communication in a global context is essential, considering the exponential development of the field. The following discussion delves into the summative evaluative practices employed by ESP practitioners at the Faculty of Sciences in Algeria, when evaluating 1st Year Computer Science students during final or makeup exams. To enrich this discussion, we incorporate findings from category-based and thematic analyses of assessment artifacts, providing a holistic perspective of the issues and patterns within the examined materials.

The assessments are predominantly criterion-referenced, focusing on measuring students' mastery of predetermined learning objectives. These assessments follow a paper-based format with a standard duration of 60 minutes. However, one significant issue revealed by the content analysis is the absence of feedback provisions, suggesting that these assessments serve solely as terminal evaluations without a formative dimension. This absence of formative assessment can hinder students' learning and improvement, particularly when the design process is not iterative and does not inform curriculum design. What was observed in the ME 2011-2012, FE 2012-2013, and FE 2014-2015, in addition to the FE 2010-2011, FE 2011-2012, and ME 2012-2013 (see Appendix A) is more of a redundant repetition of tasks and items rather than iteration of the design process, since these artifacts did not show a real enhancement or innovation. The analysis of FE 2017-2018, FE 2018-2019, FE 2019-2020, and FE 2022-2023 has revealed that the design process was somewhat revisited during iterations, since instructions, items, and tasks have been edited from one artifact to another.



## ***CHAPTER THREE Data Analysis and Discussions***

---

The thematic analysis brought to light the Department of Mathematics' significant role in organizing these assessments. While this influence is evident, it can lead to a misalignment between assessment content and the specific needs of computer science students. The assessments must be adapted to also meet the requirements of the future students of computer science domain and mathematics domain as well. The ME 2011-2012, FE 2012-2013, and FE 2014-2015 (see Appendix A) include tasks that would have been better suited for an English as a Medium for Instruction (EMI) mathematics class, representing another type of class in the ESP context. However, the primary aim of these ESP exams is to assess the linguistic component primarily, then the topical component, because they are basically subsequent evaluations of an English language course. The FE 2017-2018, FE 2018-2019, FE 2019-2020, and FE 2022-2023 rely heavily on the computer science domain and fail at integrating the mathematical component. However, they do not prioritize topic over language and include ICL tasks that assess both linguistic and topical constructs.

Assessment tasks typically consist of three to five items covering various language skills, including vocabulary, grammar, reading, and writing. However, a conspicuous absence of listening and speaking constructs in the assessment process is noted. This omission raises questions about the comprehensiveness and authenticity of the assessments, particularly given the real-world demands on computer science professionals. The analysis has shown that some tests included indirect pronunciation tests that are not contextualized with a primary focus on grammar.

The thematic analysis identified ten key themes that highlight various challenges and issues within the assessment artifacts. The research found recurring issues in the overall design of assessments, triggered by problems related to untrained, not motivated, and frequently replaced educators or assessors, as suggested by the literature (see Assassi, 2020). These issues are reflected in the structure of the assessment artifacts, task specifications, and scoring methods,

## ***CHAPTER THREE Data Analysis and Discussions***

---

among other aspects. For example, in the 2011-2012 Makeup Exam, two gap-filling tasks assess the same construct, which could have been presented in a single task, illustrating a lack of alignment and efficiency in the assessment design. Dysfunctional assessment design affects not only the reliability and validity of the assessments but also students' perception of the evaluation process. Inconsistent and poorly designed assessments may lead to frustration and disengagement among students. This, in turn, can hinder the overall learning and development experience.

The study identified concerns regarding the repeated application of the Baccalaureate English Exam model for designing weak and passage-based ESP assessments without adaptation or innovation (see Benmoussat & Benmoussat, 2018). Teachers are clearly guided by their subjective beliefs when creating such assessments (see Boubris & Haddam, 2020). This standardization influences the format and content of tasks, which may not be relevant to the learners' needs in the ESP context. Standardized assessments that do not align with the specific requirements of computer science students may lead to a gap in their language proficiency. Students may not acquire the language skills necessary for their field, potentially affecting their future career prospects and global competitiveness.

The thematic analysis also revealed a pattern of recurring tasks, items, and content, suggesting a lack of diversity and freshness in instructional materials and evaluative tools. For instance, identical extended-production tasks in different assessment artifacts indicate a repetitive rather than a reflective-iterative assessment design process. Sometimes, the task format and items are overused. Content redundancy can lead to a limited exposure to diverse language contexts and skills. Students may become overly familiar with specific task formats, reducing the challenge and variety in their assessments. This may limit their language development and readiness for real-world communication.

## ***CHAPTER THREE Data Analysis and Discussions***

---

The absence of item-based tasks, particularly multiple-choice tasks, which are widely used for testing reading comprehension, is a notable concern. The artifacts show that comprehension is mainly tested through open-ended questions and vocabulary-in-context tasks that target direct function meanings. The lack of varied item-based tasks can affect the reliability and practicality of assessments. Open-ended questions, while valuable for assessing deeper understanding, may be more time-consuming to score. This can lead to delays in providing feedback, or tendencies towards reducing the number of items, potentially impeding students' progress. The research brought to light issues related to guessing and cheating in item-based and selected-response tasks, highlighting vulnerabilities in their design. The low number of items and unclear item construction procedures may also compromise the reliability of these tasks. Challenges in selected-response tasks, such as guessing and cheating, can undermine the integrity of the assessment process. This, in turn, affects the fairness and credibility of students' language proficiency evaluations.

The thematic analysis raised concerns about the authenticity of assessments due to the overwhelming use of indirect, discrete, uncontextualized, and non-integrated task formats. The absence of integrated tasks targeting real-life competencies weakens the assessments' validity argument. Additionally, the absence of scoring rubrics and point allocation problems emerged as notable challenges in the artifacts. Assessments that lack authenticity and clear scoring criteria may leave students uncertain about their performance and how to improve.

The 1<sup>st</sup> year cohorts in the Algerian context are well known for their massive number of students, especially in STEM domains. It is, hence, very challenging to manage the instruction and evaluation of, sometimes up to 800, 1<sup>st</sup> year students, especially in the Algerian higher education context where, as reported by research, the lack of qualified and versatile ESP practitioners is critical. Therefore, it is assumed that most of the issues revealed by the content and thematic analyses emerge from assessment practicality and available resources. In addition

## ***CHAPTER THREE Data Analysis and Discussions***

---

to the design of these summative assessments, the test planning, administration, and scoring procedures are also impacted by practicality. Thus, practicality becomes the locomotive of assessment usefulness, shadowing validity, reliability, and fairness, and triggering a strong washback effect. ESP teachers may rely on traditional assessment methods, even when they are not the most appropriate for the learning objectives, due to a number of factors such as limited training in assessment methods, lack of access to resources, and heavy workloads (Davidson & Coombe, 2019; Raffas, 2023).

### **3.2.1.4 Summary**

The summative assessment practices employed by ESP practitioners at the Faculty of Sciences in Algeria, when evaluating 1st Year Computer Science students, reveal a range of challenges and issues that warrant careful consideration. The findings signal a significant gap in ESP teachers' assessment literacy. The issues, as illuminated by the thematic analysis of assessment artifacts, have implications not only for the assessment process but also for the students and their language development. In light of the findings, it is essential for ESP practitioners to revisit and adapt their assessment practices to better serve the needs of 1st Year Computer Science students. Stakeholders and policy makers, in addition to researchers and informed practitioners, have a pivotal role to play in the enhancement of the actual, somewhat problematic, ESP situation. Assessments with feedback provisions, alignment with the computer science domain, inclusion of listening and speaking constructs, and addressing the identified challenges in assessment planning and design can contribute to more effective and authentic language proficiency evaluations. Ultimately, these improvements will not only benefit students but also enhance the overall quality of education in the field of computer science.

### **3.2.2 Classroom Observation**

## ***CHAPTER THREE Data Analysis and Discussions***

---

The qualitative phase also involved a systematic observation of four (04) ESP practitioners during instruction. These practitioners teach within various departments of the same Faculty, addressing diverse first-year audiences, which extend beyond Computer Science. The primary aim was to identify the essential features of classroom formative assessment in these contexts, thereby contributing additional insights to the discussion concerning their usefulness, with a particular focus on their impact. The observation was based on the guide described in Table 2.3 (see Appendix B)

### ***3.2.2.1 Findings***

The key findings from the guided classroom observations, which aimed to document the prevalent formative assessment practices employed by ESP practitioners within the observed classrooms, are presented below. The observation centered on three ESP practitioners responsible for 1st-year students from diverse backgrounds. These teachers were selected for convenience and observed on two separate occasions.

The predominant form of classroom evaluation observed revolved around informal and unplanned questions posed by teachers. These questions were often prompted by the flow of instruction and occasionally by the need to regain student focus (e.g., T2-Ob1: "Do you have any questions?"; T3-Ob2: "Is it clear?"). It is noteworthy that the primary purpose of these questions did not appear to be the generation of constructive feedback. Instead, they were often employed to assess students' engagement with the topic.

The observational process revealed a distinct lack of diversity in classroom assessment tools and strategies. A significant proportion of class time was dedicated to instruction, with sparse instances of formative assessment, such as practice quizzes and other instructional activities. T1, in particular, used domain-related cued gap-filling activities and oral quizzes to

## *CHAPTER THREE Data Analysis and Discussions*

---

assess learners' understanding, providing authentic assessments that included real-world contextual examples.

Most of the informal and unplanned questions demonstrated a predominantly formative character. These questions were typically conducted in English, occasionally incorporating French or Arabic for clarification of new concepts. The informal questions posed by T2 and T3 indirectly targeted students' linguistic knowledge, encompassing vocabulary, background knowledge, and self-assessment of their learning progress throughout various instructional phases. These questions scarcely addressed learners' topical knowledge or their ability to apply domain-specific skills in practical scenarios. Although feedback was primarily centered on grammar and pronunciation, it did not offer substantial topical assistance, except in the case of T1.

All observed teachers demonstrated a shared commitment to providing immediate feedback to students. T1 employed visual aids and encouraged student participation, fostering an interactive learning environment. However, T2 and T3 exhibited lower levels of interactivity, contributing to a more teacher-centered classroom atmosphere. The only instance of planned formative assessment embedded within the lesson plan was observed in T1's classes. Generally, the observed classroom assessments did not appear to be integral components of the instructional process, with a lack of structured instructional activities. On occasion, instruction did include evaluative elements that primarily tested students' prior knowledge before introducing new content.

A distinct lack of teacher expertise in interpreting formative assessment data and adjusting instruction accordingly was noted in the observations of T2 and T3. Conversely, T1 demonstrated proficiency and theoretical knowledge in instructing, evaluating, and providing feedback. T1's planned instructional assessments were engaging and involved students in the

## ***CHAPTER THREE Data Analysis and Discussions***

---

learning process. In the observed sessions, students often assumed passive roles, displaying limited initiative to ask questions or actively engage in discussions concerning their learning gaps and needs. In contrast, T1's students exhibited higher interactivity.

T1 successfully integrated formative assessment into the classroom environment, creating a supportive atmosphere. Classroom tasks were presented using video projectors, with questions aligned with the evolving instruction. However, T2 and T3 exhibited challenges in adapting to instruction and assessment, leading to less supportive, somewhat teacher-centered learning environments. Nevertheless, all observed teachers displayed a readiness to provide additional instruction when students expressed particular needs. T1, in particular, asked follow-up questions after offering instructive feedback, fostering a supportive learning environment.

The majority of formative assessment practices within the observed classrooms were oriented towards providing immediate assistance to students, rather than facilitating a deeper understanding of the learning process or encouraging reflection on the curriculum. T1 demonstrated a propensity to encourage students to take responsibility for monitoring and supporting their own learning. This included suggesting that students watch specific YouTube videos to enhance their understanding of certain grammatical aspects. T1 also informed the students about the existence of online worksheets on the university e-learning platform (Moodle) for those who were interested.

These classroom observations illustrate a spectrum of formative assessment practices among ESP practitioners at the Faculty of Science, Tlemcen University. T1 emerged as a notable example of embedding formative assessment within the instructional process, encouraging a supportive and interactive learning environment. Conversely, T2 and T3 appeared to adhere to traditional, less interactive teaching approaches.

### ***3.2.2.2 Discussion of the Findings***

## ***CHAPTER THREE Data Analysis and Discussions***

---

The essential features of classroom formative assessment in the Algerian Higher Education ESP context can be discussed based on both the findings and the literature. One of the most important characteristics of the ESP situation in Algeria is that Algerian HEIs suffer mainly from the same drawbacks related to the lack of full-time, trained, ESP practitioners (Akkar & Idri, 2021; Assassi, 2020; Fehaima, 2022; Hadj Djelloul & Melouk, 2022; Khadam, 2023). It can be said, first, that classroom-based ESP formative assessment should involve a variety of tools and strategies to cater to the diverse learning needs of the Algerian students. The findings suggest that T1 was more effective in this regard, using a range of activities such as domain-related cued gap-filling activities, oral quizzes, and visual aids. Other CATs were seemingly lacking, limiting the real potential of these ESP courses.

Formative assessment should be integrated with instruction, rather than being seen as a separate activity. This allows for timely and targeted feedback to be provided to students, supporting their learning throughout the instructional process. T1 was again more effective in this regard, embedding planned formative assessment activities within their lesson plans. Such practices enhance classroom-based assessment validity and reliability, making for a positive washback effect on learners with targeted, sometimes personalized, feedback.

Teachers should have the expertise to interpret formative assessment data and adjust instruction accordingly. This requires a deep understanding of the learning process and the ability to design and implement effective evaluative interventions. T1 demonstrated a higher level of teacher expertise in this regard, using formative assessment tools and data to inform instruction and provide tailored feedback to students. The findings also underscore the necessity of teacher training and the adoption of diversified, structured formative assessment practices that better align with students' needs and learning objectives.



## ***CHAPTER THREE Data Analysis and Discussions***

---

Students should be incentivized to actively engage in the formative assessment process. This also means providing them with opportunities to reflect on their own learning, identify their strengths and weaknesses, and set goals for improvement. T1 was more effective in promoting student engagement, encouraging students to take responsibility for their own learning and providing them with support to do so.

The validity of classroom formative assessment refers to the extent to which it measures what it is intended to measure. This would involve assessing the extent to which the formative assessment practices employed by the observed teachers were accurately measuring students' learning progress. The findings suggest that T1's formative assessment practices were more valid than those of T2 and T3. This is because T1 used a variety of assessment tools and strategies that were aligned with the learning objectives, and they provided students with opportunities to demonstrate their learning in a variety of ways. T2 and T3, on the other hand, relied more heavily on informal and unplanned questions, which may not have been as effective in assessing students' learning progress.

The reliability of classroom formative assessment refers to the consistency with which it measures what it is intended to measure, centering on the extent to which the formative assessment practices employed by the observed teachers produced consistent results over time. The findings suggest that T1's formative assessment practices were more reliable than those of T2 and T3. This is because T1 used a more structured and systematic approach to formative assessment, with planned assessment activities embedded within their lesson plans. T2 and T3, on the other hand, relied more on informal and unplanned questions, which may have been more susceptible to inconsistency.

Classroom-based formative assessment practicality has been identified as another hindrance to the assessment process. Practicality, in this context, refers to the ease with which

## ***CHAPTER THREE Data Analysis and Discussions***

---

CATs can be implemented in the classroom, considering the extent to which the formative assessment practices employed by the observed teachers were feasible and time-efficient. The massive number of students in 1st year cohorts along with wide lecture halls where instruction took place do not represent favorable condition for CATs deployment and CATs-based feedback provision. The findings suggest that T1's formative assessment practices were more practical than those of T2 and T3. T1 could eventually use a variety of assessment tools and strategies that were relatively easy to implement and did not require a significant amount of time. T2 and T3, on the other hand, relied more heavily on informal and unplanned questions, which may not be best suited for relevant and personalized feedback provision, even if not really time-consuming.

The impact of classroom formative assessment refers to the extent to which it leads to improved learning outcomes, considering the extent to which the formative assessment practices employed by the observed teachers helped students to learn more effectively. The findings suggest that T1's formative assessment practices had a more positive impact on student learning than those of T2 and T3. This is because T1's assessment practices were more aligned with the essential features of formative assessment, as they also foster active learning through the provision of latent feedback based on external instructional material such as worksheets. T1's students also exhibited higher levels of engagement and interactivity, which are associated with improved learning outcomes.

### ***3.2.2.3 Summary***

The findings of this study suggest that there is a need for ESP practitioners in Algerian Higher Education to develop a deeper understanding of the essential features of useful ESP formative assessment and how to implement them in the classroom. T1's formative assessment practices provide a notable example of how formative assessment can be embedded within the

## ***CHAPTER THREE Data Analysis and Discussions***

---

instructional process to create a supportive and interactive learning environment that promotes student engagement and achievement.

### ***3.2.3 Conclusion***

The assessment of 1st year computer science students in the ESP context at the Faculty of Sciences in Algeria is a complex and challenging task. ESP practitioners must balance the need to assess students' English language proficiency with the need to assess their understanding of the specific content of their computer science courses. One of the most important aspects of effective ESP assessment is the use of a variety of tools and strategies. This allows ESP practitioners to cater to the diverse learning needs of their students and to assess a wide range of linguistic and topical KSAs.

ESP practitioners should also integrate formative assessment into their instruction. This allows them to provide timely and targeted feedback to students, supporting their learning throughout the instructional process. When designing and implementing summative assessments, ESP practitioners should keep in mind essential theoretical principles such as alignment, validity, and reliability, among many others. The impact of ESP summative assessment on student learning is highly significant. Effective ESP assessment can help students to identify their strengths and weaknesses, set goals for improvement, and develop self-assessment skills, in addition to becoming more motivated learners which will improve their overall academic and professional achievement.

ESP teachers at the Faculty of Sciences can play a vital role in supporting the learning of 1st year computer science students by developing and implementing useful assessment practices. Trained ESP practitioners can create assessments that are valid, reliable, practical, and aligned

with the learning objectives of their courses. This can have a positive impact on student learning, helping students to develop the English language skills and knowledge they need to succeed in their academic and professional careers.

### **3.3 Planning the Second Phase**

The research employs a sequential mixed-methods approach, an approach that combines both qualitative and quantitative instruments to gain a comprehensive understanding of evaluative practices in the context of language skills development (Alam & Aktar, 2019; Gogo & Musonda, 2022; Schoonenboom & Johnson, 2017). This approach allows us to explore the correlation between ESP practitioners' evaluative practices and learners' language skills development from diverse perspectives.

This exploratory sequential mixed methods design consists of two distinct phases. The first phase involves qualitative data collection and analysis, primarily utilizing document analysis and observation methods. In this phase, an in-depth examination of the first research question is undertaken, employing coding and thematic analysis to produce key insights.

The insights gained from this initial qualitative phase significantly influence the planning and execution of the second phase of sequential exploratory research (Almeida, 2018; Berman, 2017; Gogo & Musonda, 2022). The design and administration procedures of the quantitative instruments used in the present research, including surveys and tests, have taken into consideration the main findings of the preceding qualitative phase. While this phase may not have a distinct name, it holds a pivotal role within the sequential exploratory research design. It is considered a substantial and indispensable component, serving as a bridge between the qualitative and quantitative aspects of sequential designs.

## ***CHAPTER THREE Data Analysis and Discussions***

---

### ***3.3.1 Informing the Design of Learners' Survey***

The qualitative findings from this research have pointed out to a significant negative washback effect on the development of 1st year Computer Science learners' language skills. These findings have revealed various critical issues, particularly concerning the inherent dysfunctionality in the design of assessments. Notably, the absence of evaluating listening and speaking skills in both formative and summative assessments stand out as a primary concern. Additionally, the standardized task formats, which do not necessarily align with the learners' needs, further compound the complexities. The involvement of part-time teachers, lacking not only experience but also training in ESP and assessment practices, adds another layer of complexity to this situation.

In sequential mixed-methods research, the qualitative data obtained is generally used in identifying crucial variables and concepts to measure in the subsequent quantitative phase. The focus of this research revolves around exploring how various factors associated with Computer Science ESP assessments influence the development of learners' language skills based on their experiences. Consequently, the design, structure, and variables of the survey were partly informed by the findings derived from the qualitative phase.

The intensity of the negative washback effect, as suggested by the qualitative findings, prompted the survey to address a broader population, encompassing 1st, 2nd, and 3rd year Computer Science students. The data also revealed that the evaluative practices within this context do not effectively contribute to the development of domain-related language skills. The variation in their design across academic years, oscillating between extremes on the language-content continuum, causes lacks in reliability, validity, and fairness.

To discern the extent to which washback influences the teaching-learning experience based on test-takers' experiences, the survey targets all undergraduate Computer Science

## ***CHAPTER THREE Data Analysis and Discussions***

---

students. The focus is specifically directed towards the high-stakes Computer Science Final ESP Exams, taking into account the deficiencies observed in listening and speaking assessments, particularly in the first year. However, it's crucial to investigate if similar evaluative practices are employed in other academic levels or if ESP assessments at different levels have varying impacts.

The survey primarily revolves around the many factors linking ESP assessment to the development of learners' language skills. These factors include the overall exam procedures within the given context, as well as learners' self-perception of how exams have contributed to their general or domain-specific language skills development. Importantly, the qualitative data have been pivotal in creating survey items that invite learners to rate their agreement with given statements using a Likert scale.

### ***3.3.2 Informing the Design of the Tests***

The initial qualitative phase within this sequential exploratory research design has revealed insights through content and thematic analyses regarding the drawbacks within ESP assessment design. These identified issues span every facet of evaluation, encompassing content, targeted proficiency, task formats, and the overarching approach to ESP testing. While some artifacts reflect a baccalaureate model with a more traditional approach to language testing, heavily criticized in the literature, others follow a sophisticated form of testing known as Integrated Content and Language (ICL) assessment, combining language proficiency with content-specific knowledge and skills.

Artifacts rooted in the baccalaureate model exhibit a lack of alignment with learners' needs, particularly those tailored towards international market demands. They inherently embed numerous drawbacks that impede their usefulness in an ESP context, failing to facilitate the development of learners' language proficiency. Assessments inspired by the baccalaureate-model

## ***CHAPTER THREE Data Analysis and Discussions***

---

English exam, whether text-based or not, lack innovation and development, ultimately failing to propel learners' language proficiency to a higher level.

The findings have also revealed some mathematics-based assessments that combine language and content, aligning with modern theories on ESP and ICL. However, these assessments were found to sometimes emphasize mathematics over language in an English language course, which is concerning as it does not fairly evaluate students based on their language proficiency. Additionally, other forms of ICL assessment observed during the artifacts analysis show a commitment to a specific domain, such as Computer Science, using it as a framework to evaluate students' linguistic skills. Nevertheless, these assessments are marked by redundancy in task formats and a lack of substantial authenticity, failing to present a real-life, domain-specific use case of the English language.

The qualitative data derived from these analyses informed the creation of two tests administered to first-year Computer Science students in the subsequent quantitative phase. The first test, referred to as the "generic" test, shares characteristics and similarities with the analyzed artifacts, mainly offering simple instructions, scarce information on scoring procedures, and lacking feedback provisions (see Appendix E). On the other hand, the second test, or the ICL test (see Appendix F), aligns more closely with modern ESP and ICL assessment theories. This test, tailored to a language course in a Computer Science (ESP) context, prioritizes language over content and integrates topical skills and knowledge in a more authentic approach that caters to professional market needs.

Overall, both tests are narrowly dependent a domain-related topic i.e., requirements elicitation. The qualitative insights primarily influenced the design of the generic test, while the ICL test was crafted based on state-of-the-art assessment theory. The qualitative data served to identify drawbacks in the actual ICL evaluative practices such as the lack of MC tasks, prompts,

## ***CHAPTER THREE Data Analysis and Discussions***

---

and scoring rubrics, in addition to feedback, to guide the design of the second test and prevent the recurrence of these deficiencies. Both tests conclude on an ICL culminating competency that is elicited through a limited-production dialogue-completion task.

Additionally, the tests cover the same constructs but are somewhat presented differently. The generic test, however, does not include a task that integrates topical knowledge and skills through reading, besides feedback provisions. Therefore, the maximum achievable score in the Generic Test (30 points) and its duration (90 minutes) are inferior to the maximum achievable score in the ICL Test (35 points) and its duration (120 minutes). One could intuitively assume that this difference, in terms of maximum achievable points, would eventually bias the data, especially when expressing the scores using the Algerian grading system, which ranges from 0 to 20.

The final grades will be calculated based on algebraic proportionality, expressing equality between ratios. For instance, if the ratios  $a/b$  and  $c/d$  are equal, the relationship between  $a$  and  $b$  will be the same as the relationship between  $c$  and  $d$ . In this sense, proportions are used to solve problems where one of the four quantities is unknown. To solve a proportion, cross-multiplication and division are used. This means multiplying the numerator of one ratio by the denominator of the other ratio and dividing by the remaining numerator or denominator. The equations below show how students' final grades were calculated for both tests.

- Generic Test Grade Calculation:  $GTG = (GTS * 20) / 30$

Where: GTG represents the Generic Test Grade; GTS represents the Generic Test Score (obtained by the learner); 20 represents the highest grade in the Algerian grading system; and 30 represents the highest achievable score.

- ICL Test Grade Calculation:  $ITG = (ITS * 20) / 35$



## ***CHAPTER THREE Data Analysis and Discussions***

---

Where: ITG represents the ICL Test Grade; ITS represents the ICL Test Score (obtained by the learner); 20 represents the highest grade in the Algerian grading system; and 35 represents the highest achievable score.

Furthermore, the culminating tasks from both tests, which are the most relevant to the present research, are scored using the same scoring rubric (with a maximum of 15 points). Hence, the data from both tasks will not be influenced by the difference in the maximum achievable scores, maintaining reliability and validity. Ultimately, learners' scores from the third enabling task in the ICL test can also be put aside to demonstrate the reliability and validity of the data.

### ***3.3.3 Conclusion***

The qualitative insights played a fundamental role in informing the planning and execution of the subsequent quantitative phase, particularly in the design of surveys and tests. These instruments were tailored to address the identified deficiencies, targeting a wider population including 1st, 2nd, and 3rd year Computer Science students, aiming to understand the impact of ESP assessments on language skills development. Moreover, the qualitative data significantly contributed to the creation of survey items that prompted learners to rate their agreement with given statements, allowing a deeper understanding of the complex relationship between ESP assessments and language skills. Similarly, the insights gained from the first phase guided the formulation of two distinct tests - the "generic" and "ICL" tests. While the "generic" test mirrors standardized approaches with limited feedback provisions, the "ICL" test, rooted in contemporary ESP and ICL theories, integrating professional market needs into its assessment framework.

By synthesizing these qualitative findings into tangible elements of the quantitative design, this research not only illuminates the existing deficiencies within ESP assessments but

## ***CHAPTER THREE Data Analysis and Discussions***

---

also endeavors to pave the way for more authentic, fair, and aligned evaluative practices. The iterative and integrative nature of this approach marks a significant stride toward understanding the intricate dynamics between assessment designs and language skills development in the context of Computer Science ESP education.

### **3.4 Quantitative Phase**

This subsequent quantitative phase focused on collecting data through a survey and a test. This phase aimed to address the second research question related to learners' perceptions of ESP final exams and their experiential feedback concerning its impact on the growth of language and topical knowledge and skills: How do learners experience ESP assessment?

#### ***3.4.1 Learners' Survey***

The second phase started by administering a survey to undergraduate students majoring in Computer Science, encompassing those in their first (1st), second (2nd), and third (3rd) years of study. The principal aim was to investigate students' perspectives on ESP final exams, with a specific emphasis on gathering their experiential feedback regarding how these exams contributed to the enhancement of their language proficiency, subject-specific knowledge, and skill development.

##### ***3.4.1.1 Findings***

After conducting a three-week online survey, responses from a total of 367 participants were received. The primary group of respondents comprised first-year Computer Science students, constituting 43% of the total with 157 students. Second-year Computer Science students followed, representing 31% of the participants, including 114 individuals. Lastly, 96 third-year Computer Science students, making up 26% of the total respondents, contributed to the survey.

## ***CHAPTER THREE Data Analysis and Discussions***

---

Each item of the survey rubrics is explained below, and the findings are initially presented in tables using systematically cleaned data. Owing to the online survey's structured format, no incomplete or invalid data needed identification or removal. Furthermore, to enhance data readability and comprehension, all data points were rounded to the nearest integer, considering the insignificance of decimal places in this context.

### ***3.4.1.1.1 Exam Procedures***

- **Item N°1:** “The instructions for completing the English exam are clear.”

This item asked learners to rate how clear and easy to understand the instructions for completing the English exam were. If learners give this item a high rating, it means that they were able to understand the instructions without any difficulty and were able to complete the exam as instructed. If learners give this item a low rating, it means that they found the instructions to be confusing or difficult to understand, and this may have made it difficult for them to complete the exam accurately. Findings from Item N°1 are presented in Table 3.1.

- **Item N°2:** “The English examination process is organized.”

This item asked learner to rate how well-organized the English exam process was. This includes things like the clarity of the exam schedule, the availability of seating, and the efficiency of the check-in and check-out procedures. If learners give this item a high rating, it means that they found the exam process to be smooth and efficient. If learners give this item a low rating, it means that they experienced problems with the exam process, such as long wait times, confusion about where to go, or difficulty finding a seat. Findings from Item N°2 are presented in Table 3.1.

- **Item N°3:** “The English exam invigilators are professional.”

## ***CHAPTER THREE Data Analysis and Discussions***

---

This item asked learners to rate how professional the English test invigilators were. This includes things like the invigilators' attitude, helpfulness, and fairness. If learners give this item a high rating, it means that they felt that the invigilators were doing their job well and that they were creating a positive and supportive environment for the test-takers. If learners give this item a low rating, it means that they had negative experiences with the invigilators, such as the invigilators being disruptive or unhelpful, or unfair in enforcing the exam rules. Findings from Item N°3 are presented in Table 3.1.

- **Item N°4:** “I am allotted enough time to complete the English exam.”

This item asked learners to rate whether they had enough time to complete the English exam. If learners give this item a high rating, it means that they felt that they had enough time to answer all of the questions and to review their answers before the end of the exam. If learners give this item a low rating, it means that they felt rushed or that they did not have enough time to complete the exam to the best of their ability. Findings from Item N°4 are presented in Table 3.1.

- **Item N°5:** “I have problems with the English exam materials.”

This item asked learners to report any problems they had with the test materials, such as missing pages and illegible input. If learners give this item a high rating, it means that they did not have any problems with the test materials. If learners give this item a low rating, it means that they had problems with the test materials that may have interfered with their ability to complete the exam accurately. Findings from Item N°5 are presented in Table 3.1.

**Table 3. 1** *Learners’ Satisfaction with the English Final Exam Procedures*

Items	Data	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Clarity	%	9%	40%	13%	34 %	4%

## *CHAPTER THREE Data Analysis and Discussions*

<b>Organization</b>	N°	33	148	47	124	15
	%	22%	52%	16%	9%	1%
<b>Professionalism</b>	N°	82	191	59	31	4
	%	28%	33 %	13%	15%	11%
<b>Duration</b>	N°	103	122	46	55	41
	%	37%	32%	4%	18%	9%
<b>Test Material Related Issues</b>	N°	135	117	14	68	33
	%	9%	17%	7%	50%	17%
	N°	31	61	26	184	63

### *3.4.1.1.2 Reading Skill Assessment and Development*

- **Item N°1:** “The exam assesses my ability to read and understand different types of texts, such as news articles, academic essays, and fiction.”

This item assesses whether the exam is representative of the types of texts that learners are likely to encounter in the real world. If learners feel that the exam is only assessing their ability to read and understand certain types of texts, they may not feel confident in their ability to apply their reading skills to a variety of situations. Findings from Item N°1 are presented in Table 3.2.

- **Item N°2:** “The exam assesses my ability to identify the main ideas and supporting details in a text.”

This item assesses whether the exam targets learners' ability to understand the key concepts and information presented in a text. If learners are unable to identify the main ideas and supporting details in a text, they will struggle to functional meanings the text accurately. Findings from Item N°2 are presented in Table 3.2.

- **Item N°3:** “The exam assesses my ability to draw inferences and conclusions from a text.”

This item assesses whether the exam tests learners' ability to go beyond the functional or literal meaning of a text and to make their own interpretations, negotiating implied meanings. If learners are unable to draw inferences and conclusions from a text, they will be limited in their ability to think critically about the text and to apply their

## **CHAPTER THREE Data Analysis and Discussions**

understanding of the text to new situations. Findings from Item N°3 are presented in Table 3.2.

- **Item N°4:** “The exam assesses my ability to identify and understand the author's purpose and point of view in a text.”

This item assesses whether the exam includes learners' ability to negotiate pragmatic meanings. If learners are unable to identify and understand the author's purpose and point of view, they may misinterpret the text or miss out on important nuances in the author's argument. Findings from Item N°4 are presented in Table 3.2.

**Table 3. 2Exam Design and Learners’ Reading Skill Development**

Items	Data	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
<b>Authenticity / Diversity</b>	%	10%	17%	42%	24%	7%
	N°	37	62	154	88	26
<b>Direct Functional Meanings</b>	%	17%	51%	11%	17%	4%
	N°	63	186	39	64	15
<b>Implied Functional Meanings</b>	%	7%	27%	36%	16%	14%
	N°	27	97	133	58	52
<b>Implied Pragmatic Meanings</b>	%	2%	5%	13%	52%	28%
	N°	7	19	47	193	101

### **3.4.1.1.3 Writing Skill Assessment and Development**

- **Item N°1:** “The exam assesses my ability to write clear and concise sentences and paragraphs.”

This item assesses whether the exam targets learners' ability to communicate their ideas effectively in writing. If learners are unable to write clear and concise sentences and paragraphs, their writing may be difficult to understand and meanings may not be communicated effectively. Findings from Item N°1 are presented in Table 3.3.

- **Item N°2:** “The exam assesses my ability to organize my thoughts and ideas in a logical way.”

This item assesses whether the exam tests learners' ability to write clear and well-structured essays. If learners are unable to organize their thoughts and ideas in a

## ***CHAPTER THREE Data Analysis and Discussions***

logical way, their writing may be disorganized and difficult to follow, again impeding meaning. Findings from Item N°2 are presented in Table .3.

- **Item N°3:** “The exam assesses my ability to use appropriate grammar and vocabulary.”

This item assesses whether the exam targets learners' writing skills and their ability to use language correctly. If learners are unable to use appropriate grammar and vocabulary, their writing may lack clarity and appropriateness. Findings from Item N°3 are presented in Table 3.3.

- **Item N°4:** “The exam assesses my ability to write in different genres, such as essays, reports, and letters.”

This item assesses whether the exam considers learners' versatility as writers. The ability to write in different genres is essential for success in many different types of writing, from academic essays to professional letters. Findings from Item N°4 are presented in Table 3.3.

**Table 3. 3Exam Design and Learners’ Writing Skill Development**

Items	Data	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
<b>Topical Control</b>	%	15%	37%	6%	8%	34%
	N°	55	135	21	29	127
<b>Rhetorical Control</b>	%	14%	38%	5%	6%	37%
	N°	51	141	17	21	137
<b>Language Accuracy</b>	%	28%	42%	22%	5%	3%
	N°	104	153	79	18	13
<b>Authenticity / Diversity</b>	%	3%	10%	21%	50%	16%
	N°	12	37	77	183	58

### ***3.4.1.1.4 Speaking Skill Assessment and Development***

- **Item N°1:** “The exam assesses my ability to speak clearly and fluently on a variety of topics.”

This item assesses whether the exam targets learners' ability to communicate their ideas effectively in a variety of situations. If learners are unable to speak clearly and

## ***CHAPTER THREE Data Analysis and Discussions***

fluently, they may have difficulty communicating their ideas in class, in job interviews, and in other social and professional settings. Findings from Item N°1 are presented in Table 3.4.

- **Item N°2:** “The exam assesses my ability to use appropriate grammar and vocabulary in speech.”

This item assesses whether the exam focuses on learners' speaking skills and their ability to use language correctly. If learners are unable to use appropriate grammar and vocabulary in speech, their communication may be unprofessional and difficult to understand. Findings from Item N°2 are presented in Table 3.4.

- **Item N°3:** “The exam assesses my ability to organize my thoughts and ideas in a logical way when speaking.”

This item assesses whether the exam targets learners' ability to deliver clear and well-structured presentations and speeches. If learners are unable to organize their thoughts and ideas in a logical way, their speaking may be disorganized and difficult to follow. Findings from Item N°3 are presented in Table 3.4.

- **Item N°4:** “The exam assesses my ability to interact with others in a conversation.”

This item assesses whether the exam targets learners' social skills and their ability to participate in conversations effectively. If learners are unable to interact with others in a conversation, they may have difficulty building relationships and working collaboratively with others. Findings from Item N°4 are presented in Table 3.4.

**Table 3.4** *Exam Design and Learners' Speaking Skill Development*

Items	Data	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
<b>Topical Control</b>	%	0%	0%	1%	38%	61%
	N°	0	0	5	138	224
<b>Language Accuracy</b>	%	0%	0%	1%	38%	61%
	N°	0	0	4	139	224
<b>Rhetorical Control</b>	%	0%	0%	1%	38%	61%



## CHAPTER THREE *Data Analysis and Discussions*

<b>Authenticity / Diversity</b>	<b>N°</b>	0	0	2	140	225
	<b>%</b>	0%	0%	0%	38%	62%
	<b>N°</b>	0	0	0	139	228

### 3.4.1.1.5 *Listening Skill Assessment and Development*

- **Item N°1:** “The exam assesses my ability to understand spoken language in a variety of contexts, such as lectures, conversations, and presentations.”

This item assesses whether the exam focuses on learners' ability to understand spoken language in a variety of real-world situations. If learners are unable to understand spoken language in a variety of contexts, they may have difficulty following lectures, participating in conversations, and understanding presentations in academic or professional situations. Findings from Item N°1 are presented in Table 3.5.

- **Item N°2:** “The exam assesses my ability to identify the main ideas and supporting details in a spoken passage.”

This item assesses whether the exam considers learners' ability to follow the main thread of a spoken passage and to understand the key points that are being made. If learners are unable to identify the main ideas and supporting details in a spoken passage, they may have difficulty remembering the information that they have heard and they may not be able to construct knowledge based on listening. Findings from Item N°2 are presented in Table 3.5.

- **Item N°3:** “The exam assesses my ability to draw inferences and conclusions from a spoken passage.”

This item assesses whether the exam targets learners' ability to go beyond the direct functional meanings of a spoken passage and to make their own interpretations. If learners are unable to draw inferences and conclusions from a spoken passage, they may miss out on important nuances in the spoken message and they may not be able to apply their understanding of the passage to new situations. Findings from Item N°3 are presented in Table 3.5.

## ***CHAPTER THREE Data Analysis and Discussions***

- **Item N°4:** “The exam assesses my ability to follow instructions and complete tasks based on spoken information.”

This item assesses whether the exam tests learners' ability to understand and follow spoken instructions. If learners are unable to follow instructions and complete tasks based on spoken information, they may have difficulty succeeding in academic and professional settings where they are required to follow verbal instructions. Findings from Item N°4 are presented in Table 3.5.

**Table 3. 5***Exam Design and Learners' Listening Skill Development*

Items	Data	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
<b>Authenticity / Diversity</b>	%	0%	0%	0%	35%	65%
	N°	0	0	2	127	238
<b>Direct Functional Meanings</b>	%	0%	0%	0%	35%	65%
	N°	0	0	1	125	241
<b>Implied Functional Meanings</b>	%	0%	0%	0%	35%	65%
	N°	0	0	1	125	241
<b>Implied Pragmatic Meanings</b>	%	0%	0%	0%	31%	69%
	N°	0	0	1	113	253

### ***3.4.1.1.6 Language Skills in Specific Domain***

- **Item N°1:** “The English exam prepared me for using language in Computer Science academic domain such as reading papers, following an online course, or participating to a conference.”

This item assesses whether the exam targets learners' ability to use language effectively in Computer Science academic settings. This includes being able to read and understand research papers, follow online courses, and participate in conferences. Findings from Item N°1 are presented in Table 3.6.

- **Item N°2:** “The English exam prepared me for using language in Computer Science professional domain such as writing a professional document, presenting a solution, and eliciting needs.”

This item assesses whether the exam tests learners' ability to use language effectively in Computer Science professional settings. This includes being able to write

## ***CHAPTER THREE Data Analysis and Discussions***

professional documents, analyze problems, and present solutions. Findings from Item N°2 are presented in Table 3.6.

- **Item N°3:** “The English exam prepared me for using language in Computer Science socio-professional domain such as interacting with colleagues, interacting with customers, and freelancing.”

This item assesses whether the exam fostered learners' ability to use language effectively in Computer Science socio-professional settings. This includes being able to interact with colleagues, customers, and other stakeholders. Findings from Item N°3 are presented in Table 3.6.

- **Item N°4:** “The English exam helped me to achieve my domain-related language learning goals.”

This item assesses learners' overall satisfaction with the exam's ability to prepare them for their language learning goals in the Computer Science domain. Findings from Item N°4 are presented in Table 3.6.

**Table 3. 6***Exam Design and Language Skills in Domain-specific Situations*

Items	Data	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
<b>Academic TLU Domain</b>	%	0%	0%	8%	25%	67%
	N°	0	1	29	93	244
<b>Professional TLU Domain</b>	%	0%	1%	7%	23%	69%
	N°	0	5	27	84	251
<b>Social-Interactional TLU Domain</b>	%	0%	0%	5%	24%	71%
	N°	0	0	18	89	260
<b>Domain-related Language Learning Goals</b>	%	0%	4%	9%	31%	56%
	N°	0	16	33	115	203

### ***3.4.1.2 Discussion of the Findings***

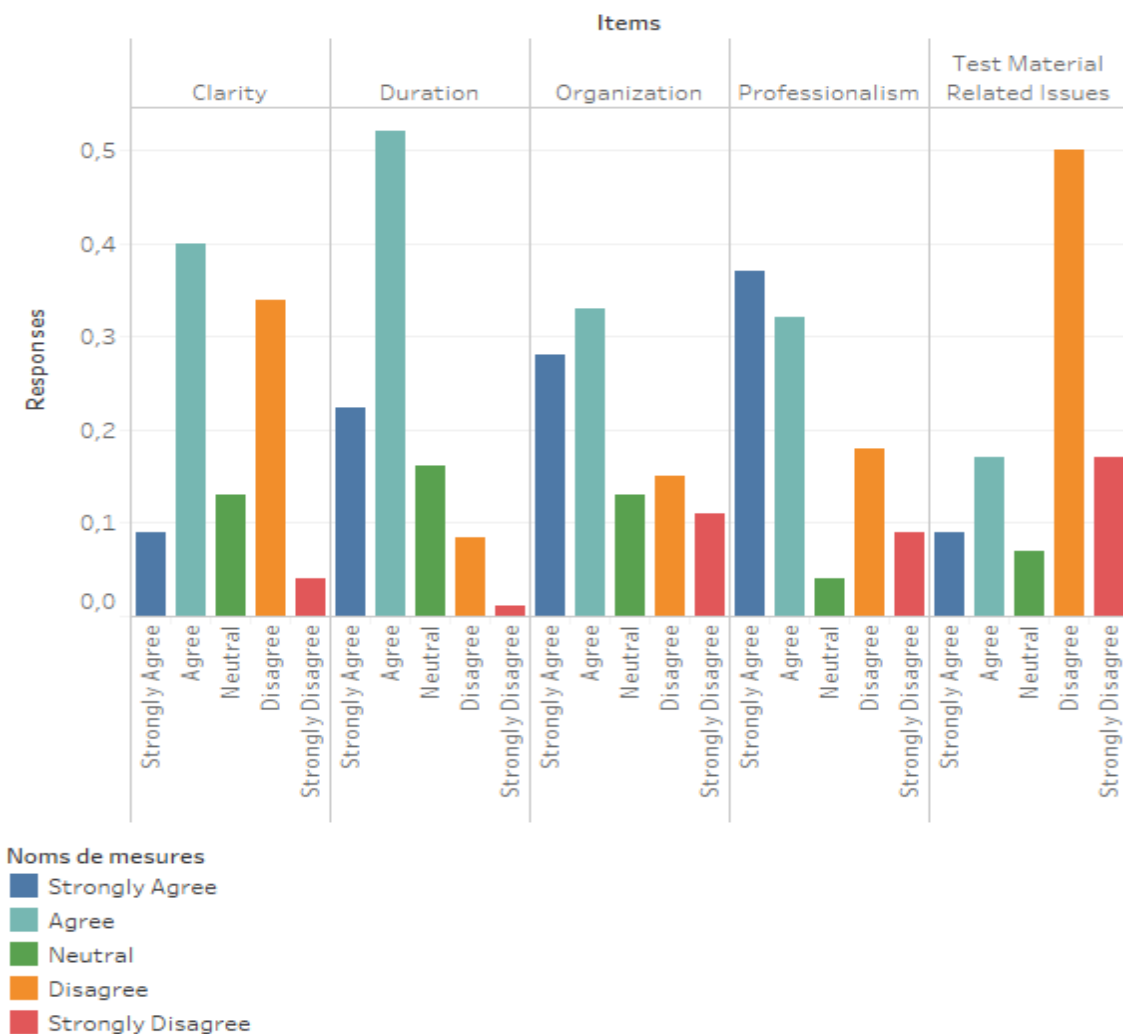
The survey data reveal that the overall procedures for the English Final Exam exhibit a rather high degree of effectiveness and practicality. Across three distinct academic levels, the majority of participants did not report significant negative experiences concerning the administration of the exam. Nevertheless, the data underscores concerns regarding learners'

## **CHAPTER THREE Data Analysis and Discussions**

comprehension of instructions, which play a pivotal role either in facilitating or impeding task completion and overall performance (Table 3.1; Figure 3.1). This points toward potential issues in assessment design and task specifications, possibly indicating a reliance on standardized task formats among learners.

**Figure 3. 1** Learners' Satisfaction with Exam Procedures

### Learners' Satisfaction with Exam Procedures



Approximately a quarter of the participants highlighted concerns related to exam organization and invigilation, suggesting a potential adverse impact on the psychological state of test-takers. Most respondents conveyed that the English Final Exam does not operate as a speed

## *CHAPTER THREE Data Analysis and Discussions*

---

test, emphasizing that respondents believe they receive sufficient time to complete the exam. Furthermore, the findings indicate that the test materials themselves did not significantly impede participants' ability to accurately finish the exam. The essence of any successful exam lies in the execution of its procedures. Therefore, it is plausible to infer that the administration's procedures represent the most practical and achievable strategy, given the contextual realities influencing the exam, such as legal stipulations, the student population, and the availability of ESP practitioners.

The survey findings illuminate a significant shortfall in addressing learners' receptive language skills. A considerable portion of respondents exhibited genuine uncertainty regarding whether the English Final Exam adequately evaluates their capacity to comprehend various types of written texts (Table 3.2). This indecision might stem from a lack of information, hindering the formation of a decisive opinion. Conversely, respondents did not avoid disclosing their true opinion on the assessment of learners' ability to comprehend spoken language in diverse contexts (Table 3.5). Strikingly, not a single respondent conveyed a positive opinion regarding the assessment of listening skills. This underscores the gravity of the issue, prompting strong opinions among the participants (Table 3.5).

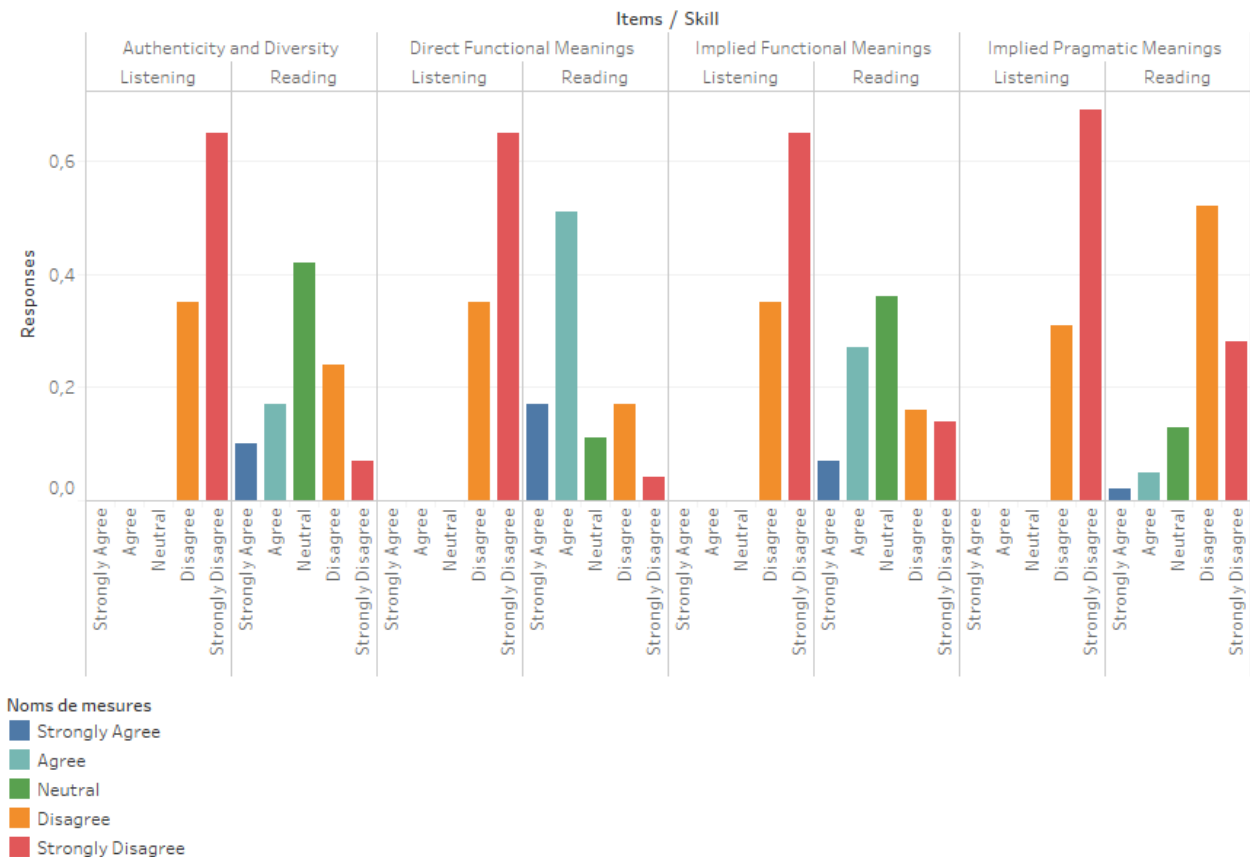
Beyond the absence of any kind of listening assessment, the results indicate that the evaluation of reading skills predominantly emphasizes the interpretation of direct, functional (literal) meanings within texts, encompassing main ideas and supporting details. This focus suggests significant limitations in both assessing and supporting learners' reading capabilities, as it requires a sole concentration on the surface structure of written discourse to complete comprehension tasks. Moreover, participants did not uniformly concur regarding the assessment of their ability to make inferences, identify differing viewpoints, and synthesize conclusions based on their reading experiences (Table 3.2).

## **CHAPTER THREE Data Analysis and Discussions**

Figure 3.2, below, juxtaposes the outcomes derived from the third rubric (focused on reading skill assessment and development) and the sixth rubric (pertaining to listening skill assessment and development). This juxtaposition reveals a consistent pragmatic approach guiding ESP assessment design in the 2<sup>nd</sup> and 3<sup>rd</sup> year Computer Science as well. The absence of listening assessments and the ineffective consideration of learners' reading skills both contribute to a substantial negative impact, particularly when compounded by a traditional testing cycle occurring four times annually.

**Figure 3. 2Learners' Receptive Skills Assessment**

### **Learners' Receptive Skills Assessment**



Furthermore, the research findings also underscore the inadequacies in addressing learners' productive language skills. The feedback polarity evident in the initial item of the fourth rubric (focused on writing skill assessment and development), as shown in Table 3.3, highlights a consistent failure in ESP Final Exams to effectively evaluate students' ability to construct clear

## ***CHAPTER THREE Data Analysis and Discussions***

---

and concise sentences and paragraphs. Participants have noted that the emphasis in writing tasks leans more toward assessing grammatical accuracy and vocabulary usage rather than fostering skills in diverse genres such as essay and report writing, or the organization of ideas and arguments in a coherent manner. These insights significantly reflect the underlying assessment design, which appears heavily reliant on item-based and selected-response task formats.

The consensus among participants strongly indicates that the utilization of learners' writing skills predominantly revolves around showcasing grammatical and lexical knowledge. This narrow focus overlooks learners' proficiency in constructing sentences, let alone higher-level tasks like essays and other writing genres. Consequently, learners are primarily tasked with filling in correct grammatical forms and vocabulary, presenting a significant setback in both assessing and cultivating their writing skills, which subsequently impacts their ability to develop proficient writing abilities.

Participants did not hold back in expressing their views on the incorporation of learners' ability to orally articulate thoughts clearly and fluently across a range of topics (Table 3.4). Strikingly, not a single respondent reported on the inclusion of speaking assessments in ESP Final Exams. The lack of neutral opinions further indicates that participants consider this issue important enough to elicit a strong collective opinion (Table 3.4). Additionally, the results reveal that neither second-year nor third-year students' speaking skills are subject to assessment, and, therefore, to motivated and self-initiated development.

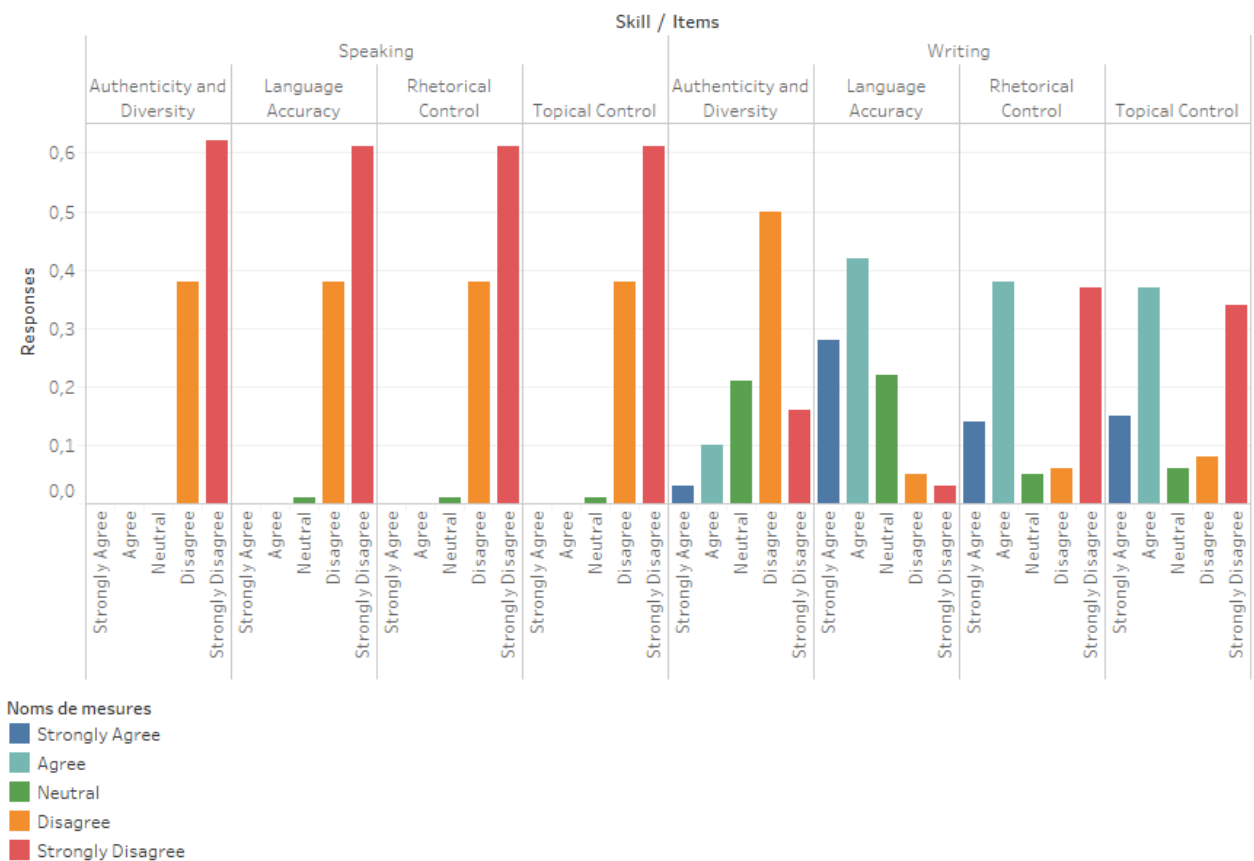
Figure 3.3 draws a link between the outcomes derived from the fourth rubric (concerning writing skill assessment and development) and the fifth rubric (pertaining to speaking skill assessment and development). This juxtaposition indicates a somewhat parallel understanding guiding ESP practitioners, underpinned by pragmatic considerations in ESP assessment, alongside influential contextual factors. The lack of speaking assessments and the limited

## *CHAPTER THREE Data Analysis and Discussions*

evaluation of learners' writing skills both contribute significantly to a strong negative washback, especially when compounded by the same recurrent and formal testing tradition.

**Figure 3. 3Learners' Productive Skills Assessment**

**Learners' Productive Skills Assessment**



The study findings were also used to gauge the intensity of respondents' opinions on the use of language skills within the Computer Science domain and the extent to which these skills are addressed by ESP assessments (Figure 3.4; Table 3.6). Participants conveyed that the current



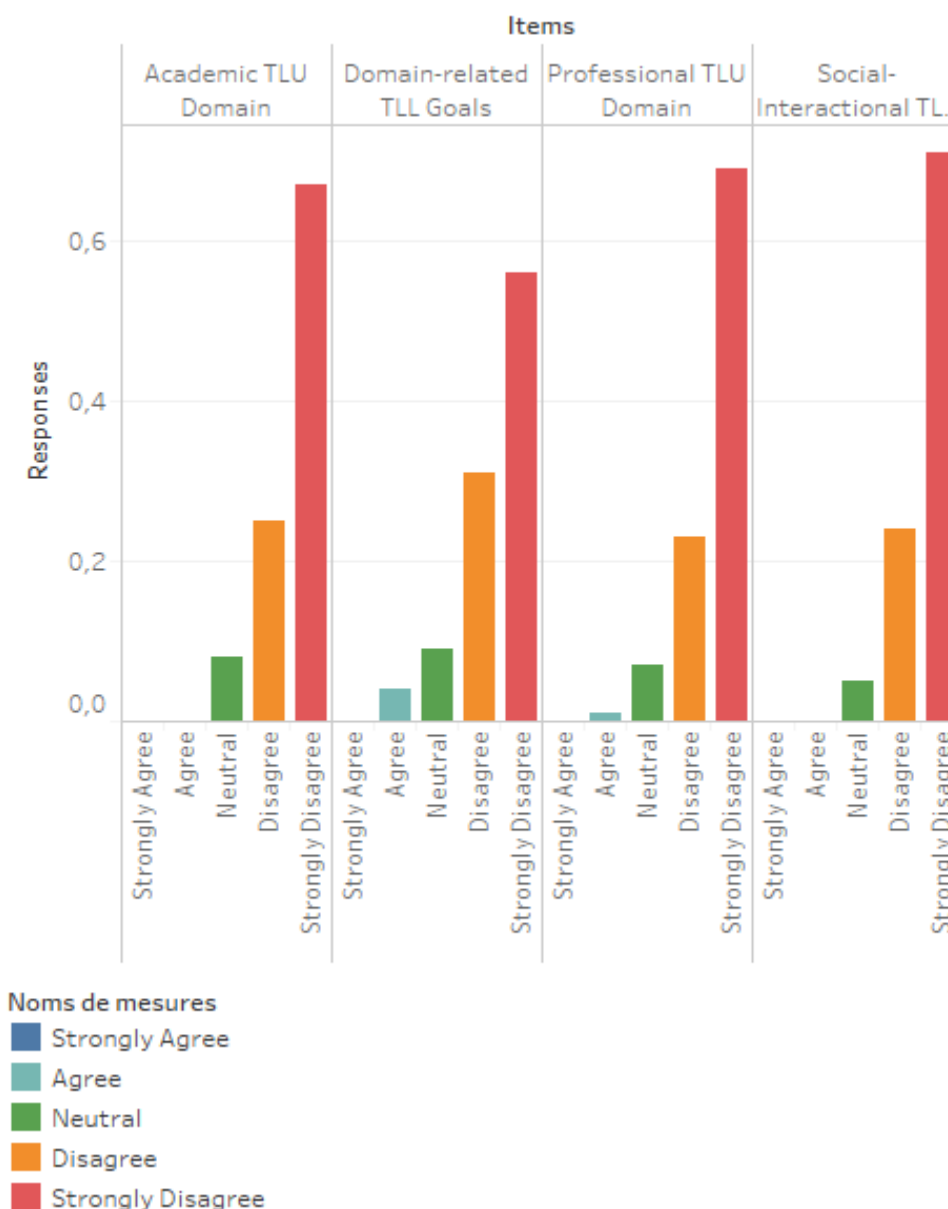
## ***CHAPTER THREE Data Analysis and Discussions***

---

ESP Final Exams inadequately equip them for using language within the academic domain of Computer Science. This deficiency suggests a lack of focus on fostering learners' capabilities in comprehending research papers, engaging with online courses, and participating effectively in academic conferences.

**Figure 3. 4***Learners' Satisfaction and Domain-related Linguistic Needs*

**Learners' Satisfaction and Domain-related Linguistic Needs**



Furthermore, the ESP Final Exams were shown to neglect preparing learners for using language within the professional domain of Computer Science as well. Proficiencies required for writing professional documents like user guides and presenting software solutions appear to be largely marginalized in the assessment design process. Additionally, learners are not adequately prepared for using language within the socio-professional realm of Computer Science, which involves training and assessment based on the quality of interactions with both colleagues and customers.

## ***CHAPTER THREE Data Analysis and Discussions***

---

The scarcity of positive and neutral opinions in this specific rubric underscores the depth of consideration given by respondents to both sides of the issue. This basically indicates a significant misalignment between ESP assessment and learners' actual needs. A considerable majority of respondents (87%) expressed a consensus that ESP assessment did not effectively contribute to their achievement of language learning goals within their specific domain. This discrepancy implies that students feel compelled to revise and prepare for exams that do not address their linguistic and topical needs and objectives. Such a scenario significantly impacts learners' motivation to engage in language learning and influences their priorities while preparing for final evaluations. This impact is particularly intensified as second and third-year ESP assessments share similar shortcomings.

### ***3.4.1.3 Summary***

The comprehensive analysis of survey data reveals a number of insights regarding the ESP assessment efficacy, alignment with learners' needs, and overall impact on language proficiency and domain-specific competencies. The analysis and discussion encompassed the examination of administration procedures, assessment of receptive and productive language skills, and alignment of exam content with the demands of the Computer Science domain.

Administering the survey across three academic levels provided a nuanced understanding of the strengths and limitations inherent in the Computer Science English Final Exam procedures. While the exams were generally regarded as effectively administered and non-time-pressured, concerns regarding learners' comprehension of instructions and the absence of a listening assessment were prominent. This shed light on potential issues in assessment design and the need for a more comprehensive approach to evaluating receptive language skills, particularly in reading and listening comprehension. The study also illuminated significant deficiencies in the assessment of learners' productive language skills. The narrow focus on

## ***CHAPTER THREE Data Analysis and Discussions***

---

grammatical accuracy and vocabulary usage in writing tasks neglected the development of intermediate to advanced writing competencies, such as essay writing and coherent argumentation. Moreover, the absence of speaking assessments indicated a remarkable oversight in evaluating and fostering oral communication skills across various topics, contributing to a substantial negative impact on learners' skill development.

Participants expressed a consensus that the current ESP assessments did not align with their specific language learning goals within the Computer Science domain. The lack of emphasis on comprehending research papers, engaging in professional communication, and interacting effectively in both academic and socio-professional contexts unveiled a significant misalignment between the assessment content and learners' actual needs. The findings, with an overwhelming majority (87%) of respondents expressing dissatisfaction with the alignment of the assessments with their language learning objectives, highlight the considerable impact on learners' motivation and examination preparation strategies. These discrepancies underscore a critical need for a more comprehensive and contextualized approach to ESP assessment design, one that aligns more effectively with the specialized language and communication requirements within the Computer Science domain. It is imperative for future assessment frameworks to address these identified shortcomings, considering the pivotal role language proficiency plays in the academic and professional development of Computer Science students.

### ***3.4.2 The Test***

The second phase continued with a testing procedure targeting first (1st) year students majoring in Computer Science (I). This procedure included a generic test, based on the qualitative findings, and an ICL test based on the LOA framework (Turner & Purpura, 2016) and the MOM of L2 proficiency (Purpura & Dakin, 2020). The principal aim was to offer more insight into the impact these two assessment approaches on learners' experience and

## ***CHAPTER THREE Data Analysis and Discussions***

---

performance, facilitating the discussion regarding their impact on language proficiency, subject-specific knowledge, and skill development.

### **3.4.2.1 Findings**

The participation of 76 first-year Computer Science students was recorded, as estimated previously on the registration list. The students were divided into two groups, with 38 students each: one group took the Generic Test, and the other group took the ICL Test. Further subdivision resulted in two subgroups within each main group, consisting of 19 students each. Notably, there were no absences, and the attendance matched the initial registration list, with no additional students beyond the recorded ones.

The detailed data derived from the ICL and Generic Tests for analysis and discussion is presented in Table 3.7 and Table 3.8 below. This information encompasses students' scores in reading comprehension (RC) tasks, grammar tasks, algorithms through reading tasks (Algo RC), and ICL writing tasks. Additionally, the tables provide insights into the GTS, GTG, ITS, and ITS scores achieved by the students. Considering the pre-established conditions of the experiment, no incomplete or invalid data needed identification or removal. Furthermore, to enhance data readability, all GTG and ITG data points were expressed using only two decimals, considering the insignificance of the other decimal places in this context.

**Table 3. 7ICL Test Students' Scores**

<b>Student</b>	<b>RC IT (* / 5)</b>	<b>GR IT (* / 10)</b>	<b>Algo RC</b>	<b>ICL WT IT (* / 15)</b>	<b>ITS (* / 35)</b>	<b>Net ITS</b>	<b>ITG (* / 20)</b>	<b>Net ITG (* / 20)</b>
----------------	--------------------------	---------------------------	--------------------	-------------------------------	-------------------------	--------------------	-------------------------	-----------------------------

## *CHAPTER THREE Data Analysis and Discussions*

	(* /5)				(* /20)			
1I	5	9	5	9	28	23	16,00	15,33
2I	5	8	5	8	26	21	14,86	14,00
3I	5	7	5	9	26	21	14,86	14,00
4I	4	5	5	6	20	15	11,43	10,00
5I	5	8	4	8	25	21	14,29	14,00
6I	4	6	4	7	21	17	12,00	11,33
7I	5	9	3	9	26	23	14,86	15,33
8I	5	10	5	11	31	26	17,71	17,33
9I	3	4	4	7	18	14	10,29	9,33
10I	5	9	5	10	29	24	16,57	16,00
11I	5	8	5	8	26	21	14,86	14,00
12I	5	9	5	9	28	23	16,00	15,33
13I	5	10	5	10	30	25	17,14	16,67
14I	5	7	5	8	25	20	14,29	13,33
15I	5	8	4	8	25	21	14,29	14,00
16I	4	6	4	7	21	17	12,00	11,33
17I	5	9	5	11	30	25	17,14	16,67
18I	5	7	4	9	25	21	14,29	14,00
19I	3	3	5	6	17	12	9,71	8,00
20I	5	10	5	13	33	28	18,86	18,67
21I	5	8	5	10	28	23	16,00	15,33
22I	4	6	4	8	22	18	12,57	12,00
23I	5	10	5	11	31	26	17,71	17,33
24I	5	7	5	9	26	21	14,86	14,00
25I	4	7	4	7	22	18	12,57	12,00
26I	5	7	5	9	26	21	14,86	14,00
27I	5	6	4	8	23	19	13,14	12,67
28I	4	6	3	8	21	18	12,00	12,00
29I	3	5	4	7	19	15	10,86	10,00
30I	5	8	4	9	26	22	14,86	14,67
31I	5	9	5	10	29	24	16,57	16,00
32I	5	8	5	10	28	23	16,00	15,33
33I	5	8	5	8	26	21	14,86	14,00
34I	5	10	5	11	31	26	17,71	17,33
35I	3	6	4	6	19	15	10,86	10,00
36I	5	7	5	9	26	21	14,86	14,00
37I	2	4	3	6	15	12	8,57	8,00
38I	5	10	5	11	31			
					26	17,71		17,33

**Table 3. 8** *Generic Test Students' Scores*

Student	RC (* /5)	Grammar (* /10)	ICL Writing Task (* /15)	GTS (* /30)	GTG (* /20)
---------	-----------	-----------------	--------------------------	-------------	-------------

## *CHAPTER THREE Data Analysis and Discussions*

1G	5	8	7	20	13,33
2G	3	6	5	14	9,33
3G	4	7	6	17	11,33
4G	5	10	8	23	15,33
5G	5	5	6	16	10,67
6G	4	5	4	13	8,67
7G	4	4	4	12	8,00
8G	5	8	7	20	13,33
9G	5	4	5	14	9,33
10G	4	6	6	16	10,67
11G	5	10	9	24	16,00
12G	2	2	3	7	4,67
13G	2	5	5	12	8,00
14G	4	3	5	12	8,00
15G	5	8	7	20	13,33
16G	4	6	6	16	10,67
17G	5	9	7	21	14,00
18G	4	4	5	13	8,67
19G	2	3	3	8	5,33
20G	5	9	8	22	14,67
21G	5	10	9	24	16,00
22G	3	4	4	11	7,33
23G	4	7	5	16	10,67
24G	3	2	4	9	6,00
25G	5	8	7	20	13,33
26G	4	5	5	14	9,33
27G	5	5	6	16	10,67
28G	5	9	8	22	14,67
29G	1	2	3	6	4,00
30G	2	3	5	10	6,67
31G	3	5	6	14	9,33
32G	3	3	5	11	7,33
33G	5	8	7	20	13,33
34G	4	9	7	20	13,33
35G	4	6	6	16	10,67
36G	3	3	5	11	7,33
37G	4	5	6	15	10,00
38G	5	10	8	23	15,33

Table 3.9 provides an overview of the data, presenting key statistical measures of central tendency such as the mean, the trimmed mean, the median, and the mode. These statistical summaries not only offer valuable insights into the distribution of data but also shed

## **CHAPTER THREE Data Analysis and Discussions**

light on additional variability characteristics. The table includes details on the standard deviation, variance, and range of the collected data.

**Table 3. 9** *Test Scores Central Tendency and Variability*

Measures	RCGTS	RCITS	GR GTS	GR ITS	Algo RC ITS	ICLWT GTS	ICLWT ITS	GTS	ITS	Net ITS	GTG	ITG	Net ITG
<b>Mean</b>	3,95	4,55	6,84	7,47	4,53	5,84	8,68	16,63	25,24	20,71	11,09	14,42	13,81
<b>10% Trimmed Mean</b>	4,00	4,61	6,89	7,56	4,56	5,83	8,64	16,72	25,31	20,75	11,15	14,46	13,83
<b>Median</b>	4	5	6,5	8	5	6	9	16	26	21	10,67	14,86	14,00
<b>Mode</b>	5	5	6	8	5	5	9	15	26	21	10	14,86	14
<b>Standard Deviation</b>	1,11	0,80	2,02	1,84	0,65	1,59	1,66	4,43	4,39	4,03	2,95	2,51	2,68
<b>Variance</b>	1,24	0,63	4,08	3,39	0,42	2,51	2,76	19,59	19,27	16,21	8,71	6,29	7,20
<b>Range</b>	4	3	9	7	2	3	7	18	18	16	12	10,29	10,67

The Generic Test results reveal that the mean scores for GTS and GTG converge toward the average, resulting in a 68% success rate (Table 3.9; Table 3.8). Despite a high number of students passing the Generic Test, their results are predominantly concentrated around the average test score and grade. In contrast, the ICL Test results demonstrate that the mean scores for ITS and ITG are significantly higher, converging toward the maximum achievable score and grade (Table 3.9; Table 3.7). Notably, the success rate is 94% for students who took the ICL Test, which indicates a better performance.

The Net ITS and Net ITG represent ICL Test scores and grades, excluding data from the third task of the ICL Test, thereby balancing the maximum achievable scores and grades in both tests. While the Net ITS and Net ITG means are slightly lower than ITS and ITG, they remain significantly higher than GTS and GTG means (Table 3.9). This suggests that the substantial difference between the means of the pairs (GTS, ITS) and (GTG, ITG) is not solely linked to the number of tasks per test, but is influenced by more impactful factors.



## ***CHAPTER THREE Data Analysis and Discussions***

---

The use of the trimmed mean, which removes 10% of the smallest and largest values from the dataset, does not reveal a significant difference between the mean and the 10% trimmed mean, indicating the absence of outliers (Table 3.9). The almost equal values for the mean, 10% trimmed mean, median, and mode of GTS imply a symmetrical distribution of scores. This symmetry is also observed in the case of ITS and Net ITS, where the four measures of central tendency (Table 3.9) exhibit minimal differences.

However, significant differences emerge when comparing the measures of central tendency for GTS, ITS, and Net ITS. For example, the GTS median indicates that 50% of students scored below 16 out of 30, while the Net ITS median reveals that 50% of students scored above 21 out of 30. Similarly, the GTS mode indicates that 15 is the most frequent score, whereas the Net ITS mode shows the highest occurrence at 21. These findings clearly highlight better performance by students who took the ICL Test, indicating the previously mentioned factors.

Table 3.9 also provides insights into the dispersion of the data. The standard deviation indicates a relatively small dispersion in the dataset. Both the standard deviation and variance suggest a symmetrical distribution of data, with students scoring close to the means of GTS, ITS, and Net ITS. The ranges indicate a good spread of values in GTS, ITS, and Net ITS, suggesting that the values are not closely clustered together, indicating a high level of variability. The same analytical approach has been applied to examine the distribution and dispersion of data related to students' scores in each individual task constituting the Generic and ICL tests.

### **3.4.2.2 Discussion of the Findings**

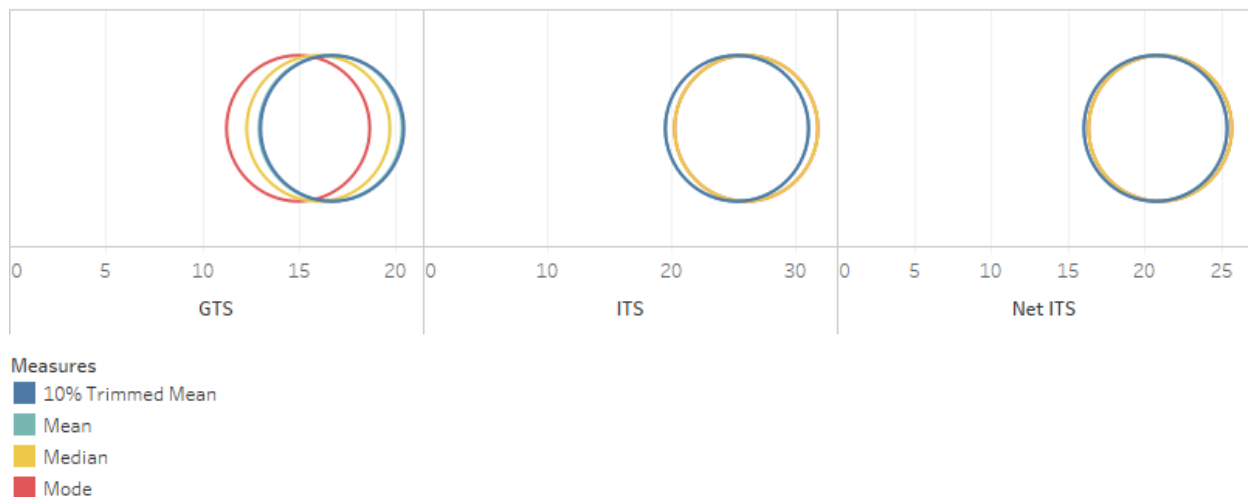
The mean, 10% trimmed mean, median, and mode of Net ITS exhibit a very close proximity, indicating a perfect symmetrical distribution of scores. This pattern extends to ITS and GTS, but with less symmetry. Figure 3.5 illustrates the potential of the ICL Test to

## CHAPTER THREE Data Analysis and Discussions

homogenize learners' collective performance around the mode (the score with the highest frequency). The Generic Test, on the other hand, failed at homogenizing the performance, and so the manifested proficiency level, of the test-takers. GTS mode, median, and mean are a little bit misaligned, indicating that the Generic Test did not have any impact on learners' initial language proficiency level. The distribution measures of the Generic Test are less symmetrical than the ICL Test. Nevertheless, this cannot be interpreted as a total asymmetry, which would have meant high heterogeneity among test-takers' initial language proficiency level.

**Figure 3. 5** Symmetrical Distribution of GTS, ITS, and Net ITS

**Symmetrical Distribution of GTS, ITS, and Net ITS**

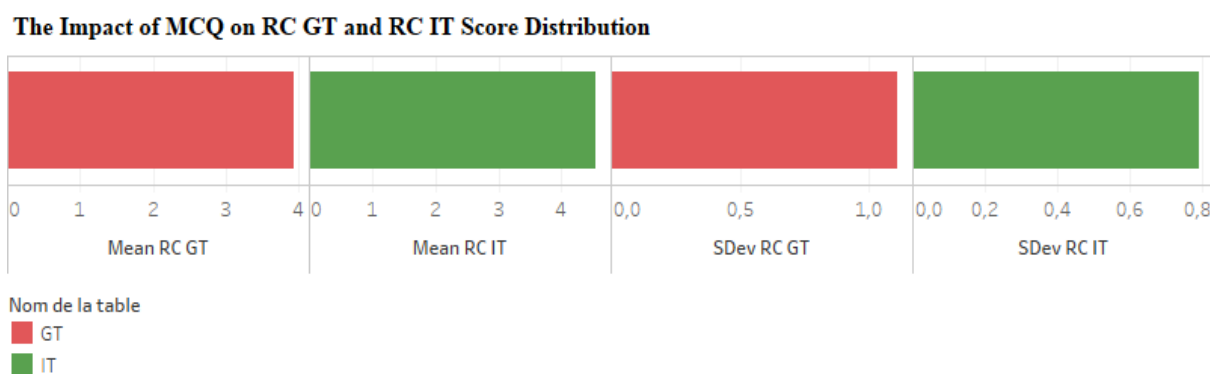


Notably, when delving into measures of central tendency for individual tasks such as reading comprehension (RC) and grammar (GR), nuances emerge. For instance, the mean, median, and mode of RC scores in both tests (RC GTS and RC ITS) demonstrate minimal differences, pointing to homogeneity in reading comprehension levels. Conversely, a subtle advantage is noted for the ICL test group in grammar tasks (GR GTS and GR ITS), despite the tasks being nearly identical. Importantly, this difference does not significantly impact the standard deviation and range for GR GTS and GR ITS, as detailed in Table 3.9.

## CHAPTER THREE Data Analysis and Discussions

The distribution measures for RC IT are slightly higher than RC GT (Figure 3.6), indicating improved performance potentially linked to the reading comprehension task format. The item-based (MC) reading comprehension task, reflected in RC IT mean, median, and mode, yielded higher scores. Furthermore, the RC GT and RC IT ranges (Table 3.9) highlight that task format and overall design contribute to error reduction. Open-ended questions, involving copying and pasting from the text, may trigger disengagement, especially when learners are uncertain about being held accountable for grammatical mistakes and typos. In contrast, Multiple-Choice Questions (MCQs) afford greater control for test designers, facilitating scoring procedures, and encourage deeper engagement by requiring students to negotiate meaning with distractors.

**Figure 3. 6** *The Impact of MCQ on RC GT and RC IT Score Distribution*



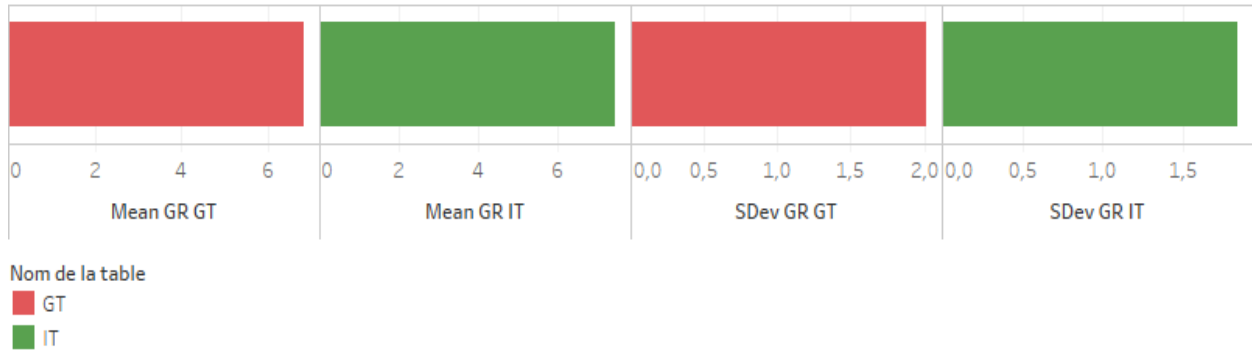
However, the impact of MCQs differed in grammar tasks (GR GT and GR IT), where the ICL Test group outperforms the Generic Test group, even with nearly identical tasks (Figure 3.7). Given the LOA core of the ICL Test, this performance advantage in GR IT is presumed to result from prior enablement. Test-takers in the ICL Test received feedback after the first task (RC IT), providing an opportunity to check answers and enhance understanding. Furthermore, the GR IT task, titled *"The Language of Requirements Questionnaire,"* clarifies that it encompasses meaning in the context of requirements elicitation, beyond mere grammar. In contrast, the GR GT task, titled *"Language Mastery,"* lacks specificity about the grammatical

## CHAPTER THREE Data Analysis and Discussions

task's context, failing to activate learners' schemata. These considerations contribute to enabled and engaged ICL Test-takers who ultimately outperformed their Generic Test counterparts.

**Figure 3.7** The Impact of MCQ on GR GT and GR IT Score Distribution

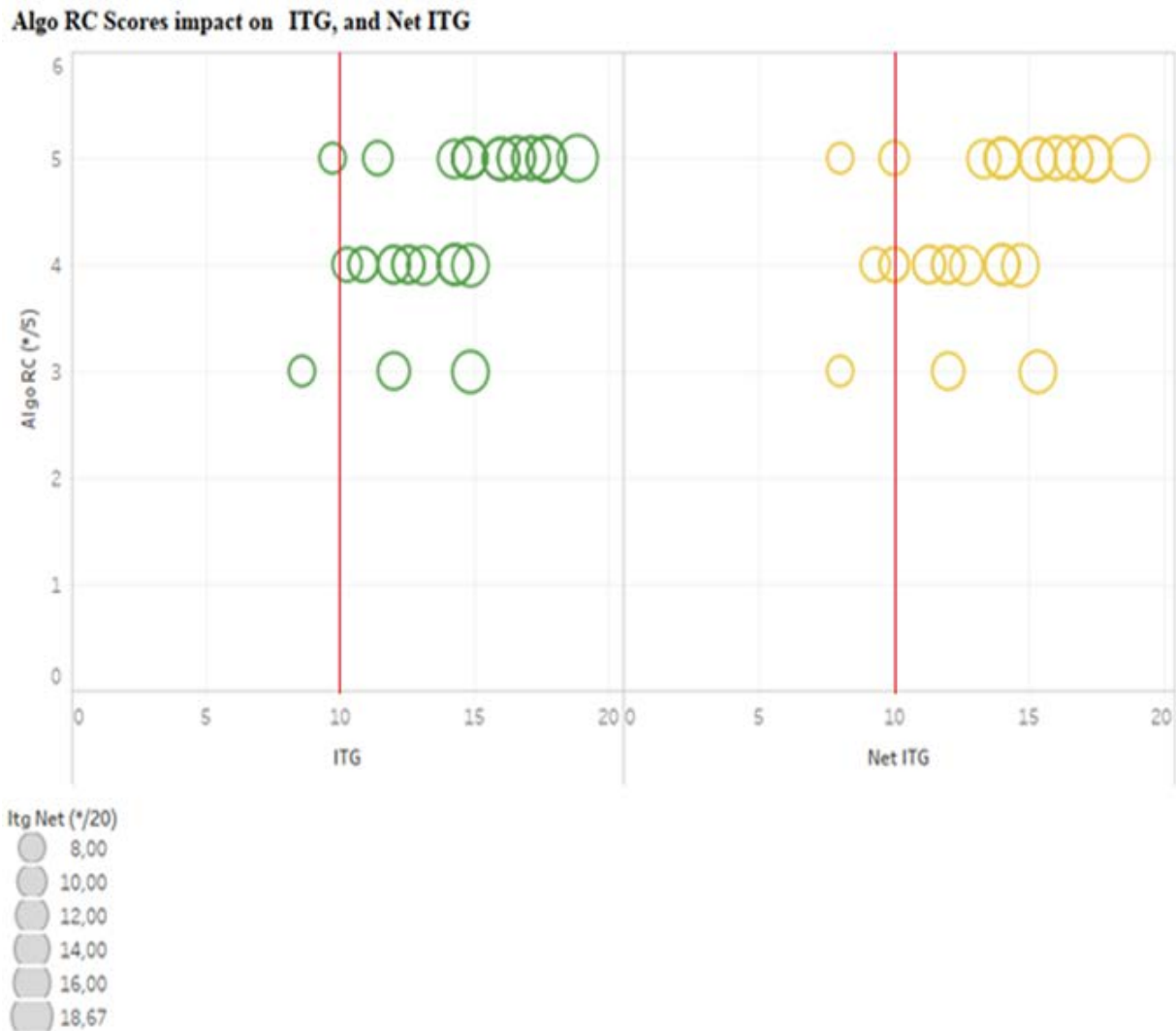
**The Impact of MCQ on GR GT and GR IT Score Distribution**



The ICL Test, characterized by an ICL task prioritizing topical knowledge over language, specifically targeting topical knowledge through reading in an item-based MCQ format (Table 9), yielded positive impacts on exam scores and grades (Figure 3.8). The case of *Student9I* shows that some learners' future depends on the integration of topic-oriented task. The ITG of *Student9I* is over the average (10,29), whereas his Net ITG is below it (9.33). This student was on the verge of failing the test, but managed to pass. He had poorly performed in RC and Grammar, but had better performed in the ICL ALGO RC task. His performance in the culminating ICL writing task shows progress, indicating the potential of topical integration in regenerating learners' motivation.

Comparing data distribution and dispersion between Generic Test and ICL Test results not only highlights superior performance in the latter but also reveals significantly higher scores in the ICL Writing Task (ICL WT IT) compared to the Generic Test (ICL WT GT) (Table 3.9). This notable difference, evident in the substantial disparity in data distribution between ICL WT GT and ICL WT IT, is strongly correlated with the test design featuring feedback provisions.

**Figure 3. 8** Algo RC Scores impact on GTG, ITG, and Net ITG

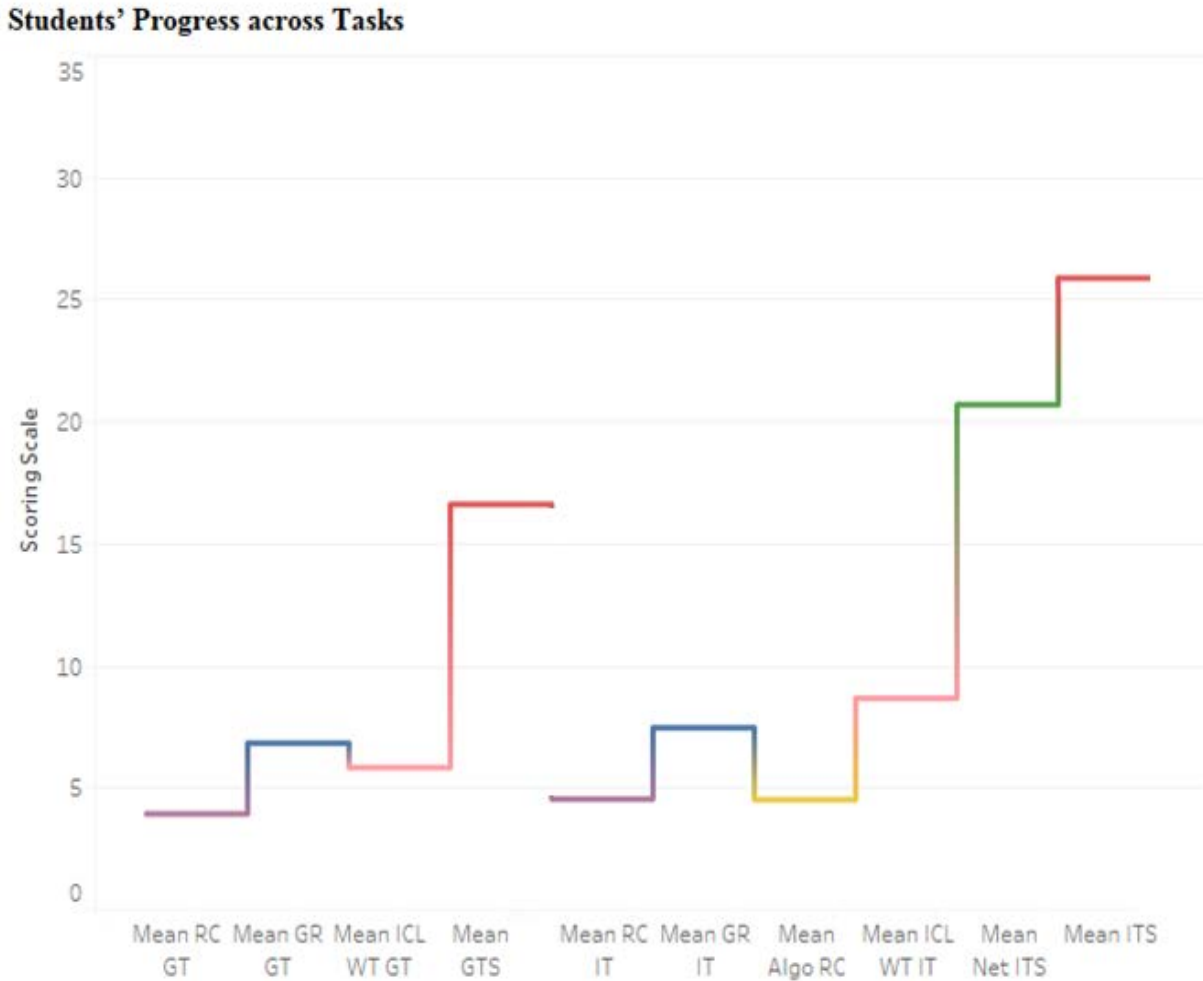


As ICL Test-takers rectify errors, learn from mistakes, and progress through subsequent tasks, feedback guides their improvement and skill development. The data indicates that several ICL test-takers demonstrated improvement, while Generic Test-takers exhibited the opposite trend. The Generic Test overlooked various performance moderators, including learner engagement and motivation, in addition to feedback. Moreover, the Generic Test lacked prompts and guidelines that were taken into consideration in the ICL Test design phase within the instructional dimension. The results also show that more students have passed the ICL Test

## CHAPTER THREE Data Analysis and Discussions

(94%). Figure 3.9, below, depicts students' progress across tasks in both the Generic and ICL Tests.

**Figure 3.9** Students' Progress across Tasks



The integration of assessment into Computer Science discourse practice, aligned with the Language of Assessment (LOA) framework and the Model of L2 Proficiency (MOM), coupled with assessment standards, significantly enhances the assessment experience and impact. The ICL Test notably benefits learners, prompting them to mobilize and apply both topical and linguistic knowledge, skills, and abilities (KSAs) to create and evaluate their requirements elicitation questionnaire. The data distribution of ITS and Net ITS significantly differs from that

## ***CHAPTER THREE Data Analysis and Discussions***

---

of GTS, underscoring the strong positive impact of the ICL Test on learners' mobilization and development of topical and linguistic skills. Implementing learning-oriented ICL tasks in ESP contexts motivates learners to mobilize and develop relevant linguistic and topical skills, therefore, boosting their performance, learning, and grades.

### ***3.4.2.3 Summary***

The analysis of the Generic Test and ICL Test results provides valuable insights into the performance and distribution of scores among English language learners. The Generic Test exhibits a concentration of mean scores around the average, resulting in a 68% success rate. In contrast, the ICL Test demonstrates significantly higher mean scores, converging toward the maximum achievable score and grade, with an impressive 94% success rate. The examination of Net ITS and Net ITG, representing scores excluding the third task of the ICL Test, highlights that their means, although slightly lower than ITS and ITG, remain notably higher than GTS and GTG means. This suggests feedback, in addition to the integrated design of the ICL Test, a pivotal role in influencing the observed differences in means.

Further analysis of data dispersion and task-specific performance reveals that the ICL Test, characterized by a focus on topical knowledge through reading, positively influences scores and grades. The integration of MCQs in reading comprehension tasks contributes to improved performance, while similar effects are not observed in grammar tasks. Feedback provisions and task-specific design in the ICL Test, compared to the Generic Test, contribute to the enhanced engagement and performance of ICL Test-takers.

Moreover, the ICL Test demonstrates a strong positive impact on learners' topical and linguistic skill mobilization and development. The success of the ICL Test is attributed not only to the LOA framework and the MOM of L2 Proficiency but also to

## ***CHAPTER THREE Data Analysis and Discussions***

---

careful adherence to assessment standards and the consideration of various performance moderators. The data distribution of ITS and Net ITS significantly differs from that of GTS, underscoring the robust positive impact of the ICL Test on learners' skill development.

The comprehensive analysis of test results and performance measures provides a nuanced understanding of the effectiveness and impact of the ICL Test compared to the Generic Test. The findings emphasize the importance of thoughtful test design, feedback provisions, and learners-centeredness in shaping successful language assessments within the context of Computer Science discourse practice.

### ***3.4.3 Conclusion***

As stated, the assessment of 1st year computer science students in the ESP context at the Faculty of Sciences in Algeria is a complex and challenging task. The quantitative data collected in the second phase of this sequential mixed-methods design show that ESP assessment has a strong and negative washback on students' language learning experience. This negative washback extends to third-year Computer Science students who felt that English Final Exams did not support them in achieving their domain-related language learning goals. Moreover, assessment practices failed at preparing students to take on domain-related real-life situations in academic, professional, and social-interactive language use domains.

The ICL Test emerged as a pivotal and advantageous component within the assessment framework, as evidenced by the comprehensive analysis of its impact on learners' performance. The symmetrical distribution of scores, closely aligned means, and consistent proficiency levels demonstrated by the ICL Test group, as compared to the Generic Test, highlight its effectiveness in maintaining quality in formal evaluations. The nuanced exploration of specific tasks, such as reading comprehension and grammar, reveals the ICL Test's superiority, particularly in ICL writing task, where it outperforms the Generic Test. The strategic use of Multiple-Choice Questions (MCQs) in the ICL Test enhances student engagement, providing greater control for



both test designers and learners. Moreover, the incorporation of an item-based MCQ format, coupled with feedback provisions, not only reduces errors but also facilitates a deeper understanding of linguistic constructs.

The ICL Test's emphasis on an integrated approach, targeting both topical knowledge and language skills, is reflected in the superior performance and progress of its takers, underlining its positive impact on learners' mobilization and development of essential skills. The higher pass rate further accentuates the ICL Test's success in catering to diverse learning needs and promoting engagement, motivation, and improvement among students. Overall, the ICL Test stands out as a robust and innovative assessment tool, aligning with contemporary theories and models in language assessment, and significantly contributing to learners' holistic development in the context of Computer Science education.

### **3.5 Discussion of the Main Findings**

The assessment of English for Specific Purposes (ESP) among 1st Year Computer Science students at the Faculty of Sciences in Algeria is a crucial component of their academic and professional journey. The increasing importance of English in technical fields in Algeria, especially in Computer Science, aligns with the global demand for proficient professionals (Ouahmiche et al., 2017). This paradigm shift towards ESP courses responds to the growing need for English language skills tailored to the specific requirements of the field (Maarouf & Lamouri, 2022).

The qualitative and quantitative data, from both phases, provide insights into the evaluative practices employed by ESP practitioners, shedding light on their design, learners' experience, and impact. The challenges identified in this study echo broader concerns highlighted in the literature about ESP courses in Algerian universities. Notably, the lack of alignment between ESP courses and learners' professional needs is a recurring

## ***CHAPTER THREE Data Analysis and Discussions***

---

issue (Assassi, 2020; Akkar & Idri, 2021). The misalignment is intensified by the shortage of qualified ESP teachers, many of whom lack formal training in ESP teaching methodologies (Abbassi, 2022).

The assessments are predominantly summative, focusing on predetermined learning objectives, reflecting a common trend in ESP courses in Algeria (Benabdi, 2022; Benmoussat & Benmoussat, 2018). However, a critical issue emerges – the absence of feedback provision. The lack of formative assessment is particularly concerning, as it hinders students' ability to learn and improve. This aligns with the broader concern raised by Saraa (2023) about the lack of clear assessment procedures and the difficulty in measuring students' learning without them.

The thematic analysis identified a pattern of recurring tasks and input, suggesting a lack of diversity and freshness in instructional materials and evaluative tools. Standardization influenced by teachers' subjective beliefs leads to weak assessments that may not align with the specific needs of computer science students. This echoes concerns raised by Khadam (2023) about the lack of cooperation between ESP tutors and subject matter specialists, which can hinder the alignment of assessments with international job industry requirements. The lack of the formative and / or learning-oriented dimensions can impede students' learning and improvement, particularly when assessments do not inform the iterative design process and further decisions (Boubris & Bouabdallah, 2023; Boumediene & Hamzaoui – Elachachi, 2017).

Thematic analysis extracts crucial insights from assessment artifacts, exposing a pattern of redundancy and a lack of innovation. It identified ten key themes, highlighting challenges and issues within the assessment artifacts. Some notable themes were the influence of the assessment designers on the overall assessment construction process and the use of the English Baccalaureate Exam blueprint, potentially leading to misalignment between assessment content and the specific needs of computer science students. Educators lacking LAL will rely on practical traditions (e.g., Baccalaureate) and on pragmatic understandings of assessment. Moreover, the absence of listening and speaking constructs in assessments raises questions about

## ***CHAPTER THREE Data Analysis and Discussions***

---

the comprehensiveness and authenticity of the evaluations, a concern also noted by Benabdi (2022).

While thematic analysis indicates some revisiting of the design process in certain artifacts, concerns linger regarding the effective integration of topical and linguistic constructs, particularly in computer science-related assessments. The study challenges the repeated application of standardized assessment models without adaptation. While standardized testing is common, the study suggests that a one-size-fits-all approach may not be suitable for Computer Science ESP i.e., standardized models cannot seamlessly transfer across diverse educational contexts without blueprint design and customization.

The study's emphasis on the learning-oriented ICL Test as a positive outlier resonates with the recommendations proposed by Abbassi (2022) for diversifying assessment methodologies. Learning-oriented ICL assessments, unlike generic tests, incorporate instructional provisions such as feedback and align with established assessment standards, positively influencing learners' engagement, performance, and overall learning experience. This underscores the importance of tailoring up-to-date and standard-based assessments to meet the unique needs of computer science students, as suggested by Benabdi (2022) and Saraa (2023).

The challenges identified in this study echo broader concerns highlighted in the literature about ESP courses in Algerian universities. Notably, the lack of alignment between ESP courses and learners' professional needs is a recurring issue (Assassi, 2020; Akkar & Idri, 2021). The impact of dysfunctional assessment practices extends beyond academic performance, influencing students' perception of the evaluation process. Inconsistent and poorly designed assessments may lead to frustration and disengagement among students, hindering their overall learning and development experience. The absence of formative assessments and the thematic issues identified in design, task

## ***CHAPTER THREE Data Analysis and Discussions***

---

specifications, and scoring methods collectively contribute to a potential barrier in the holistic development of 1st Year Computer Science domain-related linguistic skills.

Within these challenges, the ICL Test emerges as a positive outlier. Unlike the generic tests, mainly produced by part-time teachers, the ICL Test incorporates feedback provisions and aligns with the LOA framework and the MOM of L2 Proficiency. The data highlights that the ICL Test strongly and positively influences learners' performance and engagement, bridging the gap between language proficiency and computer science discourse. It is positioned as a promising approach that not only enhances performance but also contributes to a more constructive learner experience.

Survey data amplify the subjective experiences of learners during ESP assessments. Concerns about exam instructions' comprehensibility and logistics underscore potential issues in assessment design. Learners' uncertainty about the adequacy of the English Final Exam in evaluating various language skills hints at potential gaps in the current evaluation methods. The unanimous negative opinion regarding the assessment of listening skills reinforces the need for a comprehensive evaluation that aligns with the real-world demands on computer science professionals (Assassi, 2020; Saraa, 2023).

The study exposes a substantial misalignment between ESP assessment and the actual needs of 1st Year Computer Science students. Learners express dissatisfaction with the inadequacy of ESP assessments in preparing them for language use within the academic and professional domains of Computer Science. The impact of dysfunctional ESP assessment practices in Algerian higher education is far-reaching, extending beyond academic performance to influence students' overall perception of the evaluation process. The challenges identified in recent studies, as well as the present research, have highlighted the potential hindrances to the holistic development of English language skills among Algerian students.

The predominant use of summative assessments, coupled with the absence of feedback provisions, poses a critical challenge. This deficiency inhibits students' ability to learn and

## ***CHAPTER THREE Data Analysis and Discussions***

---

improve (Bezziou & Ahmed, 2013). Thematic analysis exposes a concerning pattern of redundancy and potential lack of innovation in instructional materials, limiting the effectiveness of ESP assessments. As Saraa (2023) emphasizes, the lack of clear assessment procedures makes it difficult to monitor students' learning, hindering their development of English language skills necessary for academic and professional careers.

Furthermore, the dissatisfaction expressed by learners regarding the inadequacy of ESP assessments in preparing them for language use within the academic and professional domains of Computer Science underscores the urgency for adjustments in assessment practices. This dissatisfaction can potentially lead to frustration and disengagement, influencing not only the academic performance of students but also their overall learning and development experiences. In contrast, the ICL Test demonstrated the potential for a meaningful assessment experience, positively influencing learners' performance and engagement, bridging the gap between language proficiency and domain-specific discourse. This highlights the crucial role that well-designed assessments can play in enhancing language skill development, providing constructive feedback to students, and fostering a positive learning experience. The comparative analysis of data collected through the two experimental tests also shows the potential negative impact of the tests that inspired the design of the Generic Test (see Appendix A). These assessments, which are not systematically engineered in alignment with LA theory and learners' needs, do not have the potential to homogenize and boost learners' performance. ICL assessments that abide by LA standards and are learning-oriented by design have the potential to propel learners, regardless of their L2 proficiency level.

Addressing these challenges is imperative for students to achieve their domain-related language learning goals. The misalignment between ESP courses and learners' professional needs, coupled with a shortage of qualified ESP teachers, necessitates further research on assessment practices. By introducing learning-oriented assessment,

## ***CHAPTER THREE Data Analysis and Discussions***

---

diversifying assessment tools, and focusing on learner-centered design, the impact of ESP assessments can be transformed. These measures not only facilitate language skill development but also contribute to a more constructive and engaging learning experience, aligning assessments with the evolving needs of Computer Science. The impact of ESP assessment on language skill development in Algerian higher education is substantial, with challenges that warrant attention and reform. Addressing these challenges presents an opportunity to enhance the overall learning experience and proficiency development of students in the ESP context, ensuring they are well-equipped for academic and professional success.

### **3.6 Conclusion**

The comprehensive examination of ESP assessment practices in the context of Computer Science education in Algeria reveals a myriad of challenges and opportunities for improvement. The study employed a robust mixed-methods design, combining qualitative and quantitative analyses to provide a nuanced understanding of the evaluative practices employed by ESP practitioners. The assessment artifacts, classroom observations, and survey responses collectively shed light on the current state of ESP assessment, offering valuable insights for both researchers and practitioners.

The qualitative phase illuminated the intricacies of evaluative practices through assessment artifact analysis and classroom observations. Thematic analysis uncovered recurring issues in assessment design, including a lack of formative assessment, misalignment with learners' needs, and a dearth of diversity in instructional materials. The significant role of the Department of Mathematics in organizing assessments highlighted potential misalignments between assessment content and the specific needs of computer science students.

Quantitative data, analyzed with statistical techniques, provided a quantitative lens to the evaluation of language proficiency. The symmetrical distribution of scores in both the Generic Test and ICL Test indicates consistency in English language proficiency. Noteworthy is the ICL

## ***CHAPTER THREE Data Analysis and Discussions***

---

Test's advantage in grammar tasks, attributed to prior enablement through feedback provisions. The ICL Test's emphasis on topical knowledge, as seen in the item-based MCQ format, positively impacted scores and grades, with the ICL Writing Task standing out as a particular strength.

The study's findings further revealed several shortcomings in the current ESP Final Exams. The absence of listening assessments, limited evaluation of writing and speaking skills, and a narrow focus on grammatical and lexical knowledge in writing tasks were identified as key issues. Additionally, the research highlighted a misalignment between the assessments and the language needs of computer science students in both academic and professional domains.

The essay also explored the practicality, impact, validity, and reliability of classroom formative assessment in ESP education. It emphasized the need for diversified, structured formative assessment practices aligned with students' needs and learning objectives. The study revealed that effective formative assessment practices contribute positively to student learning outcomes, engagement, and the overall assessment experience.

In essence, the research underscores the importance of re-evaluating and refining ESP assessment practices in Computer Science education in Algeria. The identified challenges call for a shift towards more formative, authentic, and learning-oriented assessment approaches, tailored to the specific needs of learners. The integration of topical knowledge and language skills, as exemplified by the ICL Test, presents a promising model for enhancing language proficiency in the context of computer science. Ultimately, the study's insights provide a foundation for future research endeavors and practical interventions aimed at improving ESP assessment practices and, consequently, enhancing the language learning experience for computer science students in Algeria.

## CHAPTER FOUR: Suggestions and Recommendations

### 4.1 Introduction

### 4.2 Incorporating Formative Assessment

#### *4.2.1 Formative Assessment Potential*

#### *4.2.2 Enriching Classroom Assessment*

#### *4.2.3 Boosting Classroom Assessment Effectiveness*

### 4.3 Incorporating Learning-oriented Assessment

#### *4.3.1 Developing LOA in the Algerian HE Context*

#### *4.3.2 A Focus on Performance Moderators*

#### *4.3.3 Feedback in LOA*

### 4.4 Diversifying Assessment Tools

#### *4.4.1 ICL Assessment in the Algerian ESP Context*

#### *4.4.2 The Need for Listening and Speaking*

#### *4.4.3 Improving ESP Classroom Logistics*

#### *4.4.4 Designing Useful LA Tests*

#### *4.4.5 Rethinking Assessment Delivery in the Age of AI*

### 4.5 Enhancing Teacher Educational Development

#### *4.5.1 Training Programs*

#### *4.5.2 The Algerian ESP Centre*

### 4.6 Fostering Research on ESP Assessment

### 4.7 Conclusion



**4.1 Introduction:**

The assessment of English for Specific Purposes (ESP) among 1st Year Computer Science students at the Faculty of Sciences in Algeria has been explored through a sequential design to ultimately uncover both challenges and opportunities within the current evaluative practices. Through this chapter, recommendations will mainly reflect on the pivotal role of effective assessment in shaping the academic and professional journey of students. The qualitative and quantitative data collected throughout this study have revealed critical issues surrounding ESP assessments, ranging from the absence of formative dimensions to concerns about the authenticity of evaluations concerning real-world demands on computer science and IT professionals.

This chapter aims to bridge the gap between identified challenges and actionable solutions, offering recommendations that have the potential to enhance the overall ESP assessment experience of Algerian students in general, in addition to providing support to ESP practitioners. The insights gained from the review of the literature as well as the data collected throughout this research work provide a foundation for informed and practical recommendations that can change the value and intensity of LA impact in ESP contexts.

By addressing the highlighted challenges, we aspire to contribute to a paradigm shift in ESP assessment, one that aligns more closely with the evolving needs of computer science professionals and fosters a more constructive and enriching learning experience for students. This chapter serves as a roadmap for implementing tangible improvements that can propel ESP assessment practices toward a more learner-centered and learning-oriented future.

**4.2 Incorporating Formative Assessment*****4.2.1 Formative Assessment Potential***

The findings from research on the deployment of formative evaluative strategies in the Algerian Classrooms show a detrimental lack of this particular type of assessment (Aouine, 2011; Boumediene & Hamzaoui – Elachachi, 2017, Mandi & Kacha, 2023; Remadna, 2016). Formative assessment plays a pivotal role in enhancing students' academic achievements and, specifically, their comprehension of texts. In light of the findings from this study, advocating for the integration of formative assessment practices within Algerian educational contexts is imperative.

The integration of formative assessment should be accompanied by a robust goal-setting process (Dixson & Worrell, 2016). Clear and well-communicated learning goals benefit both teachers and students. It enables students to visualize their learning journey, contributing to better results and quicker progress. The absence of clear goals can lower motivation and hinder the learning process. Teachers should communicate learning goals explicitly, fostering student engagement and creating a foundation for effective formative assessment.

Creating a safe and inclusive classroom environment is essential for student involvement. Utilizing scaffolding techniques and CATs, teachers can actively engage students in the learning process, considering diverse learning needs. The importance of involving all students, including low achievers, through scaffolding is emphasized. This inclusive, learner-centered, approach not only raises self-confidence but also fosters a positive attitude towards learning, contributing to increased motivation. Formative assessment emerges as a tool that significantly boosts students' motivation (Boumediene & Hamzaoui –Elachachi, 2017). Recognizing the instrumental role of

formative assessment in learning achievements is crucial to addressing issues of low motivation and interest.

In the Algerian context, fostering a culture of student involvement in the teaching-learning process is a key challenge due to various factors. However, recognizing and enhancing students' roles in formative assessment is crucial for its success. The student's contribution is multifaceted, involving goal setting, peer assessment, and self-assessment. Student involvement in formative assessment begins with goal setting. These goals involve reaching new levels of proficiency and more meaningful achievements. Teachers play a vital role in modeling and guiding students in goal-setting practices, fostering self-regulation skills and active participation in their learning (Qu & Zhang, 2013)

While peer-assessment is a valuable practice, its implementation is constrained by time limitations in the Algerian context. Peer assessment involves feedback exchange among students, promoting interaction, mutual problem-solving, and experimentation. Some techniques involve sharing specific tasks, asking follow-up questions, offering suggestions, and engaging in class discussions. Through guidance (scoring rubrics, key answers, and sample questions) and practice students can develop effective peer-assessment strategies, providing constructive feedback to their peers.

Similar to peer-assessment, self-assessment is constrained by time in the Algerian context but holds significant potential for promoting motivation and self-regulation. Self-assessment aids students in developing crucial strategies for success, including planning, evaluating, and monitoring their own learning. Techniques such as assigning homework with specific criteria for self-assessment and applying peer assessment techniques to self-assessment can be effective. The

teacher's guidance, coaching, and modeling are essential for students to engage confidently and accurately in self-assessment practices over time.

#### ***4.2.2 Enriching Classroom Assessment***

Classroom assessment techniques (CATs) are integral components of educational practices, providing instructors with valuable insight into student learning while fostering a reflective and student-centered environment (Hanson & Florestano, 2020). Within this landscape, CATs can be broadly categorized into summative and formative assessments, each serving distinct purposes. Summative assessments, which encompass tests and student ratings, are designed to assess the extent of retained knowledge at the conclusion of a learning period. In contrast, formative assessments, emphasizing reflection and continual improvement, facilitate adjustments before summative evaluations (Adams, 2004).

Since the 1980s, the notion of formative classroom assessment has gained prominence, and within this framework, CATs have emerged as distinct teaching strategies crafted to enrich learning through reflective evaluations. (Simpson-Beck, 2011). These techniques involve students in the assessment process, providing instructors with valuable feedback on the quality of learning experiences. Recognizing the shift in focus from enhancing teaching to catalyzing learning, formative assessment techniques underscore the role of students in actively engaging in their own learning (Hanson & Florestano, 2020; Simpson-Beck, 2011).

Implementing CATs involves a systematic approach, starting with planning. Instructors must carefully select techniques aligned with course goals, their teaching style, and ease of implementation. The chosen technique is then introduced to students with clarity regarding its purpose. Subsequent steps involve collecting student responses, promptly analyzing data, and

responding to students by sharing understandings gained from the assessment (Simpson-Beck, 2011). CATs, such as the Minute Paper, one-sentence summary, Memory Matrix, and Muddiest Point, cater to diverse instructional needs, differing in difficulty and time requirements (Marzano & Kendall, 2007).

Although qualitative literature underscores the efficacy of CATs, there is a scarcity of empirical evidence supporting this assertion (Simpson-Beck, 2011). The existing literature presents a nuanced view of the cause-and-effect relationship between CATs and learning outcomes. Some contend that CATs predominantly improve teaching methods and establish a favorable learning atmosphere, thereby indirectly impacting learning outcomes. However, CATs play a crucial role as effective instruments for providing formative support, fostering student self-assessment, and encouraging active engagement in the learning process (Marzano & Kendall, 2007).

Among the myriad CATs available, the Know-Want-Learned (K-W-L) chart technique stands out as a prominent method (Conderman & Hedin, 2012). This collaborative approach actively involves students with informational materials, leveraging their preexisting knowledge to establish a sound foundation for learning. The K-W-L acronym encapsulates three essential phases: "what I know," "what I want to know," and "what I have learned." This technique, rooted in activating and building upon learners' schemata, serves to establish a purpose and facilitate the synthesis of new knowledge.

The K-W-L chart technique serves as a dynamic tool for both evaluating students' learning experiences and providing instructors with valuable insights. In its application, students navigate their learning process by actively engaging with reading materials. The technique guides

them through three distinct phases: identifying existing knowledge, articulating areas of curiosity, and reflecting on the knowledge gained. By fostering a collaborative and reflective learning environment, the K-W-L chart technique enhances the overall effectiveness of CATs.

According to Simpson-Beck (2011), CATs are closely associated with assessment for learning (AFL) and formative assessment (FA). These methodologies share a common goal of systematically improving student learning and motivation. McMillan (2015) underscores the significance of establishing clear criteria to assess student learning and comprehension within the context of formative assessment. These criteria, serving as pivotal indicators of success, encompass diverse dimensions such as student responses and the quality of products and performances that ultimately determine the level of understanding.

The integration of CATs, curriculum, and instruction within the framework of assessment for learning has a positive impact on both student achievement and motivation (Hanson & Florestano, 2020; McMillan, 2015; Simpson-Beck, 2011; Thomas & Hornsey, 2014). This comprehensive educational approach prioritizes the mastery of knowledge and skills, transcending mere test scores to cultivate students' abilities to evaluate their understanding, pinpoint knowledge gaps, and devise effective learning strategies. For educators, assessment for learning represents a mindset centered on improving both student learning outcomes and motivation, fostering qualities essential for success in professional and personal pursuits. Among the various CATs available, the Know-Want-Learned (K-W-L) chart technique stands out as a potent and dynamic tool that seamlessly aligns with the objectives of formative assessment, contributing significantly to the enhancement of students' competencies. Through collaborative exploration of informational materials, this technique not only assesses learning experiences but also furnishes instructors with valuable evaluative insights.

***4.2.3 Boosting Classroom Assessment Effectiveness***

McMillan (2015) outlines five crucial steps in the formative assessment process. First, a set of clear and objective criteria for assessing student learning must be established. These criteria represent benchmarks, defining success and encompassing the features of student responses and product dimensions. Second, evidence demonstrating student learning and understanding must be collected. This objective can be accomplished through the implementation of pretests, observations during instructional activities, and the application of effective questioning techniques. Utilizing a combination of verbal and nonverbal indicators enables an accurate assessment of progress.

The third phase entails delivering timely, precise, and personalized feedback to students, tailored according to their individual performance. Effective feedback is crucial for promoting student learning, encouraging future improvement, and linking directly to established criteria. Simultaneously, CATs encourage self-assessment and peer assessment, fostering self-monitoring and self-determination. These practices empower students to reflect on their learning, identify strengths and weaknesses, and contribute to the learning process by generating ideas for improvement.

The final step of the formative assessment process revolves around implementing instructional adjustments or correctives. Teachers assume a crucial role in directing students through supplementary learning activities designed to augment comprehension and mastery. These activities ought to diverge from previous ones and may involve integrating students' own suggestions for enhancement. Through the promotion of self-directed learning and the provision

of tailored support, teachers enable students to traverse the divide between their existing knowledge and the sought-after learning outcomes within the specific domain.

Teachers also need to pay attention to the criteria underpinning the validity of formative assessment, ensuring its alignment, diagnostic value, fairness, and quality (Kane & Wools, 2019). Formative assessment should closely align with the learning objectives, which are in turn aligned with learners' and stakeholders' needs. This alignment ensures that assessment results provide meaningful insights into students' progress toward achieving standards.

Teachers can use a numerical metrics to monitor learning. However, the scores need not to be used to directly impact learners' official final and official scores and grades. The Algerian educational system is deeply marked by the teach-to-the-test approach. This may lead to dysfunctional formative assessment, whereby intentions are formative and use is summative. The primary aim of formative assessment is to enhance students' achievements, necessitating a focus on diagnostic feedback. Assessments with high diagnostic value not only reveal performance levels but also provide insights into why students perform at certain levels, guiding further improvement. The information derived from formative assessment should be constructive and diagnostic, aiding both teachers and students in refining instructional strategies. The utility of formative assessment lies in its ability to offer valuable insights for instructional improvement.

Formative assessment demands fairness in commenting on students' progress, promoting an equitable evaluation process. Moreover, Formative assessment must exhibit reliability in delivering consistent results and validity in eliciting performance and measuring constructs. Considering the cyclical nature of formative assessment, teachers play a pivotal role in instructing (teaching and re-teaching), assessing (collecting and analyzing data), and providing



feedback (generating and delivering corrective measures) to consolidate learning outcomes. Learners, on the other hand, need to be aware of their levels and understand the operational dynamics of formative assessment to actively participate in the teaching-learning process.

### **4.3 Incorporating Learning-oriented Assessment**

#### ***4.3.1 Developing LOA in the Algerian HE Context***

Unlike traditional assessment paradigms, LOA prioritizes the interpretation of language performance evidence in diverse contexts, aiming to draw inferences about learners' language knowledge, skills, and abilities (KSAs) within specific domains (Turner & Purpura, 2016). Crucially, LOA goes beyond mere outcome evaluation; it places equal weight on understanding the processes involved in language learning. This nuanced approach allows individual classroom agents, including teachers and students, to make goal-referenced decisions that propel further language processing and contribute to achieving target-like performance.

Algerian Higher Education institutions should adopt a holistic view of LOA that accentuates learning goals and individual progressions. This involves clearly delineating learning objectives, aligning assessments with these goals, and utilizing performance evaluations and feedback as tools to guide learners through their language learning journey. By integrating LOA principles into the curriculum, educators can foster a dynamic educational environment where the focus is on continual improvement and the attainment of language competencies.

LOA recognizes language learning as a highly individual cognitive process, with additional layers of complexity when situated within collaborative settings. Algerian Higher Education should encourage collaborative practices that leverage interaction to exchange meanings and jointly develop language competencies. Recognizing the socio-cognitive and

sociocultural dimensions of language learning, educators can create inclusive learning environments that celebrate diversity and engage learners in meaningful interactions, enhancing the effectiveness of LOA in practice.

While curriculum, instruction, and assessment undoubtedly impact language processing and learning success, LOA acknowledges the influence of additional factors, such as learner engagement. Algerian Higher Education institutions should take a comprehensive approach to LOA by carefully considering performance moderators influencing on the assessment process. This involves not only refining curriculum and instruction but also fostering a supportive and engaging learning environment that motivates learners and encourages active participation.

In light of the comprehensive review on learning-oriented assessment by Zeng et al. (2018) and the subsequent developments in assessment paradigms it highlights, recommendations for the implementation of LOA in Algerian Higher Education are put forward, centering on method implementation. The intentional order of LOA components—Assessment for Learning (AfL), Assessment as Learning (AaL), and Assessment of Learning (AoL)—serves as a foundational principle for effective implementation. Algerian Higher Education institutions should encourage educators to plan assessments by considering the curriculum and their students. Addressing the five essential questions—why, what, which method, how to ensure quality, and how to use the information—guides the development of assessment methods aligned with the unique needs of each stage. Emphasizing the importance of AfL and AaL throughout the learning process ensures continuous metacognition, while AoL is reserved for summative decisions.

The introduction of LOA represents a significant shift in thinking about assessment as a key contributor to enhancing learning. However, this shift necessitates a change in the mindset of

educators, administrators, parents, and society. Understanding and motivation are critical elements in this process, urging stakeholders to engage actively in the changes proposed by LOA. ESP practitioners, in particular, should enhance their knowledge and skills to effectively implement LOA principles, shifting their paradigm to view assessment as a driver of instruction and a positive influence on student learning. Principals play a vital role in supporting these changes through strong instructional leadership, fostering a culture of innovation and professional growth.

Mindset-changing is intrinsically linked to capacity-building. Algerian Higher Education institutions should invest in initiatives that focus on professional learning, leadership, and collaboration with parents and the community. Professional learning should encompass both formal and informal strategies, including in-service sessions, professional development, assessment study groups, and collaborative practices. Leadership and support strategies, such as boundary-spanning activities and developing critical friendships, are essential to create an environment where LOA can thrive. Collaboration with parents, students, and the community further strengthens the LOA process, turning it into a partnership that involves, rather than dictates, the stakeholders.

The ICL Test design and operationalization in the context of this research represent an instance of LOA implementation into computer science ESP assessment engineering. The implementation of Learning-Oriented Assessment in Algerian Higher Education requires consideration of strategies to implement this new assessment method and its frameworks. By adopting assessment practices aligned with the LOA framework, educators can create an environment that fosters continuous learning and metacognition. Changing the mindset of stakeholders, particularly educators, is crucial for embedding LOA into classroom practices. This

shift in perspective, accompanied by professional development and strong support, contributes to the whole potential of LOA, ensuring it becomes an integral part of the educational landscape. As Algerian HEIs embrace LOA, they have the potential to significantly enhance the quality of learning and prepare students for a lifetime of inquiry and self-directed learning.

#### ***4.3.2 A Focus on Performance Moderators***

Turner and Purpura's (2016) Learning-Oriented Assessment (LOA) framework introduces several critical dimensions, known as performance moderators, which significantly influence the dynamics of LA. These moderators offer a thorough insight into the dimensions that directly or indirectly influence the performance of learners, shaping the Learning-Oriented Assessment (LOA) process.

The contextual dimension emphasizes the impact of socio-political forces, socio-cultural norms, and other education-related parameters on curriculum, instruction, and assessment. Algerian HEIs should consider these contextual factors when implementing LOA, recognizing that the broader socio-cultural environment significantly influences the learning process.

At the core of LOA lies the socio-cognitive dimension, delving into the theories of learning and cognition within second language classrooms. Educators in Algerian Higher Education should align their instructional practices with theories that recognize learning as a social and participatory process. Encouraging activities that involve collaborative meaning-making, such as discussions and interactions, can enhance language processing and contribute to learning success.

The instructional dimension underscores the critical role of teachers' linguistic content knowledge, topical content knowledge, and pedagogical content knowledge in shaping the LOA landscape. Algerian educators should prioritize the continuous development of these knowledge types, understanding their impact on language processing, learning outcomes, and the effectiveness of LOA. This involves aligning language instruction with topical content and ensuring that teachers possess the necessary knowledge to identify and address language performance errors.

Moreover, acknowledging the impact of affective factors on learning success, the affective dimension of LOA delves into learners' emotions, beliefs, personality characteristics, attitudes, and motivation. In the Algerian Higher Education context, educators should consider the affective dimension as a crucial aspect of the assessment process. Tailoring LOA practices to accommodate learners' socio-psychological predispositions can enhance engagement and contribute to overall learning success. LOA also recognizes the importance of interactional dynamics in the assessment process. Algerian Higher Education institutions should focus on understanding the organization of LA within talk-in-interaction sequences. Recognizing patterns of positive and negative evaluations, coupled with scaffolded assistance during repair sequences, can contribute to the effectiveness of feedback.

As shown in Table 2.4, assessment designers have to ponder on each of the LOA framework dimensions that moderate and indicate performance. More often, proficiency and elicitation dimensions are intuitively planned. Nevertheless, Algerian Higher Education institutions seeking to implement LOA should meticulously consider performance moderators. Addressing the contextual, socio-cognitive, instructional, interactional, and affective dimensions will not only enrich the LOA experience but also foster a more inclusive and effective learning

environment. This aligns with the overarching goal of LOA: to facilitate learning through purposeful and contextually relevant assessment practices.

### ***4.3.3 Feedback in LOA***

Feedback stands out as a leading force and a key cornerstone in any assessment that privileges development, even if summative. Teachers must be trained to provide effective feedback that aligns with students' needs and proficiency levels. Differentiated feedback, tailored to individual learners, ensures the effectiveness of formative assessment (Owen, 2016). The proficiency that a teacher may aim to elicit is defined beforehand, hence, allowing teachers to plan collective feedback as well.

Timely remediation and regulation, including re-teaching, are also common manifestation of formative and learning-oriented assessments (O'Donovan et al., 2021). The teacher's role in addressing learning gaps through targeted remediation ensures the success of the whole evaluative procedure. The instructional dimension of the ICL Test, as a component of the whole learning-oriented mechanism, moderated the impact of the whole evaluative procedure to homogenize and boost students' language skills development through a better performance.

To elevate the effectiveness of Learning-Oriented Assessment (LOA), educators must prioritize the pivotal role of feedback, ensuring a harmonious integration of Assessment for Learning (AfL), Assessment as Learning (AaL), and Assessment of Learning (AoL). Educators should cultivate a thorough understanding of feedback, recognizing its multifaceted nature within the LOA framework (Lam, 2021). Professional development opportunities should emphasize the nuances of providing constructive feedback aligned with LOA principles.

Educators should strive for a balanced integration of feedback mechanisms, acknowledging the unique contributions of AfL, AaL, and AoL. Aim for an equilibrium that enhances the learning experience, with a heightened emphasis on the learning-oriented and formative aspects of feedback (Zeng et al., 2018). Learning-oriented feedback is purpose-driven (Carless, 2015). Assessment design should clearly articulate the purpose of feedback within LOA—whether it serves to inform learning progress or supports summative judgments, for instance. Feedback strategies should be tailored based on the intended purpose, ensuring that feedback aligns seamlessly with the overarching educational objectives.

ESP practitioners should embrace a shift towards learning-oriented feedback practices within evaluative practices, emphasizing ongoing assessment, diagnostic insights, and continuous improvement. Learning-oriented feedback actively contributes to students' learning journeys, fostering a dynamic feedback loop. Through LOA, a learning environment centered around feedback is fostered, encouraging students to actively engage in self-assessment and reflection, using feedback as a catalyst for their own learning progression (Zeng et al., 2018). This promotes a culture where feedback is viewed as a valuable tool for improvement. Therefore, constructive self-assessment practices are actively facilitated, empowering students to critically analyze their own work. Learning-oriented feedback, hence, guides students to refine their understanding, study strategies, performances, and take ownership of their learning journey.

Feedback represents as a dynamic element within LOA, adapting to the changing needs of both educators and learners. Educators should be provided with continuous professional development opportunities specifically focused on feedback methodologies within LOA. Training programs and workshops should equip educators with the skills to deliver timely, constructive, and purposeful feedback. By placing a strong emphasis on the role of feedback

within LOA, educators can create an enriched learning environment that not only informs students of their progress but actively engages them in a collaborative process of improvement and achievement.

#### **4.4 Diversifying Assessment Tools**

##### ***4.4.1 ICL Assessment in the Algerian ESP Context***

English for Specific Purposes (ESP) plays a pivotal role in Algerian Higher Education, equipping students with the linguistic proficiency and disciplinary knowledge essential for their ultimate success in their chosen domains. The assessment methodologies employed to evaluate ESP students' progress effectively need to evolve accordingly with the educational and professional landscapes. A transformative approach to ICL assessment is advocated throughout the present study to match diversity, authenticity, collaboration, and technological innovation.

The complex nature of ESP disciplines demands a corresponding diversity in assessment tools. Traditional exams may not adequately capture the breadth of skills and competencies required for specific professions. Integrating content and language in assessment methods allows for a more comprehensive evaluation of students' KSAs. Practical projects provide an opportunity for students to apply their language skills in real-world contexts, simulating tasks they may encounter in their future careers. Case studies foster critical thinking and problem-solving abilities, while simulations encourage collaboration and effective communication in dynamic environments.

Aligning ICL assessment with real-world demands necessitates a shift towards authentic, both task-based and construct-based, assessments. These assessments mirror the types of tasks that students will encounter in their professional lives, providing a more meaningful and



applicable evaluation of their language proficiency. Incorporating real-world tasks such as writing industry reports, conducting professional presentations, or engaging in discipline-specific communication allows for a more holistic assessment of students' abilities that aligns the TLU domain. Such tasks not only assess their linguistic KSAs but also their KSAs in the relevant professional context.

To ensure that ICL assessments accurately reflect real-world demands, collaboration with industry experts is essential. Professionals from relevant disciplines possess invaluable insights into the language skills and competencies required in actual workplace scenarios. Their expertise can inform the development of authentic assessment tasks and criteria. Engaging industry experts in the assessment design process not only enhances the validity of assessments but also establishes a valuable connection between academia and industry. This collaboration fosters a shared understanding of the skills and knowledge required for success in the professional world, contributing to the overall quality and relevance of ESP education.

In the digital age, technology offers a plethora of possibilities for innovative ICL assessments. Integrating platforms like virtual reality, online simulations, or collaborative tools can provide students with a simulated real-world environment, allowing them to navigate language challenges within the digital landscapes they are likely to encounter in their future professions. Technology-based assessments offer several advantages, including increased engagement, interactivity, and opportunities for personalized feedback. Virtual reality simulations can immerse students in realistic professional settings, while online simulations can replicate specific industry tasks. Collaborative tools can facilitate group projects and communication, fostering teamwork and effective communication skills.

By diversifying assessment tools, integrating authentic scenarios, collaborating with industry experts, and leveraging technology, Algerian Higher Education can create a dynamic assessment environment that not only evaluates language proficiency but also prepares ESP students to excel in their future professional endeavors. This transformative approach ensures that ESP education remains aligned with the evolving demands of the contemporary professional landscape, empowering students to thrive in their domain.

#### ***4.4.2 The Need for Listening and Speaking***

There is a growing recognition of the need for assessments that authentically reflect the demands of real-world professions, especially in disciplines like computer science. The study highlighted concerns about the absence of listening and speaking constructs in current evaluative practices, emphasizing the importance of aligning evaluations with the dynamic requirements of computer science professionals.

One of the primary concerns in current assessments within Algerian Higher Education is the limited inclusion of listening and speaking. These vital language skills are often marginalized in favor of written evaluations. It is imperative to address this imbalance and recognize the importance of holistic language proficiency that encompasses listening and speaking abilities. Assessments should authentically align with the real-world demands placed on computer science professionals. While written assessments have their place, the dynamic nature of the field necessitates evaluations that mirror the actual challenges faced by professionals. This involves incorporating tasks that require students to engage in speaking or writing, fostering the co-construction of knowledge through reading and listening.

A transformative shift requires educators to address the historical neglect of listening and speaking skills in ESP classes. Historically viewed as an ability that would naturally develop, listening skills have gained prominence in recent years as language input's pivotal role in language learning becomes more apparent (Mart, 2020). The conventional belief that listening skills would develop without explicit assistance has hindered the holistic growth of learners' L2 proficiency. Algerian language educators must recognize that language development without robust listening skills limits students' ability to effectively express themselves orally. Listening serves as the foundation for the four fundamental processes of oral proficiency: Conceptualization, Formulation, Articulation, and Self-monitoring (Mart, 2020).

Listening provides a fertile ground for students to encounter and internalize good language models. Exposure to diverse language models facilitates adaptation to different speech tempos and vocabulary usage. Through noticing and conscious awareness, learners can actively monitor their listening, enhance comprehension, and create meaning. This immersive experience lessens the frustration associated with language learning, fostering the development of communicative competence. The incorporation of listening and speaking activities can significantly contribute to the metacognitive development of learners. Metacognitive knowledge, involving thinking about the language learning process, empowers learners to regulate and direct their own learning. Stimulating students' reflection on their listening and speaking skills performance enhances motivation and positively influences their language performance (Nguyen & Huynh, 2021)

Utilizing schema theory in the pre-listening phase allows learners to activate prior knowledge, facilitating better comprehension. Discussions about the content before listening enhance perception and aid in connecting new information to existing knowledge. The post-

listening stage becomes an opportunity for deeper learning, as learners use the acquired information as a springboard for further language practice.

Assigning speaking activities, after listening, not only encourages practical application of language knowledge but also boosts learners' confidence. The realistic setting immerses them in authentic language use, prompting more independent language utilization. The communicative approach, emphasizing the integration of listening and speaking skills, aligns with the principle that language learners enhance their oral production by decoding messages from interlocutors.

The present study's findings underscore the total lack of integrating listening and speaking activities during instructional and evaluative activities, which inhibits the transformative impacts on the development of communicative language use. For the Algerian HE to evolve effectively, educators must move beyond traditional approaches and adhere to the symbiotic relationship between listening and speaking. Exposure to language input through listening is not just essential for conversation skills but is a cornerstone for comprehensive language development.

Encouraging the integration of listening and speaking constructs is essential for creating more authentic, comprehensive, and valid evaluations in computer science. Rather than viewing language skills in isolation, assessments should reflect the interconnected nature of communication in the field. This integration not only mirrors professional scenarios but also provides a more accurate gauge of a student's overall language proficiency.

Task design should emulate real-world scenarios where computer science professionals engage in collaborative problem-solving, requiring effective communication. Incorporating tasks that necessitate discussions, presentations, or collaborative projects can authentically assess

students' ability to articulate ideas, understand complex concepts through listening, and contribute meaningfully to knowledge co-construction.

Educators play a crucial role in reshaping assessment practices. Providing continuous professional development opportunities focused on integrating listening and speaking constructs in computer science assessments is crucial. Workshops, training sessions, and collaborative forums can empower educators to design assessments that align with industry demands and foster well-rounded language skills. Stakeholders need to foster collaboration between educational institutions and industry experts in specific domains. Understanding the communication demands faced by professionals in the field can guide the development of assessments. Industry insights can provide valuable perspectives on the essential language skills required for success in the workplace.

Stakeholders have to ensure assessments are integrated and cater to diverse learning styles. Educators need to recognize that students may have varying strengths in listening and speaking as opposed to writing or reading on which they have been training long before. Integrated-skills assessments acknowledge and accommodate these differences, providing a more equitable evaluation of students' language proficiency.

The enhancement of assessment practices in Algerian Higher Education, particularly in computer science, necessitates a paradigm shift towards authentic evaluations. By addressing the absence of listening and speaking constructs, aligning assessments with real-world demands, and fostering collaboration between educators and industry professionals, Algerian Higher Education can better prepare students for the dynamic and communication-intensive landscape of domain-related professions. Through these recommendations, the aim is to create assessments that not

only measure academic knowledge but also cultivate the domain-specific language skills essential for success in the professional realm.

#### ***4.4.3 Improving ESP Classroom Logistics***

English for Specific Purposes (ESP) teaching in the Algerian Higher Education context demands a paradigm shift to address the challenges posed by massive cohorts. Traditional methods of instruction in overcrowded lecture halls, lacking essential logistics for effective teaching and evaluation, fall short of meeting the specialized needs of ESP students. The current practice of teaching ESP in overcrowded lecture halls poses significant challenges.

These spaces do not really represent the appropriate environment to build students' confidence in listening and speaking, more often lacking essential teaching aids such as speakers, video projectors, and microphones. In contrast, classrooms provide a more conducive environment for focused listening and speaking teaching and assessment. The adaptation of teaching spaces to smaller, more manageable classrooms allows for greater interaction, personalized instruction, and the use of instructional technologies that promote language learning.

Classrooms, unlike lecture halls, offer an intimate setting that is better suited for instructive and evaluative practices of listening and speaking skills. In smaller groups, instructors can employ interactive methodologies, such as group discussions, role-playing, and peer assessments, fostering a more dynamic and engaging learning experience. This shift from the traditional lecture hall format to smaller classrooms aligns with the pedagogical principles of communicative language teaching and allows for a more personalized and effective ESP instruction.

As shown in the present study, traditional exam formats often fall short of adequately evaluating listening and speaking skills. To address this, it is crucial to review and adapt exam organization to allow for a more integrated-skills assessment that involves these essential language competencies. Moving beyond written exams, integrating listening and speaking components into assessment strategies can provide a more accurate reflection of students' communicative abilities and L2 proficiency in domain-specific contexts. Practical evaluation methods, such as oral presentations, interviews, and recorded conversations, can offer a more authentic measure of L2 proficiency.

Creating an environment conducive to listening and speaking assessments requires attention to logistics. Dedicated spaces equipped with audio-visual resources, recording facilities, and interactive online platforms are essential for evaluating these skills effectively. Investing in the necessary infrastructure ensures that ESP students have the opportunity to demonstrate their language proficiency authentically. Additionally, offering training and resources for instructors to design and conduct listening and speaking assessments, using technology, is integral to the success of this initiative (Bouabdallah, 2015).

Given the challenges posed by massive cohorts, leveraging electronic assessments (E-assessments) becomes a practical solution. E-assessments not only cope with the logistical constraints of traditional exam settings but also offer scalability for evaluating large numbers of first-year cohorts. Online platforms can facilitate both listening and speaking evaluations through recorded responses, virtual interviews, and interactive simulations. This shift to E-assessments aligns with global trends in educational technology and provides a more efficient and scalable approach to language evaluation.

Learners' speaking ability can be improved through regular video recordings and uploading, indicating the potential of e-assessment platforms (Zeng et al., 2018). In the Algerian context, adopting similar practices on e-learning platforms can provide students with a comprehensive assessment of their integrated language, and even topical, skills. E-assessments can be tailored to include tasks that require both listening and speaking, fostering a more authentic evaluation of students' communicative competence. This approach aligns with the global shift toward technology-enhanced language learning and offers a scalable solution for assessing integrated skills in massive cohorts.

While embracing e-assessment for integrated skills in the Algerian Higher Education, identifying and addressing the potential contextual issues becomes more than necessary. Educators in Algerian HEIs should receive training on leveraging e-learning platforms for integrated skills assessment. In Algerian Higher Education, where logistical and resource challenges exist, a phased implementation, different from the swift and massive Covid-19 pandemic experience, with continuous evaluation becomes essential.

Collaborative efforts between stakeholders can contribute to overcoming these limitations and scaling up the integration of e-assessments for integrated skills in ESP contexts. E-assessments, when effectively implemented, can enhance student engagement and motivation. Algerian educators should leverage the interactive features of e-learning platforms to create stimulating learning environments. Platforms like Canvas, Microsoft Teams, and Google Classroom allow for interactive discussions, collaborative projects, and multimedia content, providing students with diverse and engaging language experiences.



***4.4.4 Designing Useful LA Tests***

The landscape of education is evolving rapidly, with a growing emphasis ICL. In response to this shift, it is essential to align assessments closely with instructional goals, inform students about assessment objectives, and integrate various assessment methods that cater to the cognitive demands of diverse real-world tasks.

A fundamental recommendation is to align assessments closely with instructional goals. By ensuring that assessments directly reflect the intended learning outcomes, educators can enhance the relevance and effectiveness of the evaluation process. This alignment provides students with a clear understanding of the purpose behind each assessment and reinforces the connection between their efforts and the broader educational objectives.

Transparency in assessment is crucial for fostering a collaborative learning environment. Educators should communicate the objectives of formal assessments to students, clarifying how the results will be utilized to guide further learning. This not only demystifies the assessment process but also empowers students to take an active role in their own educational journey.

Recognizing the diverse cognitive demands of tasks, educators should design assessments that vary in complexity. Whether it's a stand-alone task or a multi-step set of tasks, the assessment design should align with the topical and linguistic KSAs required in a specific discipline. This approach ensures a more authentic evaluation of language constructs. To enhance the authenticity of assessments, educators should design tasks that closely resemble real-life challenges within a specific discipline. This not only promotes a deeper understanding of the subject matter but also prepares students for the complexities they may encounter in their future

careers. Real-world relevance adds value to the learning experience and reinforces the practical application of acquired knowledge.

Effective ICL assessments should encompass a variety of evaluation methods to cater to different learning styles and preferences. This includes multiple-choice or other selected-response tasks for objective assessments, shorter answer or other limited-production tasks for a more nuanced understanding, and written or spoken products for comprehensive evaluation. Additionally, collaborative projects provide an opportunity for students to apply their knowledge in a team setting, fostering teamwork and communication skills.

Moreover, delivering assessments online through platforms like Google Forms is a practical consideration, particularly in our digitalized era. This not only aligns with the preferences of the tech-savvy generation but also facilitates a streamlined and efficient assessment process. Online assessments can also be easily adapted to various formats, allowing for flexibility in task design.

In designing assessments, educators should consider and align their practices with the language assessment standards. Adhering to these standards ensures reliability, consistency, fairness, and quality in assessment practices, contributing to the overall effectiveness of the educational system. Scoring students' performance on essays, presentations, and other tasks should be done transparently using well-defined rubrics. Providing students with copies of these scoring rubrics enhances is a learning-oriented strategy that fosters transparency and allows them to understand the criteria used for evaluation. This transparency promotes a reliable and fair assessment process.

Beyond evaluation, the results of assessments should be used to leverage constructive feedback. Educators can provide feedback designed to promote further learning, identifying areas of improvement and guiding students towards a meaningful development of topical and linguistic KSAs. Assessment results can also be instrumental in reflecting on teaching practices and adjusting instructional strategies to better meet students' needs.

Encouraging students to submit multiple drafts of their work fosters an iterative approach to learning. This not only provides opportunities for continuous improvement but also allows students to internalize feedback, reinforcing the learning process. This iterative approach is also relevant the context of assessment design, whereby feedback is used to enhance assessment quality. The process of assessment design is of an iterative nature fostering ongoing readjustments and enhancements of evaluative practices.

Creating opportunities for students to receive feedback from peers promotes a collaborative learning environment. Peer feedback encourages communication skills, critical thinking, and a sense of community among learners. Moreover, collaborative projects, such as co-authored presentations, provide a platform for students to apply their knowledge collectively, enhancing their teamwork and problem-solving skills. Utilizing evidence of learning during instruction is a dynamic approach that involves continuous reflection and adaptation. Educators can adjust lessons based on the results of students' work, providing timely interventions and personalized support. This iterative process ensures that instruction is responsive to the evolving needs of the learners.

ESP Assessment Blueprints play a pivotal role in providing a detailed framework that aligns assessments with the unique linguistic and communicative requirements of a particular

field. These blueprints serve as a guide for educators, ensuring that assessments are not only relevant to the discipline but also tailored to the linguistic and topical KSAs required in real-world scenarios. Establishing clear guidelines for ESP assessments contributes to the transparency of evaluation criteria, promoting fairness and precision in gauging learners' language proficiency within specific professional contexts.

Furthermore, ICL can be augmented through the introduction of ICL Mock Tests. These mock tests serve as invaluable tools for preparing students for the diverse formats and collaborative dynamics they may encounter in actual assessments. ICL Mock Tests not only provide a simulated experience of the assessment environment but also enable educators to identify potential challenges and refine assessment designs accordingly through the data they provide.

By incorporating mock tests into the instructional framework, stakeholders can better understand the unique needs of their students, adjust their teaching strategies, and foster a supportive learning environment that encourages collaborative problem-solving and effective communication skills. In essence, the establishment of ESP assessment blueprints and the incorporation of ICL mock tests contribute significantly to the evolution of assessment practices, ensuring that evaluations are not only rigorous but also attuned to the specificities of both the discipline and the learners themselves.

The design and implementation of ICL assessments require a holistic and dynamic approach. By incorporating the additional recommendations outlined in this chapter, stakeholders can ensure assessments that align closely with instructional goals, engage students in the learning process, and provide a consistent evaluation of their understanding. This comprehensive

framework not only contributes to the effectiveness of ICL but also empowers educators to adapt to the evolving landscape of education and meet the diverse needs of their students.

#### ***4.4.5 Rethinking Assessment Delivery in the Age of AI***

In the rapidly evolving landscape of education, the integration of Artificial Intelligence (AI) tools, particularly chat generative models like ChatGPT and Gemini, has sparked significant debate and concern. The narrative often centers around the ethical implications of students leveraging AI for academic tasks, blurring the lines between genuine learning and the detrimental adoption of AI-generated content. A set of recommendations for rethinking Language Assessment in the Algerian Higher Education Context is needed, acknowledging the challenges posed by AI and proposing strategies to ensure the integrity of the educational process.

Research highlights the trend of students opting for the easy way out, utilizing AI to generate academic content (Cotton et al., 2023; Hassan et al., 2022; Shalevska, 2023). While AI, including tools like Google, can undoubtedly be beneficial for learning, the concern arises when it becomes a substitute for genuine effort and understanding. In rethinking Language Assessment in Algerian Higher Education, educators must recognize that AI, if misused, can erode the foundations of academic integrity. Therefore, it becomes imperative to assess the impact of AI not only on the educational tools but also on the very essence of learning.

A necessary exploration of the concept of cheating in the context of AI is required to disambiguate ethical and pedagogical conundrums. As AI tools can swiftly generate content, distinguishing between authentic student work and AI-generated output becomes a critical challenge. The definition of cheating needs refinement, aligning with the goals of assessments. If assessments aim at promoting independent thought, research, and the writing process, then undue

reliance on AI for content generation may indeed be considered a form of academic dishonesty. Therefore, any rethinking of Language Assessment in Algeria must grapple with a nuanced understanding of cheating in the AI era.

To preserve the quality and integrity of education, it is imperative to set strong standards and deterrents against the misuse of AI in assessments (Cotton et al., 2023). Educators should emphasize the importance of genuine learning, articulation of thoughts, and the development of critical thinking skills. Clear policies and consequences for AI misuse should be communicated, aligning with the seriousness attributed to traditional forms of cheating. By doing so, the education system can maintain its integrity and uphold the value of students' authentic efforts.

Like any technology, there is no inherent moral value of AI. Therefore, depending on the standpoint and the implemented proactive measures, AI can offer opportunities rather than raising challenges (Cotton et al., 2023). People are just projecting their own misuse onto the technology. In addition to setting clear policies and deterrents, stakeholders should actively promote responsible AI use. Providing students with clarifications on how AI can be included in the teaching-learning process and used to promote learning. AI should be presented as a tool for augmentation, assisting in research, summarization, and learning reinforcement. By fostering an understanding of AI as a supplementary learning aid rather than a shortcut, educators can guide students towards ethical and responsible AI use, ensuring its positive integration into the educational journey.

The integration of GPT-like AI into educational settings prompts researchers and stakeholders to explore innovative teaching methodologies. Collaboration emerges as a cornerstone for effectively navigating this paradigm shift (Kostka & Toncelli, 2023). Research

underscores the significance of teachers engaging in collaborative practices, such as co-teaching, sharing teaching materials, and participating in professional learning communities. In the context of AI integration, educators in Algerian higher education should adopt a collaborative approach to explore GPT-like AI potential applications. By collaboratively experimenting with AI tools, teachers can not only enhance their collective knowledge but also address the ethical and pedagogical implications of AI use in assessments.

Moreover, the collaborative spirit should extend beyond the confines of individual classrooms. Educational institutions need to establish comprehensive responsible use policies that account for ethical considerations related to academic integrity, privacy, and authorship. Building on the concept of communities of practice, educators, administrators, and stakeholders should actively engage in ongoing discussions to refine policies and share insights about AI integration (Kostka & Toncelli, 2023).

This collective approach ensures that responsible AI use is embedded in the broader educational framework, fostering a culture of transparency and adaptability. In addition to collaborative efforts, reflection emerges as a key aspect in redefining language assessment. Educators should involve students in discussions about the ethical use of AI tools and its role in supporting the learning process. Reflective practices should not only include student perspectives but also integrate teachers' insights at all levels of AI implementation, ensuring a balanced and inclusive approach to decision-making.

Embracing pedagogical flexibility and experimentation is essential. The transformative potential of AI in fostering critical thinking and creativity should be recognized (Cotton et al., 2023). Assessment methodologies need to evolve to focus on the learning process, emphasizing

critical digital literacy skills over rote memorization. As educators rethink learning activities in the AI era, they should explore new assessment approaches that encourage learners to utilize AI-generated information as part of problem-solving, analysis, and evaluation (Rudolph et al., 2023). By doing so, educators can catalyze a paradigm shift towards a learning-centric educational system that prepares students for the challenges and opportunities presented by the evolving landscape of AI.

In response to the challenges posed by AI, a reevaluation of assessment methods is warranted (Abd-Elaal et al., 2019). Flipping the classroom, emphasizing in-class writing and interactive learning, could be a strategic move. This approach not only minimizes the opportunity for AI misuse but also fosters a more engaging and participatory learning environment. To combat the potential threat posed by AI-generated written content, oral assessments should be integrated into the evaluation process. Students should be made aware that articulating their understanding, answering questions, and engaging in discussions are integral aspects of assessment. This ensures that genuine comprehension and expression are assessed, going beyond the limitations of written assignments susceptible to AI manipulation.

Rethinking Language Assessment in Algerian Higher Education is an imperative task to adapt to the age of generative AI. The recommendations proposed aim to strike a balance between leveraging AI for educational enhancement and safeguarding the quality, effectiveness, and overall integrity of the teaching-learning process. By setting clear standards, adapting assessment methods, and emphasizing the essence of genuine learning, Algerian educators can cope with this technological evolution without compromising the foundational principles of education. It is through such proactive measures that the Algerian Higher Education Context can



successfully meet the challenges of the AI era while preparing a generation of authentically educated individuals.

#### **4.5 Enhancing Teacher Educational Development**

##### ***4.5.1 Training Programs***

Assessment literacy, a pivotal component of effective teaching, encapsulates teachers', test developers', and stakeholders' understanding of the measurement of students' learning. In the Algerian higher education context, where summative assessments play a prominent role, it is imperative to cultivate assessment literacy among educators. A comprehensive framework for teachers' training and development needs to be delineated drawing on insights from international standards and research on assessment literacy.

To establish a robust foundation, teacher training programs should encompass the three dimensions of language assessment literacy proposed by Davies (2008) - knowledge, principles, and skills. Knowledge entails an awareness of measurement and language; principles involve ethical practices, validity, and reliability, while skills focus on training in appropriate assessment methodologies (Fulcher, 2012). A holistic approach to assessment literacy should integrate these dimensions, empowering ESP practitioners with a multifaceted understanding of assessment.

Aligning with global best practices, teacher training initiatives should incorporate the eight prominent research-based assessment measures identified by DeLuca et al. (2016a) and stated in the literature review. These measures range from understanding assessment purposes to fostering assessment ethics. Ensuring teachers' competence in these areas not only enhances their capacity to take informed decisions but also reinforces the ethical and fair conduct of

assessments. Acknowledging the evolution of assessment literacy standards, training programs must address contemporary approaches, emphasizing competencies such as ICL and LOA design (DeLuca et al., 2016a). A forward-looking training model should prepare educators to navigate the dynamic landscape of assessment, incorporating up-to-date and learner-centered approaches.

Recognizing the role of assessment-literate educators in the teaching-learning experience, training programs should emphasize their impact. Research indicates that despite the recognition of the importance of assessment literacy, many educators struggle to articulate research-based knowledge (Davidson & Coombe, 2019). Algerian Higher Education training initiatives should bridge this gap, providing teachers with practical tools and strategies, such as assessment blueprints and informative videos, to enhance their assessment literacy. Boubris and Haddam (2020) stress the impact of teachers' beliefs, often subjective and experience-based, on teaching, learning, and assessment. Effective teacher training should include modules that address these beliefs, fostering an objective understanding of assessment. This involves creating awareness of the potential biases in beliefs and instilling evidence-based practices.

To promote assessment literacy, teacher development programs should integrate expert trainers and advisors. This involves providing teachers with an authentic opportunity to understand the various types of assessment and their inherent purposes, in addition, to their design, operationalization, delivery, and scoring. The collaboration with language assessment (LA) experts holds immense potential. This collaboration can yield valuable assessment blueprints and artifacts that can eventually serve several purposes. The artifacts, crafted with expertise, can be utilized as mock tests, strategically integrating assessment-driven approaches into the learning process. These mock tests become invaluable tools, supporting learners in their

targeted language proficiency development. The practical application of assessment artifacts in the learning environment enhances students' familiarity with authentic and valid assessment formats, aligning their language learning with domain-specific proficiency goals.

Moreover, the assessment blueprints produced in collaboration with LA experts can extend their utility beyond individual assessments. These blueprints, which encapsulate the essential elements of effective language assessment, can be incorporated into pedagogical manuals. Teachers, particularly those in need of support in designing English for Specific Purposes (ESP) instruction and evaluation, can benefit from these guides. By integrating the blueprints into manuals, educators receive structured guidance, ensuring that ESP instruction is aligned with rigorous assessment standards. This collaborative approach not only enriches teacher expertise but also reinforces a cohesive and standards-driven approach to language instruction in the Algerian higher education context.

Training programs must also equip educators with digital AI literacy skills to effectively integrate AI into their teaching practices. Training programs, online courses, and continuous engagement with AI literature are essential elements of upskilling. The development of a community of practice focused on AI literacy can facilitate knowledge sharing and keep educators abreast of the latest advancements. As Algerian higher education embraces generative AI, investing in the digital literacy of educators becomes paramount to harness the full potential of these tools responsibly.

Most importantly, teacher training programs should instill in educators the ability to advocate for improvements in the assessment system. This requires comprehending the strengths and limitations of the system and acknowledging how it affects both student motivation and

learning results. Training should empower educators to actively contribute to discussions on enhancing the assessment system for better educational outcomes. A robust teacher training and development program in Algerian higher education should go beyond imparting theoretical knowledge. It should focus on practical competencies, ethical considerations, and the evolving landscape of assessment. Teachers' initiatives should be incentivized to foster a competitive and innovative atmosphere. By directly, or indirectly, addressing ESP practitioners' assessment literacy, such programs can empower educators to navigate the complexities of assessments effectively, thereby enhancing the overall quality of teaching and learning experiences in higher education.

#### ***4.5.2 The Algerian ESP Centre***

The establishment of English for Specific Purposes (ESP) Centre in Algeria holds immense potential for fostering educational advancement, professional development, and economic growth. As Assassi (2021) reported, the initial intent of the first ESP Centre project was to facilitate Algerian students' enrollment in full-time postgraduate programs in collaboration with British institutions. The dissolution of these ESP centers, as highlighted in his paper, underscores legal issues encountered during the initiative. A Reasonable argument would be in favor of reinstating and reinforcing ESP Centre projects in Algeria for the numerous advantages it can provide.

Based on Assassi (2020), the ESP Centre aligns perfectly with the demands of a transitional period in Algeria, marked by economic openness and a surge in foreign businesses. The increased demand for specialized classes in fields such as tourism, industry, agriculture, and computing sciences underscores the urgency of re-establishing ESP Centers to cater to evolving

educational needs. The ESP Centre can produce useful material, such as instructional and evaluative materials, informative videos, and individual support to ESP practitioners, including subject-matter instructors.

On top of addressing the pedagogical shortcomings of the ESP situation in the Algerian Higher Education context, the ESP Centre has the capacity to compensate for the deficit of permanent ESP practitioners. It should also strive to create a collaborative environment between language and subject-matter instructors to participate in the enhancement of the overall quality of the teaching and learning experience, in the midst of Algerian HEIs shift towards English as a Medium for Instruction (EMI).

The ESP Centre, on the other hand, should benefit from financial support to incentivize contributors and boost its overall impact on education and society. The ESP Centre emerges as a strategic move, resulting in tailored syllabi, curricula, and training programs, in addition to the creation of an ESP network and collaboration opportunities. These position it as a crucial component in fulfilling many of the objectives to which the Algerian Higher Education stakeholders, including learners, are aiming.

#### **4.6 Fostering Research on ESP Assessment**

The necessity for more research on ESP assessment practices in the Algerian Higher Education emerges as a pivotal avenue for academic, economic, and societal advancements. The following recommendations delineate potential research trajectories that align with the nature of ESP assessment, aiming to contribute substantively to the enhancement of educational practices and the socio-economic fabric of Algeria.

The foundation for robust research on ESP assessment in Algerian Higher Education lies in conducting descriptive, exploratory, and comparative studies across diverse ESP contexts. Researchers should undertake systematic investigations into evaluative practices and student performances, participating in the creation a comprehensive database. By comparing findings across various contexts, a nuanced and holistic picture of ESP assessment in Algeria can be developed. This comparative approach enables researchers to identify characteristics, best practices, challenges, and contextual nuances, laying the groundwork for evidence-based improvements. These studies can be informed by the comprehensive framework proposed by Bachman and Palmer (2010), encompassing various assessment standards such as reliability, validity, and impact.

Another exciting avenue for future research involves the implementation of action research, focusing on innovative assessment designs within the Algerian Higher Education. Researchers should actively engage in designing and implementing novel assessment approaches, gauging their effectiveness in real-world educational settings. This approach allows for the exploration of innovative assessment methodologies and the subsequent research into their impact and outcomes. Such research not only fosters continuous improvement but also sets the stage for simultaneous corrective measures, leading to assessment reforms and paradigm shifts in ESP assessment practices.

To truly gauge the effectiveness of ESP assessments, there is also a need for long-term research that delves into the impact of these assessments on students' language development and career readiness over time. Beyond the realm of academic assessment, researchers should incorporate a socio-economic impact analysis into ESP assessment studies. This involves

evaluating the broader implications of assessment practices on the socio-economic fabric of Algeria.

Researchers can explore how effective ESP assessments contribute to the development of a skilled workforce, subsequently influencing economic growth and societal advancement. For instance, researchers can conduct predictive validity studies to assess the ability of ESP assessments to predict future language proficiency and career success. This longitudinal perspective provides valuable insights into the enduring impact of assessment practices on students' academic and professional trajectories. Additionally, it offers a basis for discussing the overall usefulness of ESP assessments, paving the way for strategic changes that align with the evolving socio-economic landscape of Algeria.

Another area where research on LA in the Algerian context is lacking is in assessment literacy. Assessment literacy refers to the knowledge and skills that teachers and students need to develop, use, and interpret assessments effectively. This encompasses comprehending the various forms of assessments, their objectives, and the proper methods for selecting and employing them effectively (Ashraf & Zolfaghari, 2018). Research is needed to determine whether the Algerian teachers possess the necessary assessment literacy skills, and the factors that may interfere with their development, such as limited training in assessment methods, lack of access to resources, and heavy workloads (Davidson & Coombe, 2019).

The recommendations provided offer a comprehensive roadmap for advancing research on ESP assessment in Algerian Higher Education. By embracing diverse research methodologies and exploring the long-term impact of evaluative practices, researchers can actively participate in framing educational policies, fostering innovation, and contributing to the socio-economic

progress of Algeria. This collaborative effort between researchers, educators, and policymakers holds the potential to catalyze positive transformations in the overall Algerian ESP situation.

#### **4.7 Conclusion**

Building on the comprehensive analysis, this research lays the groundwork for actionable recommendations. Incorporating feedback provisions is identified as a key improvement area, addressing the absence of formative and learning-oriented dimensions. Diversifying assessment tools and moving away from standardized models is recommended to better align assessments with the specific needs of computer science students. Addressing issues related to assessment design, task specifications, and scoring methods is underscored, aiming to create a more robust and learner-centered evaluation process.

The findings are relevant for ESP practitioners, curriculum designers, and policymakers seeking to enhance language assessment practices in ESP contexts. Practicality-driven assessment practices underscore the need for resource considerations in large cohorts, urging institutions to address challenges in educator training and retention. The study's focus on a specific context (Algerian Higher Education) may limit its generalizability to other ESP contexts. The study acknowledges practicality constraints, but the depth of these constraints and their impact on assessment design could benefit from further exploration. Future research should explore innovative assessment designs that integrate ICL learning-oriented elements, ensuring a balance between practicality and the validity of assessments.

The findings offer valuable insights into the challenges and opportunities in Algerian ESP assessment, paving the way for improved practices and future research endeavors. The study underscores the pivotal role of teacher educational development through training programs and



the creation of an ESP center. The detailed discussions, in the third chapter, highlight the intricacies of ESP assessment for 1st Year Computer Science students. It underscores the need for a paradigm shift in assessment practices, emphasizing ICL nature, learning-oriented design, and embedded alignment with the evolving needs of computer science professionals. The present research work sets the stage for meaningful improvements that can enhance the overall learning experience and proficiency development of students in the Algerian ESP context.

# **GENERAL CONCLUSION**

Worldwide, Language assessment (LA) in English for Specific Purposes (ESP) contexts is witnessing a paradigm shift towards Integrated Content and Language (ICL) methodologies and Learning-oriented Assessment (LOA) for their evidenced impact on learners' linguistic and topical skills development. As far as the Algerian Higher Education Institutions (HEIs) context is concerned, the whys and the whens of ESP assessments are clearly mandated through text laws and administrative circulars. However, very little information is communicated about how to carry out evaluative practices in ESP. This results in a one-size-fits-all tradition, depriving assessment from its formative, learning-oriented, and domain-reflective functions.

Moreover, the complexity of assessment theory and the challenges identified by other researchers in relation to the Algerian context portend dysfunctionalities in ESP assessment design, operationalization, delivery, and scoring, in addition to its overall quality and alignment with learners' domain-related language learning goals. Hence, the motive driving this research was to explore ESP teachers' evaluative practices and their potential effect on language skills development among 1st year Computer Science students, at Tlemcen University. As those students face many difficulties in integrating linguistic and topical skills in domain-specific situations, the researcher believes that ICL instruction and assessment can improve their overall language proficiency in the field-specific Target Language Use (TLU) domains.

The work was initiated by providing the theoretical background underlying the multitude of assessment-related concepts needed to address the research questions. In the first chapter we laid the ground for ICL formative, summative, and learning-oriented assessments, with a glimpse into assessment engineering, assessment standards, and assessment literacy. The second chapter was devoted to the description of the ESP situation in the Algerian HEIs; besides a thorough description of the research design was offered. In chapter three, the main research findings were exposed, analyzed, and discussed in relation to the research questions and hypotheses. The fourth

and final chapter was concerned with providing a set of recommendations which are likely to enhance ESP assessment procedures, and the overall ESP situation in the Algerian HE context.

The research adhered to a sequential exploratory design, with two major phases. Initially, the researcher explored ESP teachers' evaluative practices, at the level of the Faculty of Sciences, through document analysis and classroom observation. The qualitative data collected throughout the first phase grounded the description of ESP assessment situation, including its potential impact on the development of learners' domain-related language skills, as well as the design of the quantitative instruments deployed during the second phase. The quantitative phase, combining findings from a survey addressed to all undergraduate Computer Science learners and an experimental test addressed uniquely to first-year students, helped in delineating the potential impact of ESP assessment methodologies and practices on the development of domain-related language skills among Computer Science learners and their overall satisfaction with the ESP assessment situation.

Regarding the first research question which was concerned with the current ESP evaluative practices, findings from the first phase of the research have shown that the assessment of 1st Year Computer Science students' language proficiency at the Faculty of Sciences is characterized by certain patterns and challenges. The assessment practices primarily involve paper-based summative evaluations, lasting 60 minutes. The absence of formative assessment and feedback provision has revealed as a significant concern, potentially hindering students' learning and improvement. The design process, as observed in various artifacts, tends to lack iteration, leading to redundant repetitions of tasks and items rather than genuine enhancement or innovation. Issues, related to untrained, unmotivated, and frequently replaced educators, may be at the root of the misalignment between assessment content and the specific needs of computer science students. Some assessments appear more suitable for an EMI mathematics class, indicating a need for greater integration of relevant topical and linguistic components in ESP

exams. Assessment tasks typically cover vocabulary, grammar, reading, and writing, but there is a conspicuous absence of listening and speaking constructs. This raises questions about the validity, reliability, authenticity, and impact of the assessments, particularly in the context of real-world demands on computer science professionals. The qualitative findings from the first phase support the first research hypothesis.

To answer the second research question, which was concerned with undergraduate Computer Science experience and satisfaction with ESP assessment situations, the learners' survey was used to gather relevant data. The Dysfunctional assessment design was shown to impact students' perception of the evaluation process, potentially leading to frustration, dissatisfaction, and disengagement. Computer Science students' experiences with ESP assessments, at the Faculty of Sciences, demonstrate a generally effective and practical administration of the English Final Exam across different academic levels. While concerns regarding exam organization and invigilation exist for about a quarter of participants, the majority feel they have sufficient time to complete the exams. However, a notable issue is highlighted concerning learners' comprehension of instructions, suggesting potential challenges in assessment design and task specifications. The survey unveiled shortcomings in addressing receptive language skills, particularly in listening assessment. The evaluation of reading skills focuses predominantly on direct functional meanings within texts, potentially limiting the assessment of higher-order reading skills. In terms of productive language skills, the assessment of writing skills is criticized for emphasizing grammatical accuracy over diverse genres, hindering the development of proficient writing abilities. Strikingly, speaking skills are not subject to assessment, impacting students' motivation and development. The survey findings indicated a significant misalignment between ESP assessments and learners' actual needs within the Computer Science domain, as evidenced by the scarcity of positive opinions regarding the exams' contribution to achieving domain-related language learning goals. This misalignment, coupled with recurrent testing traditions and deficiencies in addressing language skills needed in

both academic and professional contexts, adversely affects students' motivation and priorities in language learning. These findings confirm the second research hypothesis.

The third and final research question addressed the impact of ESP assessments practices on 1st-year Computer Science students. This impact was evidenced through a detailed analysis of various measures and test results. The measures employed as performance indicators revealed that the, learning-oriented by design, ICL Test effectively homogenized learners' collective performance around the mode. The Generic Test, inspired from the findings of the qualitative phase, however, fails to have a significant impact on learners' initial language proficiency level. The impact of task formats, specifically open-ended questions versus MCQs, becomes apparent, with MCQs contributing to improved performance in reading comprehension tasks. Interestingly, the advantage of MCQs is reversed in grammar tasks, where the ICL Test group outperforms the Generic Test group, highlighting the importance of task specificity and feedback provisions. The ICL Test, characterized by an ICL task prioritizing topical knowledge, shows positive impacts on exam scores and grades, particularly in the culminating competency. Feedback provisions in the ICL Test contribute to students' rectification of errors, improvement, and skill development, as opposed to the Generic Test, which lacks such features. The integration of assessment into Computer Science discourse practice, aligned with the LOA framework and the Meaning-Oriented Model of L2 Proficiency (MOM), significantly enhanced the assessment experience and positively impacted learners' mobilization and development of both topical and linguistic skills. These findings, supporting the third research hypothesis, proved that the ICL Test's emphasis on learning-oriented tasks motivated learners and boosted their overall performance, learning, and grades compared to the Generic Test.

The assessment of English for Specific Purposes (ESP) among 1st Year Computer Science students at the Faculty of Sciences in Algeria is a critical aspect of their academic and professional journey. The study highlights the broader challenges in ESP courses, particularly the recurring issue of a lack of alignment between ESP courses and learners' professional needs.

The impact of dysfunctional assessment practices extends beyond academic performance, influencing students' perception of the evaluation process and hindering their overall learning and development experience. The predominant use of summative assessments, coupled with the absence of feedback provisions, poses a significant challenge to students' learning and improvement. Thematic analysis exposes a concerning pattern of redundancy and potential lack of innovation in instructional materials, limiting the effectiveness of ESP assessments. However, amidst these challenges, the study identifies the Integrated Content and Language (ICL) Test as a positive outlier, demonstrating a meaningful impact on learners' performance and engagement. The dissatisfaction expressed by learners regarding the inadequacy of ESP assessments in preparing them for language use within the academic and professional domains of Computer Science underscores the urgency for adjustments in assessment practices. Introducing learning-oriented assessments, diversifying assessment tools, and focusing on learner-centered design are proposed as measures to transform the impact of ESP assessments, facilitating language skill development and contributing to a more constructive and engaging learning experience aligned with the evolving needs of Computer Science. The study emphasizes the imperative need for reform to enhance the overall learning experience and proficiency development of students in the ESP context, ensuring their readiness for academic and professional success.

The research findings culminated in a set of comprehensive recommendations aimed at improving ESP assessment practices for ESP students in the Algerian HE context. Firstly, the study emphasized the potential of formative assessment and proposed its incorporation to enhance students' learning and improvement. This includes enriching classroom assessment techniques and boosting their effectiveness. Additionally, there was a strong emphasis on incorporating learning-oriented assessment (LOA) frameworks into assessment design. Recommendations in this area span developing LOA, focusing on performance moderators, and incorporating constructive feedback. Diversifying assessment tools is another key recommendation, with a specific focus on the integration of Integrated Content and Language

(ICL) assessments, addressing the need for listening and speaking evaluations, improving ESP classroom logistics, designing useful ICL tests, and rethinking assessment delivery in the age of technology and Artificial Intelligence (AI). To support these changes, the study advocates for the enhancement of teacher educational development through training programs and the establishment of an Algerian ESP Centre. Lastly, fostering research on ESP assessment is encouraged to continually inform and refine assessment practices in alignment with the dynamic needs of Algerian ESP students.

Despite the significance of the present research work, it is essential to acknowledge its inherent limitations. Firstly, the study does not extensively investigate the underlying issues that hinder ESP assessment in the Algerian context. Instead, it primarily concentrates on understanding the tangible impact of ESP assessment practices and endeavors to offer practical support based on existing literature and empirical findings. Secondly, the research lacks the inclusion of previous records of students' achievements to infer their linguistic proficiency levels and to provide additional context for some of its findings. Moreover, the researcher encountered some difficulties related to teachers' willingness to participate in the research by granting access to their classrooms for observation. Finally, the study's scope is confined to Computer Science students, at the Faculty of Sciences, Tlemcen University, Algeria, which might restrict the generalizability of some of its findings to broader contexts.

The landscape of Language Assessment in English for Specific Purposes is undergoing a transformative shift globally, marked by a growing emphasis on Integrated Content and Language methodologies, Learning-oriented Assessment frameworks, and models of L2 proficiency. While the Algerian Higher Education Institutions have clear mandates on the when and why of ESP assessments, the lack of guidance on how to conduct evaluative practices results in a conventional, one-size-fits-all approach. This research, driven by the aim to explore ESP teachers' evaluative practices and their impact on language skills development among 1st-year Computer Science students, uncovered patterns and challenges in ESP assessment. It emphasized



the need for a paradigm shift towards formative, learning-oriented, and domain-reflective assessments. Despite the identified challenges, the study recognizes the positive impact learning-oriented ICL assessments, highlighting their potential of to enhance both topical and linguistic knowledge, skills, and abilities among learners. The study concludes with a call for comprehensive reforms, offering a detailed set of recommendations. This comprehensive roadmap, fueled by a commitment to innovation and learner-centered design, aims to not only address the current challenges but also pave the way for a more constructive and engaging learning experience in ESP, aligned with the evolving needs of ESP students in Algeria. The study, in essence, sets to initiate the call for transformative change, inspiring educators, administrators, and researchers to collectively elevate ESP assessment practices to new heights.

# **BIBLIOGRAPHY**

- Abbassi, A. (2022). A Reflection on Language Assessment in ESP: Instructors' and Learners' Perceptions in Focus. *مجلة بحالات*, 4(1), 12-23.
- Abd-Elaal, E. S., Gamage, S. H., & Mills, J. E. (2019, January). Artificial intelligence is a tool for cheating academic integrity. *In 30th Annual Conference for the Australasian Association for Engineering Education (AAEE 2019): Educators becoming agents of change: Innovate, integrate, motivate* (pp. 397-403).
- Acharya, A. S., Prakash, A., Saxena, P., & Nigam, A. (2013). Sampling: Why and how of it. *Indian Journal of Medical Specialties*, 4(2), 330-333.
- Adams, P. (2004). Classroom assessment and social welfare policy: Addressing challenges to teaching and learning. *Journal of Social Work Education* 40(1), 121-42.
- Afshar, H. S., Tofighi, S., Asoudeh, M., & Ranjbar, N. (2018). The Impact of Alternative Assessment Knowledge, Teaching Experience, Gender, and Academic Degree on EAP Teachers' Assessment Literacy. *Iranian Journal of English for Academic Purposes*, 7(2), 1-15.
- Aftab, A., Qureshi, S., & William, I. (2014). Investigating the washback effect of the Pakistani Intermediate English Examination. *International Journal of English and Literature*, 5(7), 149-154.
- Ahmed, R., & Rahman, K. A. (2019). Review of empirical studies of washback in language testing. *BMJ*, 3(1), 150-162.
- Akkar, S., & Idri, N. (2021). A Diagnostic View on the Status of ESP at the Vocational Training Institute of Bejaia, Algeria: The Case of Health, Safety and Environment Trainees. *Algerian Journal of Manuscripts*, 17, Special Issue, 136-155
- Alam, M., & Aktar, T. (2019). Assessment Challenges & Impact of Formative Portfolio Assessment (FPA) on EFL Learners' Writing Performance: A Case Study on the Preparatory English Language Course. *English Language Teaching*, 12(7), 161-172.
- Alaye, A., & Tegegne, W. (2019). Critical review on the meaning, purposes and techniques of integrative language skills teaching approach. *Critical Review*, 58, 58-03.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Algozzine, B., & Hancock, D. (2017). *Doing case study research: A practical guide for beginning researchers*. Teachers College Press.
- Alias, M. (2005). Assessment of learning outcomes: validity and reliability of classroom tests. *World Transactions on Engineering and Technology Education*, 4(2), 235.
- Allen, D. (2016). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language testing in Asia*, 6(1), 1-20.

- Almeida, F. (2018). Strategies to perform a mixed methods study. *European Journal of Education Studies*.
- Amaliani, P. (2017). *The effect of using KWL chart technique on students' reading comprehension of descriptive text*. (Unpublished master's thesis). Syarif Hidayatullah Jakarta University, Jakarta, Indonesia.
- Anderson, L.W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman, New York
- Andrade, H.L., Cizek, G.J. (Eds.), (2010). *Handbook of Formative Assessment*. Routledge, Oxford.
- Aouine, A. (2011). *English Language Assessment in the Algerian Middle and Secondary Schools: A Context Evaluation* (Doctoral dissertation, Mouloud MAMMERI University of Tizi-Ouzou Faculty of Letters and Humanities).
- Armstrong, M. C. F., Marín, M. I. R., & Patarroyo, L. M. (2004). Analysis of assessment instruments used in foreign language teaching. *Íkala, revista de lenguaje y cultura*, 9(15), 171-201.
- Ashraf, H., & Zolfaghari, S. (2018). EFL Teachers' Assessment Literacy and Their Reflective Teaching. *International Journal of Instruction*, 11(1), 425-436.
- Assassi, T. (2020). The Status of ESP in Algeria: The Need for Highly Specialized Courses of English. *Humanization Journal for Researches and Studies*. 11 (20), 439-455.
- Atamena, N. (2023). Innovative Teaching Strategies in the Digitalized World during the Covid-19 Epoch: The Case of Islamic Studies. *ASJP*.
- Babbie, E. R. (2010). Quantitative methods. *The Practice of Social Research*. 12th Ed. Belmont, Ca.
- Bachman, L., & Adrian, P. (2022). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Bachman, L. F., & Damböck, B. (2018). *Language assessment for classroom teachers*. Oxford, UK: Oxford University Press.
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in Practice*. Oxford: Oxford University Press.

- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Barnes, M. (2017). Washback: Exploring what constitutes “good” teaching practices. *Journal of English for Academic Purposes*, 30, 1-12.
- Beikmahdavi, N. (2016). Washback in language testing: Review of related literature first. *International Journal of Modern Language Teaching and Learning*, 1(4), 130-136.
- Belmihoub, K. (2018). English in a multilingual Algeria. *World Englishes*, 37(2), 207-227.
- Benabdi, F. (2022). ESP Course Design and Teaching Professional Development in Algeria. *ALTRALANG Journal*, 4(02), 447-456.
- Benmoussat, S., & Benmoussat, N. D. (2018). ELT in Algeria: The Hegemony of the Teach-to-the-Test Approach. *English Language and Literature Studies*, 8(2), 63-68.
- Benmoussat, N. D., Benmoussat, S. (2017). The Teach-to-the-Test Approach: A Curse a Blessing or a Blessing in Disguise for Algerian EFL Students. *IOSR Journal of Research & Method in Education (IOSR-JRME)*, 7(6), 69-73.
- Berman, E. A. (2017). An exploratory sequential mixed methods approach to understanding researchers’ data management practices at UVM: Integrated findings to develop research data services. *JeSLIB*, 6(1), 1-24. DOI: e1104 doi:10.7191/jeslib.2017.1104
- Bezziou, M. D., & Ahmed, B. (2013). Investigating Washback Effects on The Students’ Performance in Standardized Tests.
- Bin Mubayrik, H. F. (2020). New trends in formative-summative evaluations for adult education. *Sage Open*, 10(3), 2158244020941006.
- Bonner, S. M. (2013). Validity in classroom assessment: Purposes, properties, and principles. *SAGE handbook of research on classroom assessment*, 87-106.
- Bouabdallah, F. H. (2015). Integrating Technology in the ESP Classroom in the Department of Biology–Tlemcen University (Algeria). In *International Conference on Language Learning and Teaching, HCT Dubai Men’s College, UAE*. <https://www.21caf.org/uploads/1/3/5/2/13527682/76-84.hadda.m.pdf>.
- Boubris, A. A., & Bouabdallah, F. H. (2023). ICLHE Task Design: Case of L1 Computer Science, Tlemcen University. *Revue plurilingue : Études des Langues, Littératures et Cultures*, 7(1).
- Boubris, A. A., & Haddam, F. (2020). Reading Assessment: A Case Study of Teachers’ Beliefs and Classroom Evaluative Practices. *Arab World English Journal (AWEJ)*, 11(4), 236-253. DOI: <https://dx.doi.org/10.24093/awej/vol11no4.16>

- Boumediene, A., & Hamzaoui – Elachachi (2017). The Effects of Formative Assessment on Algerian Secondary School Pupils' Text Comprehension. *Arab World English Journal*, 8 (3). DOI: <https://dx.doi.org/10.24093/awej/vol8no3.12>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal*, 9(2), 27-40.
- Berndt, A. E. (2020). Sampling methods. *Journal of Human Lactation*, 36(2), 224-226.
- Brookhart, S.M. (2009). *Grading, second ed.* Merrill, New York
- Brown, H. D. (2003). *Language assessment: Principles and classroom practices.* Longman.
- Bryman, A. (2012). *Social research methods.* New York: Oxford University Press.
- Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, 69, 963-976.
- Carr, N. (2011). *Designing and Analyzing Language Tests.* Oxford, UK: Oxford University Press.
- Cerezo, L. (2011). Introducing language assessment and testing. Inglés: Investigación, innovación y buenas prácticas. *Teacher development*, 119-142.
- Charvade, M. R., Jahandar, S., & Khodabandehlou, M. (2012). The Impact of Portfolio Assessment on EFL Learners' Reading Comprehension Ability. *English Language Teaching*, 5(7), 129-139.
- Cheng, L, Watanabe, Y., & Curtis, A. (2004) (Eds.), Washback in language testing. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study.* *Studies in language testing*, 21. Cambridge: Cambridge University Press.
- Cheng, L., & Curtis, A. (2012). Test impact and washback: Implications for teaching and learning. *Cambridge guide to second language assessment*, 89-95.
- Cilliers, F. J., Schuwirth, L. W., Adendorff, H. J., Herman, N., & Van der Vleuten, C. P. (2010). The mechanism of impact of summative assessment on medical students' learning. *Advances in health sciences education*, 15(5), 695-715.
- Cohen, A. D. (2021). Test-taking strategies and task design. In *The Routledge handbook of language testing* (pp. 372-396). Routledge.
- Cohen, L., Manion, L., & Morrison, K. (2002). *Research methods in education.* Routledge.
- Cohen, L., Manion, L., & Morrison, K. (2017a). Observation. In *Research methods in education* (pp. 542-562). Routledge.
- Cohen, L., Manion, L., & Morrison, K. (2017b). Tests. In *Research methods in education* (pp. 563-585). Routledge.

- Conderman, G., & Hedin, L. (2012). Classroom assessments that inform instruction. *Kappa Delta Pi Record*, 48(4), 162-168. doi:10.1080/00228958.2012.733964
- Cong-Lem, N. (2019). Portfolios as learning and alternative-assessment tools in EFL context: A review. *CALL-EJ*, 20(2), 165-180.
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: what do we need to learn, unlearn, and relearn? *Language Testing in Asia*, 10(1), 1-16.
- Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1-12.
- Cross, T., & Palese, K. (2015). Increasing learning: Classroom assessment techniques in the online classroom. *American Journal of Distance Education*, 29(2), 98-108.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25, 327-348.
- Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Sage.
- Davidson, F., & Lynch, B. K. (2008). *Testcraft: A teacher's guide to writing and using language test specifications*. Yale University Press.
- Davidson, P., & Coombe, C. (2019). Language Assessment Literacy in the Middle East and North Africa (MENA) Region. *Arab Journal of Applied Linguistics*, 4(2), 1-23.
- Dawadi, S. (2021). Factors Influencing Washback of a High-stakes English-as-a-Foreign-Language Test. *TESL-EJ*, 25(3).
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, 21(4), 248-266. DOI: 10.1080/10627197.2016.1236677
- DeLuca, C., LaPointe-McEwan, D. & Luhanga, U. (2016b). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251-272. DOI: <https://doi.org/10.1007/s11092-015-9233-6>
- Denzin, N. K., & Lincoln, Y. S. (2011). *The SAGE handbook of qualitative research*. London: SAGE
- Desheng, C., & Varghese, A. (2013). Testing and Evaluation of Language Skills. *IOSR Journal of Research & Method in Education*, 1(2), 31-33.
- Deygers, B., & Malone, M. E. (2019). Language assessment literacy in university admission policies, or the dialogue that isn't. *Language Testing*, 36(3), 347-368.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. John Wiley & Sons.

- Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into practice*, 55(2), 153-159.
- Dolin, J., Black, P., Harlen, W., & Tiberghien, A. (2018). Exploring relations between formative and summative assessment. In *Transforming assessment* (pp. 53-80). Springer, Cham.
- Donmoyer, R. (2012). Can qualitative researchers answer policymakers' what-works question? *Qualitative Inquiry*, 18(8), 662-673. DOI: [http](http://dx.doi.org/10.1177/1075547012468111)
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.
- Douglas, D. (2012). ESP and assessment. *The handbook of English for specific purposes*, 367-383.
- Elyza, F. (2018). PORTFOLIO ASSESSMENT IMPACT TO THE STUDENTS' PARAGRAPH WRITING. *Visipena*, 9(1), 193-202.
- Faiza, H. (2010). Conducting a Needs Analysis for an ESP Course Design: The Case of Fourth Year Students of Cellular and Molecular Biology. *International Journal of the Humanities*, 8(9).
- Fehaima, A. (2022). Translation-based task in an ESP classroom: Tlemcen University as a case study. *Contemporary Educational Research Journal*, 12(4), 246-255.
- Flick, U. (2014). *An introduction to qualitative research* (5th ed.). London: Sage Publications Ltd.
- Fowler Jr, F. J. (2013). *Survey research methods*. Sage publications.
- Fulcher, G. (2012). Assessment Literacy for the Language Classroom. *Language Assessment Quarterly*, 9(2), 113-132.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. New York: Routledge.
- Gogo, S., & Musonda, I. (2022). The Use of the Exploratory Sequential Approach in Mixed-Method Research: A Case of Contextual Top Leadership Interventions in Construction H&S. *International journal of environmental research and public health*, 19(12), 7276.
- Göçtü, R. (2013). Comparison of traditional and alternative assessment in English Language Teaching in high schools (Georgian case). *Unpublished doctoral thesis, University of International Black Sea, Tbilisi, Georgia*.
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile Issues in Teachers Professional Development*, 20(1), 179-195.
- Giraldo, F. (2019). Language Assessment Practices and Beliefs: Implications for Language Assessment Literacy. *HOW*, 26(1), 35-61. <https://doi.org/10.19183/how.26.1.481>
- Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, 13(2), 39-51.



- Green, R. (2017). *Designing listening tests: A practical approach*. London: Palgrave Macmillan.
- Hadj Djelloul, K., & Melouk, M. (2022). Meeting Students' Needs: ESP Teaching at the Department of Political Sciences. *African Educational Research Journal*, 10(2), 150-160.
- Hakim, P. K., & Srisudarso, M. (2020). A washback study on portfolio assessment. *ELT in Focus*, 3(1), 9-14, Doi: 10.35706/eltinf.v3i1.3693
- Hampel, R. (2010). Task design for a virtual learning environment in a distance language course. *Task-based language learning and teaching with technology*, 131-153.
- Hanson, J. M., & Florestano, M. (2020). Classroom assessment techniques: A critical component for effective instruction. *New Directions for Teaching and Learning*, 2020(164), 49-56.
- Haque, M. (2010). Sampling methods in social research. *Global Research Methodology*, 8(5), 1-6.
- Hassan, A., Ali, S. A., & Abdulkhaleq, N. M. S. (2022, May). Artificial intelligence in educational examinations. In *European, Asian, Middle Eastern, North African Conference on Management & Information Systems* (pp. 73-85). Cham: Springer International Publishing.
- Hatipoğlu, Ç. 2016. The impact of the university entrance exam on EFL education in Turkey: Pre-service English language teachers' perspective. *Procedia-Social and Behavioral Sciences*, 232, 136-144.
- Hidri, S. (2018). Discrete point and integrative testing. *The TESOL encyclopedia of English language teaching*. John Wiley & Sons. <https://doi.org/10.1002/9781118784235.eelt0375>.
- Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language testing*, 29(3), 395-420.
- Hinds, D. (2002). Research instruments. In *The researcher's toolkit* (pp. 57-58). Routledge.
- Hoadjli, A. C. (2013). Investigating the Validity of Achievement Tests in the Algerian Secondary Schools: The Case of EFL Classes in the Biskra Region. *Revue des Sciences Humaines – Université Mohamed Khider Biskra*, 29.
- Hogan, A., & Daw, J. (2014). Mode of Delivery: A Classroom Assessment Technique Comparison between Verbal and Non-Verbal Communication. *Journal of Instructional Research*, 3, 21-26.
- Houston, D., & Thompson, J. N. (2017). Blending Formative and Summative Assessment in a Capstone Subject: 'It's not your tools, it's how you use them'. *Journal of University Teaching & Learning Practice*, 14(3), 2.
- Hughes, A. (1989) *Testing for language teachers*, Cambridge: Cambridge University Press.

- Hughes, A. (2003). *Testing for Language Teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Hussain, S., Tadesse, T., & Sajid, S. (2015). Norm-referenced and criterion-referenced test in EFL classroom. *International Journal of Humanities and Social Science Invention*, 4(10), 24-30.
- Hüseyin, Ö. Z., & Özturan, T. (2018). Computer-based and paper-based testing: Does the test administration mode influence the reliability and validity of achievement tests? *Journal of Language and Linguistic Studies*, 14(1), 67-85.
- Iraji, H. R., Enayat, M. J., & Momeni, M. (2016). The effects of self-and peer-assessment on Iranian EFL learners' argumentative writing performance. *Theory and Practice in Language Studies*, 6(4), 716-722.
- Jang, E. E., Wagner, M., & Park, G. (2014). Mixed Methods Research in Language Testing and Assessment. *Annual Review of Applied Linguistics*, 34, 123-153. DOI: <http://dx.doi.org/10.1017/S0267190514000063>
- Jannati, S. (2015). ELT Teachers' Language Assessment Literacy. *The International Journal of Research in Teacher Education*, 6(2), 26-37.
- Johnson, K. (2000). What task designers do. *Language teaching research*, 4(3), 301-321.
- Кошарна, Н. (2020). Students' Critical Thinking Skills Development withing Pedagogical Specialties of Master's Educational Level in the Content of the Discipline Professional Communication in a Foreign Language. *Освітологічний дискурс*, 2(29), 14-25.
- Kalra, R., Sundrarajun, C., & Komintarachat, H. (2017). Using portfolio as an alternative assessment tool to enhance Thai EFL students' writing skill. *Arab World English Journal (AWEJ)*, 8(4), 292-302. DOI: <https://doi.org/10.1017/S0267190517000063>
- Kamiya, M. (2006). The role of communicative competence in L2 learning. *Sophia Junior College Faculty Bulletin*, 26, 63-87.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: Praeger.
- Kane, M. T., & Wools, S. (2019). Perspectives on the validity of classroom assessments. In *Classroom assessment and educational measurement* (pp. 11-26). Routledge. DOI: [10.4324/9780429507533-2](https://doi.org/10.4324/9780429507533-2)
- Kavakli, N., & Arslan, S. (2019). Towards a Continuum from Know-How to Show-How for Developing EFL Student-Teachers' Assessment Literacy. *International Online Journal of Education and Teaching*, 6(1), 223-232.

- Khadam, G. (2023). ESP and Business English for Postgraduate Students in the Department of Economics, Tahri Mohamed University, Bechar, Algeria. *Arab World English Journal (AWEJ)*, 14(1), 476-485. DOI: <https://dx.doi.org/10.24093/awej/vol14no1.30>
- Kherraz, M. (2022). Needs Assessment: The Cornerstone of an Effective ESP Teaching. The case of Computer Science Master Students at Dr. Moulay Tahar University-Saida, Algeria. *ESP Teaching Today: Current Practices, Challenges*, 143-156.
- Kiliçkaya, F. (2016). Washback effects of a high-stakes exam on lower secondary school English teachers' practices in the classroom. *Lublin Studies in Modern Languages and Literature*, 40(1), 116-134.
- Kirkpatrick, R., & Zang, Y. (2011). The negative influences of exam-oriented education on Chinese high school students: Backwash from classroom to child. *Language testing in Asia*, 1(3), 1-10.
- Kirmizi, Ö, & Kömeç, F. (2016). An Investigation of Performance-Based Assessment at High Schools. *Üniversitepark Bülten*, 5(1-2), 53-65. DOI: <https://doi.org/10.22521/unibulletin.2016.512.5>
- Koh, K., Burke, L. E. C. A., Luke, A., Gong, W., & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language teaching research*, 22(3), 264-288.
- Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100-120.
- Lalaoua, M., Kouicem, O. D., & Saraa, L. (2021). The Washback Effect of the Baccalaureate Examination on Teaching Practices The Case of Third year Secondary School (Doctoral dissertation, Djijel University).
- Lam, D. M. (2021). Feedback as a learning-oriented assessment practice: Principles, opportunities, and challenges. *Learning-Oriented Language Assessment*, 85-106.
- Lane, S. (2015). Performance assessment: The state of the art. *Beyond the bubble test: How performance assessments support 21st century learning*, 131-184.
- Leavy, P. (2022). *Research design: Quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches*. Guilford Publications.
- Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, 17(2), 168-182.
- Lichtman, K., & VanPatten, B. (2021). Was Krashen right? Forty years later. *Foreign Language Annals*, 54(2), 283-305.

- Liu, X., & Yu, J. (2021). Relationships between learning motivations and practices as influenced by a high-stakes language test: The mechanism of washback on learning. *Studies in Educational Evaluation*, 68, 100967.
- McMillan, J.H., (2014). *Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction*, sixth ed. Pearson, Boston.
- McMillan, James H. (2015). *International Encyclopedia of the Social & Behavioral Sciences // Classroom Assessment.*, (), 819–824. doi:10.1016/b978-0-08-097086-8.92074-9
- McNamara, T. (2014). 30 years on—Evolution or revolution?. *Language Assessment Quarterly*, 11(2), 226-232.
- Maarouf, A. Z., & Lamouri, R. (2022). *Exploring the potential shift to English language in the Algerian higher education: Initial thoughts and expectations* (Doctoral dissertation, Université Ibn Khaldoun-Tiaret-).
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329-344.
- Mandi, A., & Kacha, K. (2023). *Investigating Teachers' formative assessment on students' academic performance: A case study of middle school Teachers at the Wilaya of Tissemsilt* (Doctoral dissertation, Université IBN KHALDOUN-Tiaret).
- Manias, E., & McNamara, T. (2015). Standard setting in specific-purpose language testing: What can a qualitative study add? *Language Testing*, 33(2), 235-249. DOI: <http://dx.doi.org/10.1177/0265532215608411>
- Mart, Ç. T. (2020). Integrating Listening and Speaking Skills to Promote Speech Production and Language Development. *Mextesol Journal*, 44(2), 2.
- Marzano, R.J., Kendall, J.S., 2007. *The New Taxonomy of Educational Objectives*, second ed. Corwin Press, Thousands Oaks, CA
- Meriem, O., & Bouyakoub, N. (2020). Teaching and testing English in the Algerian educational system. *Arab World English Journal (AWEJ) Volume*, 11.
- Mertler, C. A., & Campbell, C. (2005). Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the "Assessment Literacy Inventory". *Online Submission*.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*, 3rd ed. (pp. 13-103). New York: Macmillan. Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*, 3rd ed. (pp. 13-103). New York: Macmillan.

- Min, S., He, L., & Zhang, J. (2020). Review of recent empirical research (2011–2018) on language assessment in China. *Language Teaching*, 53(3), 316-340. doi:10.1017/S0261444820000051
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language testing*, 19(4), 477-496. DOI: <https://doi.org/10.1191/0265532202lt241oa>
- Monib, W. K., Karimi, A. Q., & Nijat, N. (2020). Effects of alternative assessment in EFL classroom: A systematic review. *American International Journal of Education and Linguistics Research*, 3(2), 7-18.
- Morgan, H. (2022). Conducting a qualitative document analysis. *The Qualitative Report*, 27(1), 64-77. DOI: <https://doi.org/10.46743/2160-3715/2022.5044>
- Moss, C. M. (2013). Research on classroom summative assessment. *SAGE handbook of research on classroom assessment*, 235-256.
- Murray, D. G. (2013). *Tableau your data!: fast and easy visual analysis with tableau software*. John Wiley & Sons.
- Nassirdoost, P., & Mall-Amiri, B. (2015). The Impact of portfolio assessment on EFL learners' vocabulary achievement and motivation. *Journal for the Study of English Linguistics*, 3(1), 38-50.
- Neff, P., & Rucynski Jr, J. (2013). Tasks for Integrating Language and Culture Teaching. *In English Teaching Forum* (Vol. 51, No. 2, pp. 12-23). US Department of State. Bureau of Educational and Cultural Affairs, Office of English Language Programs, SA-5, 2200 C Street NW 4th Floor, Washington, DC 20037.
- Nguyen, N. H. V., & Huynh, N. T. (2021, March). The Effects of Integrating Listening and Speaking Skills Into Moodle-Based Activities. In *Proceedings of the Asia CALL International Conference* (Vol. 533, No. 978-94-6239-343-1, pp. 166-174).
- Nimehchisalema, V., & Bhattib, N. (2019). A review of literature on language assessment literacy in the last two decades (1999-2018). *assessment*, 8(11), 44-59.
- Nodoushan, M. A. S. (2021). Washback or backwash? Revisiting the status quo of washback and test impact in EFL contexts. *Studies in English Language and Education*, 8(3), 869-884.
- Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, 51(9), 1352-1375. Doi: 10.1080/00140130802170387
- O'Donovan, B. M., den Outer, B., Price, M., & Lloyd, A. (2021). What makes good feedback good?. *Studies in Higher Education*, 46(2), 318-329.

- Ouahmiche, G., Beddiaf, A., & Beddiaf, A. (2017). Reflections on the Linguistic Landscape and the Prospects of English Language Teaching in Algeria. *International Journal of Language and Linguistics*, 5(3-1): 15-23. Doi: 10.11648/j.ijll.s.2017050301.13
- Olajide, S. B. (2018). Test as Tool for Effective Research in the Humanities. *Bulgarian Journal of Science and Education Policy*, 12(2), 402-417.
- Oo, T. Z., Nguyen, L. A. T., & Habók, A. (2021). Review of the Assessment of Students' Language Skills in English as a Foreign Language: Knowledge, Skills, and Principles. *The International Journal of Assessment and Evaluation*, 29(1), 1.
- Owen, L. (2016). The Impact of Feedback as Formative Assessment on Student Performance. *International Journal of Teaching and Learning in Higher Education*, 28(2), 168-175.
- Paltridge, B, & Starfield, S. (Eds). (2013). *The Handbook of English for Specific Purposes* (Vol.592). West-Sussex: Wiley-Blackwell.
- Papakammenou, I. R. (2018). Examining washback in EFL multi-exam preparation classes in Greece: A focus on teachers' teaching practices. In *Teacher involvement in high-stakes language testing* (pp. 321-339). Springer, Cham.
- Pardede, P. (2020). Integrating the 4Cs into EFL Integrated Skills Learning. *Journal of English Teaching*, 6(1), 71-85.
- Parkes, J. (2013). Reliability in classroom assessment. SAGE handbook of research on classroom assessment, 107-123.
- Perera, T. S. (2020). Alternative assessment: the motivational impact upon ESL learners. *GARI International Journal of Multidisciplinary Research*, 6(06), 60-66.
- Pierce, L. V. (2002). *Performance-Based Assessment: Promoting Achievement for Language Learners*. Center for Applied Linguistics, 26(01), 1-3.
- Popham, J. W. (2004). Why assessment illiteracy is professional suicide. *Educational Leadership*, 62(1), 82-83.
- Popham, W. J. (2013). *Classroom assessment: What teachers need to know* (7th ed.). Boston, MA: Pearson.
- Pratolo, B. W., & Zahrani, N. A. (2020). Dynamic Assessment Effect on Speaking Performance of Indonesian EFL Learners. *International Journal of Evaluation and Research in Education*, 9(3), 778-790.
- Purpura, J. E. (2004). *Assessing Grammar—Ch. 5--Designing test tasks to measure L2 grammatical ability*, pp. 100-145. Cambridge: Cambridge University Press.

- Purpura, J. E. (2004). *Assessing Grammar—Ch. 6: Learning-oriented assessments of grammatical ability*, pp. 156-180. Cambridge: Cambridge University Press.
- Purpura, J. E. (2011). Quantitative research methods in assessment and testing. *Handbook of Research in Second Language Teaching and Learning*, 2, 731-751. DOI: <http://dx.doi.org/10.4324/9780203836507.ch44>
- Purpura, J. E. (2017). Assessing meaning. In E. Shohamy & L. Or (Eds.), *Encyclopedia of Language and Education*, Vol. 7. *Language Testing and Assessment*. New York, NY: Springer International Publishing.
- Purpura, J. E., & Dakin, J. W. (2020). Assessment of the linguistic resources of communication. In C. Chapelle (Ed.), *The Concise Encyclopedia of Applied Linguistics: Assessment and Evaluation* (pp. 1-10). Oxford, UK: Wiley.
- Purpura, J. E., & Graziano-King, J. (2004). Investigating the foreign language needs of professional school students in international and public affairs: A case study. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 4(1), 1-33.
- Qu, W., & Zhang, C. (2013). The analysis of summative assessment and formative assessment and their roles in college English assessment system. *Journal of Language Teaching and Research*, 4(2), 335.
- Rad, M. R. (2019). The Impact of EFL Teachers' Assessment Literacy on Their Assessment Efficiency in Classroom. *Britain International of Linguistics Arts and Education (BioLAE) Journal*, 1(1), 9-17.
- Raffas, A. (2023). Language Assessment Literacy for Classroom Teachers: assessing Algerian university teachers' language assessment literacy. *Journal of Human and Society sciences*, 12(02), 455-476.
- Rahman, K., Seraj, P. M. I., Hasan, M., Namaziandost, E., & Tilwani, S. A. (2021). Washback of assessment on English teaching-learning practice at secondary schools. *Language Testing in Asia*, 11(1), 1-23.
- Rahman, M. S. (2017). The advantages and disadvantages of using qualitative and quantitative approaches and methods in language “testing and assessment” research: A literature review. *Journal of Education and learning*, 6(01), 102-112. DOI:10.5539/jel.v6n1p102
- Randel, B., & Clark, T. (2013). Measuring classroom assessment practices. *SAGE handbook of research on classroom assessment*, 145, 163.
- Rasinger, S. M. (2013). *Quantitative research in linguistics: An introduction*. A & C Black.

- Remadna, A. (2016). *The Role of Formative Assessment in Enhancing EFL Students' Writing Skill The case of: First year EFL Students at Mohamed Kheidher University of Biskra*. (Doctoral dissertation).
- Rossi, O. (2021). The language test development process: main principles and suggestions for classroom practice. *Testing, Evaluation, and Assessment Today*, 4, 53-57.
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1).
- Saraa, N. (2020). Teachers' Challenges while Designing an ESP Syllabus. *International Arab Journal of English for Specific Purposes*, 3(1), 83-92.
- Saraa, N. (2023). Probing ESP Teaching Context: Case of Biology Learners at Ibn-khaldoun University, Tiaret, Algeria. *Psychological & Educational Studies*, 16(2), 510-524.
- Şişman, E. P., & Büyükkarci, K. (2019). A review of foreign language teachers' assessment literacy. *Sakarya University Journal of Education*, 9(3), 628-650.
- Sallee, M. W., & Flood, J. T. (2012). Using qualitative research to bridge research, policy, and practice. *Theory Into Practice*, 51(2), 137-144. DOI: <http://dx.doi.org/10.1080/00405841.2012.662873>
- Saville, N. (2010). Developing a model for investigating the impact of language assessment. *Research Notes*, 42, 2-8.
- Schedl, M., & Malloy, J. (2013). Writing items and tasks. *The companion to language assessment*, 2, 796-813.
- Schissel, J. L. (2018). Test impact and washback. *The TESOL Encyclopedia of English language teaching*. DOI: 10.1002/9781118784235.eelt0353
- Schoonenboom, J., & Johnson, R. B. (2017). How to construct a mixed methods research design. *Kolner Zeitschrift fur Soziologie und Sozialpsychologie*, 69(Suppl 2), 107.
- Shalevska, E. (2023). AI LANGUAGE MODELS, STANDARDIZED TESTS, AND ACADEMIC INTEGRITY: A CHAT (GPT). *International journal of Education Teacher*, 26, 17-25.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-317.
- Shohamy, E. (2020). *The power of tests: A critical perspective on the uses of language tests*. Routledge.
- Simpson-Beck, V. (2011). Assessing classroom assessment techniques. *Active Learning in Higher Education*, 12(2), 125-132. doi:10.1177/1469787411402482



- Snow, M. A. (2005). A model of academic literacy for integrated language and content instruction. In *Handbook of research in second language teaching and learning* (pp. 693-712). Routledge.
- Soomro, A. H., & Shah, S. Z. A. (2016). Effects of washback on high school teachers of English. *International Journal of English and Education*, 6(2), 201-210.
- Spaan, M. (2006). Test and item specifications development. *Language Assessment Quarterly: An International Journal*, 3(1), 71-79.
- Starman, A. B. (2013). The case study as a type of qualitative research. *Journal of Contemporary Educational Studies/Sodobna Pedagogika*, 64(1).
- Stoynoff, S. (2012). Looking backward and forward at classroom-based language assessment. *ELT journal*, 66(4), 523-532.
- Sukyadi, D., & Mardiani, R. (2011). The washback effect of the English national examination (ENE) on English teachers' classroom teaching and students' learning. *k@ ta lama*, 13(1), 96-111.
- Sultana, N. (2018). A brief review of washback studies in the south Asian countries. *The Educational Review, USA*, 2(9), 468-474.
- Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9(1), 1-14.
- Taylor, G. R. (Ed.). (2005). *Integrating quantitative and qualitative methods in research*. University press of America.
- Taylor, L. (2014). A report on the review of test specifications for the reading and listening papers of the Test of English for Academic Purposes (TEAP) for Japanese university entrants. *Tokyo: Eiken Foundation of Japan*.
- Thomas, J., & Hornsey, P. E. (2014). Adding rigor to classroom assessment techniques for non-traditional adult programs: A lifecycle improvement approach. *Journal of Instructional Research*, 3.
- Tierney, R.D. (2013). Fairness in classroom assessment. In: McMillan, J.H. (Ed.), *Sage Handbook of Research on Classroom Assessment*. Sage Publishing, Thousand Oaks, CA, pp. 125-144.
- Toksöz, I., & Kilickaya, F. (2018). Review of journal articles on washback in language testing in Turkey (2010-2017). *Lublin Studies in Modern Languages and Literature*, 41(2), 184.
- Torres, J. O. (2019). Positive impact of utilizing more formative assessment over summative assessment in the EFL/ESL Classroom. *Open Journal of Modern Linguistics*, 9(1), 1-11.

- Tosuncuoglu, I. (2018). The Importance of Assessment in ELT. *Journal of Education and Training Studies*, 6(9). DOI: 10.11114/jets.v6i9.3443
- Tsushima, R. (2015). Methodological diversity in language assessment research: The role of mixed methods in classroom-based language assessment studies. *International Journal of Qualitative Methods*, 14(2), 104-121. DOI: <http://dx.doi.org/10.7748/nr.21.6.28.e1251>
- Turner, C. E., & Purpura, J. E. (2016). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Baneerjee (Eds.), *Handbook of second language assessment* (pp. 255–272). *De Gruyter*, 10, 9781614513827-018.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749–770.
- Walters, F. S. (2010). Cultivating assessment literacy: Standards evaluation through language-test specification reverse engineering. *Language Assessment Quarterly*, 7(4), 317-342.
- William, D. (2011). What is assessment for learning? *Studies in educational evaluation*, 37(1), 3-14.
- Wilson, A. (2014). Being a practitioner: An application of Heidegger's phenomenology. *Nurse Researcher*, 21(6), DOI: 28-33. <http://dx.doi.org/10.7748/nr.21.6.28.e1251>
- Yi-Ching, P. (2009). A review of washback and its pedagogical implications. *VNU Journal of Foreign Studies*, 25(4).
- Zainal, Z. (2007). Case study as a research method. *Jurnal kemanusiaan*, 5(1).
- Zandi, H., Kaivanpanah, S., & Alavi, S. M. (2014). The effect of test specifications review on improving the quality of a test. *Iranian Journal of Language Teaching Research*, 2(1), 1-14.
- Zeng, W., Huang, F., Yu, L., & Chen, S. (2018). Towards a learning-oriented assessment to improve students' learning—a critical review of literature. *Educational Assessment, Evaluation and Accountability*, 30, 211-250.

# **APPENDICES**

**Appendix A: Selected Assessment Artifacts**

**1st year MI Final Exam 2010-2011**

Aboubekr Belkaïd University

Year: 2010/2011

English Exam:

Duration: 01 Hour 15 Min.

The Text:

First year Final

Teachers in some secondary schools in Britain are worried that their jobs may become impossible shortly unless something can be done to restore discipline in the classrooms. In the problem schools, mostly in large cities, a small minority of teenage pupils deliberately disrupt lessons to such an extent that teachers can no longer teach their classes effectively.

Some within the teachers unions consider that the permissive nature of modern society is responsible. Small children who are continually encouraged to express their individually without restriction are naturally reluctant to accept school discipline when they grow older.

Perhaps the problem can be solved by improving facilities for the psychological guidance of these difficult children or by better cooperation between the schools and the parents – for the parents may be mainly responsible for the aggressive behaviour of their off-spring. But some of us believe that there ought to be a return to more "old-fashioned" methods. At present in some schools teachers may not even slap a child who misbehaves, but I personally feel that caning should be re-introduced and that this might produce the desired results.

Ex-teacher, Beckenham.

I). **COMPREHENSION**

Answer the following questions according to the text:

- Which problem is the writer raising?
- Who may be responsible for the misbehaviour of children?
- Do all teachers in Britain beat their pupils?
- Does the writer agree on the corporal punishment?

II). **Mastery of language**

A. **Lexis**

a) Find in the text words, phrases, or expressions that are closest in meaning to:

Beating – re-establish – adolescents.

b) Find in the text words, phrases, or expressions that are opposite in meaning to:

possible – refuse.

B. **Syntax:**

Rewrite the second sentences so that it means the same as the first one:

Teachers may not slap a child.

A child .....

Caning should be re-introduced by teachers.

Teachers .....

"Why did you leave the house yesterday?"

He asked her .....

III). **WRITTEN EXPRESSION**

This is a conversation. Complete what B says:

A: Do you know what happened at the school last week?

B: .....

A: What did the administration do?

B: .....

A: Two days suspension for the pupil!

B: .....

A: What about the teacher?

B: .....

**1st year MI Makeup Exam 2010-2011**

About Bekr Belkaïd University  
Mathematics department  
First year: 2010/2011

Duration: **01** hour;  
Full name: .....  
Section:.....

**English Exam**

**Activity 1:** Fill each gap with a word from the box:

"marginalized-their-educational-right-compulsory-for-commitments";

The right to education is a fundamental human.....; It is essential and indispensable .....the exercise of all other human rights and for development. Education is the primary vehicle by which economically and socially .....adults and children can lift themselves out of poverty and obtain the means to participate fully in .....communities.

But in spite of all the.....made by governments for providing education for all, especially free and .....basic education, millions of children still remain deprived of .....opportunities.

**Activity 2:** Combine these pairs of sentences using the appropriate linking word. Make the necessary changes when necessary.

- He practices a lot of sport. He <sup>keeps</sup> ~~keep~~ so fit.  
.....
- The world wildlife fund raised money. It wanted to save animals and plants from extinction.  
.....
- You will feel better, you have to take your medicine regularly.  
.....
- She slammed the door behind her. She was furious.  
.....

**Activity 3:** Underline the silent letters in the following words.

Sandwich-dumb-wrist-talk-knock-hymn-pseudonym-christmas-answer-guide;

**Activity 4:** Sum up your project in 6 lines.

.....  
.....  
.....  
.....  
.....  
.....

**1st year MI Final Exam 2011-2012**

**Abou Bekr Belkaïd University.**

**Year:** 2011/2012.

**Mathematics department.**

**Duration:** 1 Hour.

**Read the passage carefully then do the activities:**

All around the world, people are travelling more. They are also spending more money on travel. What is the most popular country that people go to? France is the most popular, the United States is the second, Spain was third. Italy and Britain came next, and China was sixth.

Where do the big spenders go? In 1996, Americans went mostly to Mexico (20%) and Canada (13%), but they also went to Europe and the Caribbean. Germans went mostly to other countries in Europe, about (0%) of them came to the United States, Japanese travelers went all over the world.

Why do people travel so much? Some of them want to learn about other cultures. Some people travel for business. In United States, many people are immigrants. They travel to their home countries to visit their families.

**Section One:** Reading Comprehension.

- 1) How many interrogative sentences are there in the text?
- 2) Find in the text words or expressions whose definitions follow:
  - a) Liked or enjoyed by a large number of people.
  - b) To go to see a person or a place for a period of time.
  - c) People who come to another country to live.
- 3) What do the underlined words refer to in the text:
  - a) Went there in 1996 : (§ 1)
  - b) About (2%) of them came : (§ 3)
- 4) Spot the mistake and correct it:
  - a) I love travelling on train.
  - b) This is an area rarely visit by Europeans.
  - c) Where you are going for your holidays this year?
- 5) Reorder the following words to make a meaningful sentence.  
be/ across/ Sahara/ exciting / travelling / would / the/ adventure / an.

**Section Two:** complete the following sentences.

- Put your own address at the .....on the right.
- Addresses generally follow the rule of .....house number, then street, then town.
- Don't put your .....with the address.
- In formal letters and business letters, put the name and address of the ....., you are writing to on the .....side of the page.
- Letters which begin with Dear Sir.. Usually finish .....
- One does not normally .....a letter to a stranger with an enquiry about health.

**Good Luck**

**1st year MI Makeup Exam 2011-2012**

Abou Bekr Belkaïd University  
Mathematics Department  
First year;

Full name:  
Section:  
Duration: (01) hour;

**English Exam**

**Activity One:** Fill in the blanks.

- 1) A..... is a ..... with six sides;
- 2) A four-sides figure is called a .....
- 3) The distance around a circle is called its.....;
- 4) The line drawn from one side of the circle to the other, passing through the point of origin is called the .....
- 5) A straight line joining the ends of an arc is called a .....

**Activity Two:** Fill in the blank spaces in the following sentences.

- 1) Any number to the ..... of 0 (nought) is equal to.....;
- 2) If the plan of building is drawn to a ..... of 1: 65, one centimetre on the plan .....sixty -five centimetre on the building;
- 3) To .....a vulgar fraction to a decimal fraction, we simply ..... the numerator by the denominator;
- 4) The result of a division problem is called the .....
- 5) Any number consists of combinations of .....

**Activity Three:**

- 1) Divide thirty-six pounds into three parts in the ratio 6:5:1;  
.....
- 2) If teen litres of oil weigh eight kilograms, and a litre of water weighs one kilogram, what is the ratio of the relative density of oil and water?  
.....
- 3) Which fraction with a denominator of sixteen is in proportion to one over four?  
.....

**Activity Four:** Summarize your project in (6) lines.

.....  
.....  
.....  
.....  
.....  
.....

**1st year MI Final Exam 2012-2013**

Full Name:.....;  
 Section N° :.....

**ENGLISH EXAM**

First Year; (2012/2013)  
 Duration: 01<sup>h</sup>;

**Exercise N° 01:** (10 pts)

Fill the blank spaces in the following sentences:

- 1) When we have .....an equation, we should .....our answer.
- 2) A.....is the product of two binomials.
- 3) The smallest number which is exactly .....by two or more numbers is called their .....
- 4) Any number to the .....of 0 (nought) is equal to .....
- 5) An integer plus a fraction makes a .....
- 6) To .....a vulgar fraction to a decimal fraction, we simply .....the numerator by the denominator.
- 7) An improper fraction exists when the .....is greater than the .....
- 8) Any number consists of combinations of .....
- 9) Eighteen subtracted .....twenty equals .....
- 10) Addition is a mathematical .....in which one quantity is .....to another.

**Exercise N° 02:**

- 1) Write five definitions of the angles (02, 5 pts).

- .....  
 - .....  
 - .....  
 - .....  
 - .....

- 2) In geometry, is there any difference between a point and a dot why? (01 pt)

.....

- 3) Write the verbal equivalents of the following mathematical symbols. (02, 5 pts)

1.  $X^n$
2.  $\propto$
3.  $\approx$
4.  $\equiv$
5.  $\pm$

- 4) Give definitions of the following terms. (04 pts)

- a) Chord .....
- b) Radius.....
- c) Circumference.....
- d) Diagonal.....



1st year MI Makeup Exam 2012-2013

Abou Bekr Belkaïd University  
 FULL NAME:.....

First Year; 2012/2013  
 Time allowed: 01 hour

English Exam

Reading: Read the text carefully then do the activities below:

A) Comprehension

Text

During the past decade, there has been a considerable increase in the number of natural disasters which have caused hardship and suffering in many countries. Disasters are not confined to particular regions, nor do they discriminate between rich and poor countries. For the past years alone, floods in France, Italy, Algeria and Vietnam as well as earthquakes in El Salvador and India have reminded us of the need to take global preventive measures against disasters.

Although we continue to refer to these disasters as "natural", they are man-made. "The increasing emission of carbon dioxide and other gases, deforestation and desertification play an important role in the increase of natural disasters", says Dr Briceno, the director of the ISDR.

The most significant lesson we have learnt from disasters is that solidarity and close coordination with other people are necessary to promote a culture of prevention.

- 1) Are the statements true, false or not mentioned?
  - a) Natural disasters affect especially poor countries .....
  - b) Carbon dioxide and gases are emitted from factories .....
  - c) Solidarity with other people helps to prevent disasters .....
- 2) Answer the questions according to the text :
  - a) What message do natural disasters send to all people?  
 .....
  - b) Are natural disasters really natural?  
 .....
  - c) What have we learnt from disasters?  
 .....
- 3) In which paragraph is it mentioned that cutting trees is one of the causes of natural disasters?  
 In.....§
- 4) What or who do the underlined words refer to in the text?
  - a) - Which (§1): →.....; b) We (§3): →.....;

B) Text Exploration

1. Find in the text words or expressions that are closest in meaning to:
  - a) augmentation (§1):..... b) precaution (§3):.....
2. Rewrite sentence (b) so that it means the same as sentence (a):
  - a) "What have you learnt from disasters?" the teacher asked them.  
 The teacher asked them .....
  - b) He said that he had seen him near the station.  
 He said.....
3. Underline the silent letter: autumn - aisle - foreigner.
4. Fill in the gaps with one of the words from the list below:

arms - terrible - behind - insignificant - effects - produce - civilisations - both.

The damages that natural disasters.....either on people or nature are .....but if we compare them with the bad.....that some aspects of our.....bring they look:.....and limited .....in duration and in size.

Good Luck

**1st year MI Final Exam 2013-2014**

*Abou Bekr Belkaid University Tlemcen –English Exam-*  
**Mathematics Department**  
 First Year

Family name :  
 First name :  
 Student's number :

Exercise N° 01: Answer by true or false:

- Put your own address at the top on the left.
- Put the date directly under the address.
- In U.S.A it is common to write the month first and to put a comma before the year.
- Don't put your name "the sender" with the address.
- Letters which begin Dear Sir .....usually finish yours faithfully.
- Addresses generally follow the rule of "the biggest first" town, then street, then house number.
- One does normally begin a letter to a stranger with an enquiry about health.
- Sign with your first name (formal) or your full name (informal).

Exercise N° 02: Circle or underline the silent letters in the following words:

- Hustle; - Plumber; - Aisle; - Psalm; - Signature; - Friend; -Knee; - Right; - Cupboard;
- Exhibition; - Muscle; - Autumn; - Corps ;

Exercise N°03: Select the correctly punctuated sentence:

- 1) Can you tell me what your address is
  - a) Address is!
  - b) Address is?
- 2) John's wife Margot has resigned from her position as Head Nurse.
  - a) John's wife, Margot, has resigned from her position as Head Nurse.
  - b) John's wife, Margot has resigned from her position, as Head Nurse.
- 3) Ouch yelled Cecilia you're stepping on my toes!
  - a) Ouch, yelled Cecilia: you're stepping on my toes!.
  - b) "Ouch!" yelled Cecilia. "You're stepping on my toes!"
- 4) After our team won the national championship in 1999 the coach found recruiting much easier.
  - a) No punctuation needed
  - b) In 1995, the coach
- 5) He has a two year old son.
  - a) Two-year- old
  - b) Two year- old.

*Good Luck*

## 1st year MI Final Exam 2014-2015

Full name:

Mathematics department  
English Exam  
First year

Duration: 1 hour

Exercise one: Using single words, fill in the blank spaces in the following sentences.

- \* 67 is a number. It is a whole number or . . . . .
- \* 3, 5, 7 are . . . . . numbers. 2, 12, 38 are . . . . . numbers.
- \* Eighteen subtracted . . . . . twenty equals . . . . .
- \* The result of a division problem is called the . . . . .

Exercise two:

- \* Divide thirty-six pounds into three parts in the ratio 6:5:1
- \* Five families have a total of 100 sheep. How many sheep will six families have if the numbers are in proportion?
- \* Divide one hundred and forty sheep into two groups in the ratio of 3:4
- \* Which fraction with a denominator of sixteen is in proportion to one over four?

Exercise three: Fill in the blank spaces in the following sentences

- \*  $\sqrt{64}$  means . . . . . of 64.
- \*  $\sqrt{x}$  means . . . . . of  $x$ .
- \* To divide powers we . . . . . the . . . . .
- \* Any number to the . . . . . of 0 (nought) is equal to . . . . .
- \*  $\frac{4}{5}$  is a fraction, 4 is the . . . . . and 5 is the . . . . .

good luck

**1st year MI Final Exam 2017-2018**

<b>L1 M1</b>	<b>English Exam</b>	<b>1<sup>st</sup> Semester - 17/01/2018</b>
<b>Full Name:</b>		
<b>Section:</b>	<b>Signature:</b>	

**Activity 1 (06 pts):** Fill in the paragraph below with the appropriate words from the following list. Not all the words can be used.

- |                    |                 |            |
|--------------------|-----------------|------------|
| 1. virtual machine | 4. CPU          | 7. devices |
| 2. CLI             | 5. terminal     | 8. GUI     |
| 3. programs        | 6. dual-booting | 9. Kernel  |

“A computer would be fairly useless without an OS. Operating systems allow users to execute all sorts of ..... on their computers. .... is the fundamental part of an operating system responsible for providing access to the machine's hardware. Running more than one OS on a single computer is possible with ..... or by using a ..... The ..... allows the operators to interact with the system by using the mouse to manipulate the visual elements on the screen, while the ..... refers to the way for interacting with the system by typing text commands into the ..... window.”

**Activity 2 (03 pts):** Identify and underline the subjects (S), verbs (V), and objects (O), as well as the adjectives (Adj) and adverbs (Adv) in each of the following sentences:

1. Computer-users prefer an icon-based interface.
2. The server was set to automatically log off users.
3. The programmer tested his Java code on different operating systems.

**Activity 3 (03 pts):** Using the appropriate conjunction, combine the following pairs of sentences. Make the necessary changes:

- Pair 1:           *“Since” – “Therefore” – “Whereas”*
  - CDs have almost become obsolete.
  - Modern storage devices have replaced CDs.

.....

.....

- Pair 2:           *"For" – "But" – "So"*
  - A computer would be fairly useless without an OS.
  - Today almost all computers come with an OS pre-installed.

.....

.....

.....

- Pair 3:           *"If" – "Yet" – "Because"*
  - The computer user could not log on.
  - The computer user forgot his username and password.

.....

.....

**Activity 4 (08 pts):** Put the verbs between brackets into the correct tense:

1. My family ..... (buy) some land recently. They ..... (build) a summer house there at the moment.
2. He ..... (watch) the news on TV every day and it ..... (help) him with his English.
3. Yesterday, my internet browser ..... (crash) again, when I ..... (upload) the files.
4. When he ..... (found) Microsoft, Bill Gates was only years old. He ..... (already / write) his first computer program six years earlier.
5. She said that one day she ..... (retire) from teaching and ..... (spend) her new free time learning about computers.
6. I ..... (not sleep) at all last night, because my roommate ..... (listen) to music all night.
7. She ..... (not see) her father since he ..... (start) to work in Algiers.
8. By the end of the month, we ..... (complete) the first semester.

**Good Luck!**



Sentence 2 (Condition):

.....  
.....

Sentence 3 (Contrast):

.....  
.....

**Task 4 (4.5 pts):**    *Put the verbs between brackets into the correct tense.*

1. Computers have evolved throughout the late 20th century; they (to become) more interactive.

.....  
.....

2. The hardware (to be) what makes a computer powerful.

.....  
.....

3. Researchers from different disciplines (to participate) in the development of AI.

.....  
.....

4. My internet browser (to crash) again while I (to surf) on the net.

.....  
.....

5. Ten years from now, even teenagers (to be) able to develop a computer program.

.....  
.....

**Good Luck!**

**1st year MI Final Exam 2019-2020**

<b>L1M1</b>	<b>English Exam</b>	<b>27/01/2020 - 3:00 to 4:00 p.m</b>
<b>Full Name:</b>		
<b>Section:</b>	<b>Signature:</b>	

**Task 1 (06 pts):** *Fill in the following paragraph with the appropriate words from the list below.*

- |             |             |          |
|-------------|-------------|----------|
| 1. software | 3. operator | 5. pixel |
| 2. hardware | 4. data     | 6. input |

- a. "Quite simply, computer ..... is the physical components that a computer system requires to function."
- b. "An ..... device is any hardware device that sends data to a computer."
- c. "Flash memory is a form of nonvolatile memory that retains ..... in the absence of a power supply."
- d. "An OS is a GUI or CLI software link between the computer and the ....."
- e. "Scripts, applications, programs and a set of instructions are all terms used to describe ....."
- f. "A ..... is the smallest unit of a digital image or graphic."

**Task 2 (3 pts):** What do the following acronyms stand for?

Example: BIOS → Basic Input-Output Services

- |                |                |
|----------------|----------------|
| a. CPU → ..... | c. CLI → ..... |
| b. GUI → ..... | d. OS → .....  |

**Task 3 (06 pts):** *Complete each of the following sentences with the correct form of the adjectives between brackets.*

1. The ..... you study for the exams, the ..... you will do. (**hard / great**)
2. Python is one of ..... programming languages. (**easy**)
3. This computer is considerably ..... your old one. (**powerful**)
4. According to Elon Musk, AI is as ..... as nuclear exploitation. (**threatening**)



5. The ..... is a CPU clock speed, the ..... a computer will be. (**high / fast**)

6. Besides being a language, Mathematics is ..... English. (**universal**)

**Task 4 (05 pts):** *Put the verbs between brackets into the correct tense.*

1. Researchers from different disciplines (to participate) ..... in the development of computer sciences.

2. The program (to crash) ..... but the data (to be) ..... not lost.

3. Mobile phones have evolved; they (to become) ..... more interactive.

4. Ten years from now, even teenagers (to be) ..... able to develop a computer program.

**Good Luck!**

1st year MI & I Final Exam 2022-2023

<u>L1 MI/I</u>	<u>English Final Exam</u>	<u>2022/2023</u>
<u>Full Name:</u>	<u>Signature:</u>	

**Task N°1 (9 pts):** Complete the following passage with the appropriate words from the list below.

- |                 |                |             |
|-----------------|----------------|-------------|
| 1. instructions | 4. peripherals | 7. storage  |
| 2. binary       | 5. operator    | 8. input    |
| 3. keyboard     | 6. malware     | 9. circuits |

“Computer ..... are any of various devices (including sensors) used to enter information and ..... into a computer for storage or processing and to deliver the processed data to the .....

These devices are commonly divided into three kinds: input devices, output devices, and ..... devices. An ..... device converts incoming data and instructions into a pattern of electrical signals in ..... code that are comprehensible to a digital computer. An output device reverses the process, translating the digitized signals into a form intelligible to the user.”

**Task N°2 (06 pts):** Complete each of the following sentences with the correct form of the adjectives between brackets.

1. The ..... you practice, the ..... you learn. (**hard / much**)
2. Python is one of ..... programming languages. (**easy**)
3. This computer is considerably..... your old one. (**expensive**)

**Task N°3 (05 pts):** *Label the pictures below*



1-



2-



3-



4-



5-

**Good Luck!**

**Appendix B: Observation Guide**

Observation N°:

Teacher:

Settings:

<b>Aspect</b>	<b>Observation</b>
Assessment Type	
Input Language	
Targeted Language KSA	
Targeted Topical KSA	
Individual Needs	
Timely Feedback	
Integration into Lesson Plan	
Emphasis on Progress	
Student Goal Participation	
Teacher Inquiry	

## ***APPENDICES***

---

Alignment with Goals	
Teacher Expertise	
Seamless Integration	
Focus on Student Welfare	
Next Steps Guidance	
Enhanced Understanding	
Student Responsibility	
Multiple Data Sources	
Side Notes	

## **Appendix C: Learners' Appreciation Survey**

Dear student,

As part of our ongoing efforts to enhance the quality of English language education within the field of computer science, we invite you to participate in this research. This survey is designed to gather insights into your experiences with final English exams, as well as your language skills and needs. Your participation in this research will provide valuable data that can lead to positive changes in your academic experience.

Participation in this survey is entirely voluntary, and you have the freedom to choose whether or not to participate. Your decision to participate or not will not impact your academic standing or any other aspect of your education. Your responses will be kept confidential and anonymized, ensuring your privacy is protected.

We understand that your time is precious, and we truly appreciate your willingness to participate in this research. Your perspectives are vital in driving positive changes, and your contributions will help create a more effective and tailored English language education system within computer science programs.

If you choose to participate, please answer the survey questions honestly and to the best of your abilities. Your insights are invaluable, and we are committed to using this information responsibly and ethically to enhance your learning experience.

Please click on the space corresponding to your chosen response.

Thank you.

**Background Information**

Are you a 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> year Computer Science Student?

- 1<sup>st</sup> year
- 2<sup>nd</sup> Year
- 3<sup>rd</sup> Years

**English Exam Procedures**

Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The instructions for completing the English exam are clear.					
The English examination process is organized.					
The English exam invigilators are professional.					
I am allotted enough time to complete the English exam.					
I have problems with the English exam materials.					

**Reading Skill Assessment and Development**

Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The exam assesses my ability to read and understand different types of texts, such as news articles, academic essays, and fiction					
The exam assesses my ability to identify the main ideas and supporting details in a text.					
The exam assesses my ability to draw inferences and conclusions from a text.					
The exam assesses my ability to identify and understand the author's purpose and point of view in a text.					



**Writing Skill Assessment and Development**

Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The exam assesses my ability to write clear and concise sentences and paragraphs.					
The exam assesses my ability to organize my thoughts and ideas in a logical way.					
The exam assesses my ability to use appropriate grammar and vocabulary.					
The exam assesses my ability to write in different genres, such as essays, reports, and letters.					

**Speaking Skill Assessment and Development**

Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The exam assesses my ability to speak clearly and fluently on a variety of topics					
The exam assesses my ability to use appropriate grammar and vocabulary in speech					
The exam assesses my ability to organize my thoughts and ideas in a logical way when speaking					
The exam assesses my ability to interact with others in a conversation					

**Listening Skill Assessment and Development**

Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The exam assesses my ability to understand spoken language in a variety of contexts, such as lectures, conversations, and presentations					
The exam assesses my ability to identify the main ideas and supporting details in a spoken passage.					
The exam assesses my ability to draw inferences and conclusions from a spoken passage.					
The exam assesses my ability to follow instructions and complete tasks based on spoken information.					

**Use of Language Skills in Specific Domains**

Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
<p>The English exam prepared me for using language in Computer Science academic domain such as reading papers, following an online course, or participating to a conference.</p>					
<p>The English exam prepared me for using language in Computer Science professional domain such as writing a professional document, presenting a solution, and eliciting needs.</p>					
<p>The English exam prepared me for using language in Computer Science socio-professional domain such as interacting with colleagues, interacting with customers, and freelancing.</p>					
<p>The English exam helped me to achieve my domain-related language learning goals.</p>					

**Appendix D: Scoring Rubric for Discount Scheme Dialogue Completion Task**

Criteria	Proficient (5)	Advanced (4)	Intermediate (3)	Needs Improvement (2)	Beginner (1)
Topical Control	<ul style="list-style-type: none"> <li>- Demonstrates deep understanding of the technical aspects related to implementing a discount scheme.</li> <li>- Asks a comprehensive set of technical questions covering all relevant areas.</li> </ul>	<ul style="list-style-type: none"> <li>- Shows a good grasp of the technical aspects and asks several relevant questions.</li> <li>- Most technical areas are covered adequately.</li> </ul>	<ul style="list-style-type: none"> <li>- Displays an acceptable understanding of the technical aspects and asks a reasonable number of relevant questions.</li> <li>- Covers some important technical areas.</li> </ul>	<ul style="list-style-type: none"> <li>- Displays limited understanding of the technical aspects and asks only a few relevant questions.</li> <li>- Some important technical areas may be overlooked.</li> </ul>	<ul style="list-style-type: none"> <li>- Lacks understanding of the technical aspects and fails to ask relevant technical questions.</li> <li>- Important technical areas are entirely omitted.</li> </ul>
Language Accuracy	<ul style="list-style-type: none"> <li>- Uses clear and precise language throughout the dialogue.</li> <li>- Excellent grammar, vocabulary, and sentence structure.</li> <li>- Communicates ideas effectively</li> </ul>	<ul style="list-style-type: none"> <li>- Generally uses clear language with a few minor issues.</li> <li>- Good grammar, vocabulary, and sentence structure.</li> <li>- Ideas are mostly communicated effectively.</li> </ul>	<ul style="list-style-type: none"> <li>- Uses somewhat clear language, but may have noticeable issues that occasionally hinder communication.</li> <li>- Adequate grammar, vocabulary, and sentence structure.</li> <li>- Ideas are conveyed reasonably.</li> </ul>	<ul style="list-style-type: none"> <li>- Language is unclear at times, hindering communication.</li> <li>- Some noticeable grammar, vocabulary, or sentence structure issues.</li> <li>- Ideas may not be consistently clear.</li> </ul>	<ul style="list-style-type: none"> <li>- Language is incomprehensible or severely flawed, impeding communication.</li> <li>- Major grammar, vocabulary, and sentence structure issues.</li> <li>- Ideas are poorly conveyed.</li> </ul>

<p>Rhetorical Control</p>	<ul style="list-style-type: none"> <li>- Demonstrates excellent rhetoric skills, actively listening to the client and responding appropriately.</li> <li>- Maintains a polite and professional tone throughout.</li> <li>- Dialogue flows smoothly.</li> </ul>	<ul style="list-style-type: none"> <li>- Displays proficient rhetoric skills, actively engaging with the client and responding adequately.</li> <li>- Maintains a mostly polite and professional tone.</li> <li>- Dialogue is generally smooth.</li> </ul>	<ul style="list-style-type: none"> <li>- Shows acceptable rhetoric skills, with adequate engagement with the client and appropriate responses.</li> <li>- Tone is generally polite and professional.</li> <li>- Dialogue flows reasonably.</li> </ul>	<ul style="list-style-type: none"> <li>- Shows limited rhetoric skills, with occasional lapses in active engagement with the client.</li> <li>- Tone may be inconsistent.</li> <li>- Dialogue may have some disruptions.</li> </ul>	<ul style="list-style-type: none"> <li>- Lacks effective rhetoric skills, failing to actively engage with the client or respond appropriately.</li> <li>- Tone is unprofessional or inappropriate.</li> <li>- Dialogue is disjointed and incoherent.</li> </ul>
---------------------------	--	--	---	---	---

## Appendix E: Generic Test

### What Questions Do I Ask During Requirements Elicitation?

This article emphasizes the significance of utilizing a requirements questionnaire to enhance the effectiveness of requirements elicitation sessions and save time. A requirements questionnaire is a structured list of questions organized according to project features or objectives. It aids in refining the understanding of high-level requirements and engaging stakeholders more effectively. By investing time in preparing a well-thought-out questionnaire, follow-up meetings can be reduced. The questionnaire should aim to discover requirements rather than merely gather them.

When creating a requirements questionnaire, it is advisable to work through each feature individually. For instance, questions like "How will your stakeholders use this feature?" or "What are the steps involved in this feature?" help in comprehending its purpose and implications. Questions related to "Where" can focus on the location of feature access or visibility of results. "When" questions delve into the timing of feature usage or completion requirements. Questions about "Who" identify the users, input providers, output recipients, or key stakeholders to consult. "What" questions explore the feature's functionalities, desired outcomes, tracking needs, or alternative scenarios. Lastly, "Why" questions confirm the alignment between requirements and project needs.

It's important to note that the questions in the requirements questionnaire should not be asked one-by-one in a linear manner. Instead, select a few core questions to initiate discussion. As the stakeholders express their vision, utilize the remaining questions as a guide to ensure a comprehensive conversation. Typically, only about half of the questions will be directly asked, while the other half is addressed indirectly through the conversation. By employing a requirements questionnaire and following a structured approach to requirements elicitation, clients can actively participate, and the project requirements can be thoroughly understood and documented, leading to more successful outcomes.

Adapted from Brandenburg, L., & Brandenburg, L. (2023). What Questions Do I Ask During Requirements Elicitation? *Bridging the Gap | We'll Help You Start Your Business Analyst Career*.

## I. Reading Comprehension (05/30 pts)

Answer the following questions according to the text.

1. Which type of questions in a requirements questionnaire focus on the location or visibility of a feature?
2. What kind of questions, in a requirements questionnaire, explore the functionalities and desired outcomes of a feature?
3. What is the main objective of asking "Why" questions in a requirements questionnaire?
4. How should the questions in a requirements questionnaire be asked?
5. Which of the following best describes the purpose of the text?

## II. Language Mastery (10/30 pts)

Circle the correct answer.

1. \_\_\_\_\_ you hope to achieve by implementing a program in your bookstore?
  - a. What do
  - b. How do
  - c. Why do
2. \_\_\_\_\_ the program enhance your customers' shopping experience?
  - a. How should
  - b. How may
  - c. How must
3. Are you primarily interested \_\_\_\_\_ customers to purchase more books or increasing customer loyalty?
  - a. in incentivize
  - b. in incentivized
  - c. in incentivizing



4. Do you have any specific discount schemes or promotional offers in mind that \_\_\_\_\_ the program to support?
  - a. you should like
  - b. you would like
  - c. you could like
  
5. \_\_\_\_\_ any particular features or functionalities you would like the program to have, such as tracking customer purchases or managing inventory?
  - a. Is there
  - b. Are there
  - c. Were there
  
6. \_\_\_\_\_ you currently handle customer loyalty or rewards programs?
  - a. How are
  - b. How have
  - c. How do
  
7. \_\_\_\_\_ of reporting or analytics would be beneficial for you to have access to through the program?
  - a. How kind
  - b. What kind
  - c. Why kind
  
8. \_\_\_\_\_ have a preferred timeframe or budget for implementing this program?
  - a. Do you
  - b. Are you
  - c. Will you
  
9. Should the program \_\_\_\_\_ through a mobile app, website, or both?
  - a. be accessing
  - b. be accessible
  - c. be accessibility

10. If you have any further questions or ideas during the development process, please \_\_\_\_\_ to reach out to me.

- a. you have the freedom
- b. you are free
- c. feel free

### III. Written Expression (15/30 pts)

- Complete the dialogue between the developer and the client, focusing on asking relevant questions and actively listening to the client's responses.

- Developer:

Client: Good morning! It's my pleasure. I'm Adam, the owner of the retail store.

- Developer:

Client: Certainly! We have a store that sells various items, and we want to introduce a discount scheme to incentivize customers to purchase multiple items.

- Developer:

Client: We want to offer a 10% discount to customers if they purchase 3 or more items. However, if they buy fewer than 3 items, no discount should be applied.

- Developer:

Client: Yes, exactly!

- Developer:

Client: We currently use a Python-based point-of-sale system, so it would be ideal if the discount calculation algorithm can be implemented in Python.

- Developer:

Client: The prices are entered manually at the checkout.

- Developer:

Client: It would be great if the system can display the total cost before and after the discount, so customers can see the savings they're making.

- Developer:

Client: That sounds great! Thank you for addressing all my questions and concerns.

- Developer:

Client: Thank you so much for your help! I'm looking forward to seeing the implementation.

- Developer:

Client: I will definitely do that. Thank you once again!

- Developer:

Client: You too! Goodbye for now.

## Appendix F: ICL Test

### Requirements Elicitation

#### Task 1: Maximizing Client's Engagement through Requirements Questionnaires

Read the passage and circle the correct answer (05/35 pts).

#### What Questions Do I Ask During Requirements Elicitation?

This article emphasizes the significance of utilizing a requirements questionnaire to enhance the effectiveness of requirements elicitation sessions and save time. A requirements questionnaire is a structured list of questions organized according to project features or objectives. It aids in refining the understanding of high-level requirements and engaging stakeholders more effectively. By investing time in preparing a well-thought-out questionnaire, follow-up meetings can be reduced. The questionnaire should aim to discover requirements rather than merely gather them.

When creating a requirements questionnaire, it is advisable to work through each feature individually. For instance, questions like "How will your stakeholders use this feature?" or "What are the steps involved in this feature?" help in comprehending its purpose and implications. Questions related to "Where" can focus on the location of feature access or visibility of results. "When" questions delve into the timing of feature usage or completion requirements. Questions about "Who" identify the users, input providers, output recipients, or key stakeholders to consult. "What" questions explore the feature's functionalities, desired outcomes, tracking needs, or alternative scenarios. Lastly, "Why" questions confirm the alignment between requirements and project needs.

It's important to note that the questions in the requirements questionnaire should not be asked one-by-one in a linear manner. Instead, select a few core questions to initiate discussion. As the stakeholders express their vision, utilize the remaining questions as a guide to ensure a comprehensive conversation. Typically, only about half of the questions will be directly asked, while the other half is addressed indirectly through the conversation. By employing a requirements questionnaire and following a structured approach to requirements elicitation, clients can actively participate, and the project requirements can be thoroughly understood and documented, leading to more successful outcomes.

Adapted from Brandenburg, L., & Brandenburg, L. (2023). What Questions Do I Ask During Requirements Elicitation? *Bridging the Gap / We'll Help You Start Your Business Analyst Career*.

11. Which type of questions in a requirements questionnaire focus on the location or visibility of a feature?
  - a. How questions
  - b. Where questions
  - c. Why questions
  
12. What kind of questions, in a requirements questionnaire, explore the functionalities and desired outcomes of a feature?
  - a. Who questions
  - b. When questions
  - c. Why questions
  
13. What is the main objective of asking "Why" questions in a requirements questionnaire?
  - a. Confirming the alignment between requirements and project needs
  - b. Identifying the steps involved in a feature
  - c. Exploring alternative scenarios of a feature
  
14. How should the questions in a requirements questionnaire be asked?
  - a. All questions should be asked directly
  - b. Questions should be asked in a linear manner
  - c. Only a few core questions should be asked
  
15. Which of the following best describes the purpose of the text?
  - a. To provide a step-by-step guide for conducting requirements elicitation sessions
  - b. To highlight the importance of utilizing a requirements questionnaire
  - c. To explain the different types of stakeholders involved in a project

### **Task 2: The language of requirements questionnaire**

Circle the correct answer (10/35 pts).

16. \_\_\_\_\_ you hope to achieve by implementing a program in your bookstore?
  - a. What do
  - b. How do
  - c. Why do
  
17. \_\_\_\_\_ the program enhance your customers' shopping experience?

- a. How should
  - b. How may
  - c. How must
18. Are you primarily interested \_\_\_\_\_ customers to purchase more books or increasing customer loyalty?
- a. in incentivize
  - b. in incentivized
  - c. in incentivizing
19. Do you have any specific discount schemes or promotional offers in mind that \_\_\_\_\_ the program to support?
- a. you should like
  - b. you would like
  - c. you could like
20. \_\_\_\_\_ any particular features or functionalities you would like the program to have, such as tracking customer purchases or managing inventory?
- a. Is there
  - b. Are there
  - c. Were there
21. \_\_\_\_\_ you currently handle customer loyalty or rewards programs?
- a. How are
  - b. How have
  - c. How do
22. \_\_\_\_\_ of reporting or analytics would be beneficial for you to have access to through the program?
- a. How kind
  - b. What kind
  - c. Why kind

23. \_\_\_\_\_ have a preferred timeframe or budget for implementing this program?
- a. Do you
  - b. Are you
  - c. Will you
24. Should the program \_\_\_\_\_ through a mobile app, website, or both?
- a. be accessing
  - b. be accessible
  - c. be accessibility
25. If you have any further questions or ideas during the development process, please \_\_\_\_\_ to reach out to me.
- a. you have the freedom
  - b. you are free
  - c. feel free

**Task 3: Understanding algorithms**

Read the algorithm and circle the correct answer(05/35 pts).

```
1 def calculate_total_cost(item_prices):
2
3     """
4     Calculate the final cost by applying a discount if 3 or more items are purchased.
5
6     Args:
7     |   item_prices (list): A list of the prices of the purchased items.
8
9     Returns:
10    |   float: The final cost of the items after applying the discount (if applicable).
11
12    """
13    total_cost = sum(item_prices) # Calculate the total cost without discount
14
15    if len(item_prices) >= 3: # If 3 or more items are purchased
16        |   discount = 0.1 # 10% discount
17    else:
18        |   discount = 0 # No discount
19
20    discounted_amount = total_cost * discount # Calculate the discount amount
21    final_cost = total_cost - discounted_amount # Calculate the final cost
22
23    return final_cost
24
25
26 if __name__ == "__main__":
27     num_items = int(input("Enter the number of items purchased: "))
28     item_prices = []
29
30     for i in range(num_items):
31         price = float(input("Enter the price of item {} : ".format(i+1)))
32         item_prices.append(price)
33         amount_to_pay = calculate_total_cost(item_prices)
34
35     print("The amount to pay after discount: {:.2f}".format(amount_to_pay))
36
37
```

26. If a customer purchases 5 items with prices [10, 20, 30, 40, 50], what will be the discounted amount based on the algorithm?
- a. 35
  - b. 25
  - c. 15



27. Which of the following statements accurately describes the discount algorithm implemented for the retail store's discount scheme?
- a. The algorithm applies a 10% discount to customers who purchase 3 or more items and a 5% discount to customers who purchase fewer than 3 items.
  - b. The algorithm applies a 20% discount to customers who purchase 3 or more items and a 5% discount to customers who purchase fewer than 3 items.
  - c. The algorithm applies a 10% discount to customers who purchase 3 or more items and no discount to customers who purchase fewer than 3 items.
28. How will the discount information be displayed to customers during the checkout process?
- a. Only the final discounted price will be displayed
  - b. Only the original total cost will be displayed
  - c. No discount information will be displayed
29. What is the purpose of the function "**calculate\_total\_cost**" in the algorithm?
- a. To determine the number of items purchased
  - b. To calculate the total cost of the items before discounts
  - c. To display the final cost after the discounts are applied
30. What language did the developer use to implement the code?
- a. Python
  - b. C#
  - c. Java

## Task 4: Writing a requirements questionnaire

Assume the role of a developer who is tasked with implementing a discount scheme for a retail store. The goal is to encourage customers to purchase multiple items by offering discounts based on the number of items bought. Your task is to engage in a dialogue with the store owner (the client) to gather information about their needs and requirements regarding the discount scheme.

- Complete the dialogue between the developer and the client, focusing on asking relevant questions and actively listening to the client's responses (15/35 pts).
- 

### Guidelines:

- Introduce yourself and express gratitude for the client's time.
  - Gather information about the store and the specific discount scheme requirements.
  - Clarify and confirm the client's requirements to ensure accuracy.
  - Discuss technical aspects, such as programming language, data input method, and user interface preferences.
  - Conclude by summarizing the gathered information and expressing commitment to addressing the client's needs.
- 

### Requirements Interview:

- Developer:

Client: Good morning! It's my pleasure. I'm Adam, the owner of the retail store.

- Developer:

Client: Certainly! We have a store that sells various items, and we want to introduce a discount scheme to incentivize customers to purchase multiple items.

- Developer:

Client: We want to offer a 10% discount to customers if they purchase 3 or more items. However, if they buy fewer than 3 items, no discount should be applied.

- Developer:

Client: Yes, exactly!

- Developer:

Client: We currently use a Python-based point-of-sale system, so it would be ideal if the discount calculation algorithm can be implemented in Python.

- Developer:

Client: The prices are entered manually at the checkout.

- Developer:

Client: It would be great if the system can display the total cost before and after the discount, so customers can see the savings they're making.

- Developer:

Client: That sounds great! Thank you for addressing all my questions and concerns.

- Developer:

Client: Thank you so much for your help! I'm looking forward to seeing the implementation.

- Developer:

Client: I will definitely do that. Thank you once again!

- Developer:

Client: You too! Goodbye for now.

## ملخص:

التقييم اللغوي حاسم لفهم قدرات الطلاب وتشكيل أساليب التدريس في التعليم العالي الجزائري. ومع ذلك، يشكل النقص في مهارات التقييم بين معلمين الإنجليزية لأغراض محددة والاعتماد على الأساليب التقليدية تحديات كبيرة. يركز هذه البحث على ممارسي اللغة الإنجليزية لأغراض محددة وتأثيرهم على طلاب الإعلام الآلي في جامعة تلمسان. باستخدام منهج مرحل، تحلل الدراسة 12 موضوع تقييم بشكل نوعي وتراقب ثلاثة ممارسين للغة الإنجليزية لأغراض محددة، بالإضافة إلى إجراء استبيان كمي مع 367 طالبًا جامعيًا واختبارات أداء لـ 76 طالبًا في السنة الأولى. تكشف النتائج عن فجوات حيوية في ممارسات التقييم، بما في ذلك قضايا الانسجام والتعليق ونقص التنوع في الأساليب والمواد التقييمية، مما يؤثر سلبيًا على تطوير مهارات اللغة ذات الصلة بالمجال. وبالإضافة إلى ذلك، يسلط البحث الضوء على تعقيدات تقييم اللغة الإنجليزية لأغراض محددة، مثل نقص الممارسين المؤهلين وعدم انسجامها مع احتياجات المحترفين، وتصميمات التقييم غير الفعالة. تشدد الدراسة على ضرورة معالجة هذه التحديات لتعزيز بيئة تقييم عالية الجودة. تشمل التوصيات تزويد المعلمين بتدريب في مهارات التقييم، واعتماد تقييم متكامل للمحتوى واللغة جنبًا إلى جنب مع مبادرات مبتكرة أخرى، وإنشاء مركز جزائري للغة الإنجليزية لأغراض محددة. تدعو الدراسة إلى جهد تعاوني بين المعلمين في مجال اللغة والمواضيع، والإداريين، والباحثين لتعزيز جودة التعليم والتقييم. تسهم النتائج في فهم أوسع لممارسات تقييم اللغة الإنجليزية لأغراض محددة، مقدمة رؤى قيمة للممارسين ووصفي السياسات ومصممي المناهج الساعين لتحسين تقييم اللغة في التعليم العالي الجزائري

**الكلمات المفتاحية:** الإعلام الآلي، الإنجليزية لأغراض محددة، مؤسسات التعليم العالي، التقييم، تأثير، المحتوى المتكامل واللغة، مهارات متكاملة، تقييم موجه نحو التعلم، جامعة تلمسان

## Summary:

Language assessment is crucial for understanding learners' abilities and shaping teaching methods in Algerian Higher Education. However, a lack of assessment literacy among teachers and reliance on traditional methods pose challenges. This research focuses on English for Specific Purposes practitioners and their impact on Computer Science students at Tlemcen University. Using a two-phased approach, the study analyzes 12 assessment artifacts qualitatively and observes three English for Specific Purposes practitioners, while also conducting a quantitative survey with 367 undergraduate students and performance tests for 76 first-year students. Findings reveal gaps in ESP assessment, affecting alignment, feedback, and diversity in methods and materials. Challenges include a shortage of qualified practitioners and misalignment with professional needs. Recommendations include teacher training, adopting innovative assessments, and establishing an Algerian English for Specific Purposes Centre. Collaboration among educators, administrators, and researchers is crucial for enhancing instruction and assessment quality, providing valuable insights for improvement in Algerian Higher Education.

**Keywords:** Computer Science, ESP, HEIs, Impact, ICL Assessment, Integrated Skills, LOA, Tlemcen University

## Résumé:

L'évaluation linguistique est cruciale pour comprendre les compétences des apprenants et façonner les méthodes d'enseignement dans l'enseignement supérieur algérien. Cependant, un manque de compétence en évaluation parmi les enseignants et la dépendance à des méthodes traditionnelles posent des défis. Cette recherche se concentre sur les praticiens de l'anglais à des fins spécifiques (ESP) et leur impact sur les étudiants en informatique à l'Université de Tlemcen. En utilisant une approche en deux phases, l'étude analyse qualitativement 12 artefacts d'évaluation et observe trois praticiens de l'ESP, tout en réalisant une enquête quantitative auprès de 367 étudiants de premier cycle et des tests de performance pour 76 étudiants de première année. Les résultats révèlent des lacunes dans l'évaluation de l'ESP, affectant l'alignement, la rétroaction et la diversité des méthodes et des matériaux. Les défis comprennent une pénurie de praticiens qualifiés et un désalignement avec les besoins professionnels. Les recommandations incluent la formation des enseignants, l'adoption d'évaluations innovantes et la création d'un Centre algérien de l'anglais à des fins spécifiques. La collaboration entre les éducateurs, les administrateurs et les chercheurs est cruciale pour améliorer la qualité de l'enseignement et de l'évaluation, fournissant des perspectives précieuses pour l'amélioration de l'enseignement supérieur en Algérie.

**Mots-clés:** Informatique, AOS, EES, Impact, Evaluation CIL, Compétences Intégrés, EOA, Université de Tlemcen.