

***MEASURING THE EFFECT OF IMPLEMENTING CONTENT AND LANGUAGE-INTEGRATED LEARNING (CLIL) IN THE INSTITUTE OF LANGUAGES FOR INTERNATIONAL INTEGRATION (ILANI): AN EXPERIMENTAL RESEARCH PROJECT***

**Charlotte Simo Dgidge**

Department of General and Scientific Studies  
Bandjoun Fotso Victor University Institute of Technology, University of Dschang,  
Cameroon

E-mail: [dgidgec@yahoo.com](mailto:dgidgec@yahoo.com) or [charsim2000@gmail.com](mailto:charsim2000@gmail.com)

**Abstract**

Content and Language Integrated Learning (CLIL) was originally defined in 1994, and launched in 1996 by UNICOM, the University of Jyväskylä, and the European Platform for Dutch Education, to describe educational methods where “subjects are taught through a foreign language with dual-focused aims, namely the learning of content, and the simultaneous learning of a foreign language” (Darn, 2006, p.4). Springing forth from the fact that few experimental studies have been carried out on CLIL at different levels of the educational system and virtually none in language centers, this paper aims to provide insights into experimental research on a CLIL project for communicative skills in the context of a language center. The participants in the experiment were 50 learners of different age groups, different nationalities, and different academic and professional backgrounds, learning English as a foreign language at the Institute of Languages for International Integration located in the town of Buea in the South West Region of Cameroon, a country in central Africa. The research design involved observing, conducting interviews, and administering a CLIL test with a focus on communicative skills and the target disciplines. The data collected were statistically processed and the data analysis techniques employed were mean and standard deviation (SD) and ANOVA. The data provided support for the efficacy of implementing CLIL in a language center as the experimental group outperformed the control group in terms of both communicative skills and content knowledge and showed a considerably higher positive attitude towards the CLIL course than their peers in the control group. This paper could therefore serve as a stepping stone on which many similar yet profound experimental pieces of research on

CLIL could follow suit in language centers to ascertain its effectiveness in all other aspects and to bring all the stakeholders implicated in language valorization to applying CLIL in the teaching/learning process in language centers.

**Keywords:** *CLIL, ILANI, Language center, Communicative skills*

## **1. Introduction**

Scholars generally agree that language should be learned for communication, and teaching language through contextual communication effectively accomplishes this (Hadley, 2000). Mohan even emphasizes by declaring that "any educational approach that considers language learning alone and ignores the learning of subject matter is inadequate" (1986, p.1).

Crandall (1998) further argues that being able to use language within a specific context cannot be accomplished without integrating contextual content and language learning. Many of the arguments in favor of CLIL are deeply rooted in second language acquisition (SLA) research. Specifically, it is suggested that its implementation allows for the following four elements: 1) the creation of optimal conditions for naturalistic language learning; 2) the provision of a clearly defined purpose for using the L2; 3) the development of a positive effect for L2 learning focusing on meaning rather than form; and 4) an increase in the amount of exposure to the target language (TL) (Dalton-Puffer, 2007; Dalton-Puffer & Smith, 2007).

This is further emphasized by Krashen (1982), Lightbown & Spada (2006), and Swain (2000), who opine that L2 can be effectively acquired more in conditions that resemble the acquisition of L1 and the focus of instruction being laid on meaning rather than form, the language input being per the learners' level of competence, while opportunities for engaging in the purposeful use of the TL in a student-friendly setting are provided. Thus, there is a gap in the focus of language instruction as the learners are provided with every opportunity to engage in meaningful exposure and use of the TL through content instruction of their academic subject as part of CLIL (García, 2008; Naves, 2009).

As such, TL does not only become the medium of instruction but also of communication. Coyle (2008) makes it clearer when he views language as a "learning tool" which operates

in three ways: "language of," "language for," and "language through": Language of learning refers to the vocabulary and sentence structures that students need to learn the topic taught in the CLIL lesson; language for learning refers, at its simplest, to the use of the foreign language focusing on learning the content of the lesson, for example, answering, asking, and reporting things learned during the lesson; and Language through learning is based on an idea according to which efficient learning only happens through thinking and language. The language is used for supporting and enhancing learning simultaneously when applying things already learned. It should be noted that CLIL methodology includes the following six components: 1) its multiple foci (on language, learning, and cognition); 2) the construction of safe and enriching learning environments; 3) the use of authentic materials and interactions; 4) the promotion of active learning; 5) the use of macro-and micro- scaffolding in students' learning to enhance their autonomy; and, 6) the promotion of co-operation among students and teachers (Hammond, 2001).

Studies by several authors (Cummins, 1981; Currie, 1993; Shih, 1986) have demonstrated that communicative skills are indispensable for the success of learners in the secondary and tertiary levels of learning. Christie (1985) sums up the voice of these writers by saying that language is the "hidden curriculum" of academic and professional life and that success in school or at work is largely a language matter. Those who fail in schools are those who fail to master genres of communication. This is why this paper seeks to measure the effect of implementing CLIL on the language teaching/learning process in the Institute of Languages for International Integration (ILANI) to evaluate its impact on the acquisition of communicative competence which is of undeniable importance for academic and/or professional success.

The present study attempted an exploration of the integration of CLIL in the language curriculum of English language learners in the context of a language center through the implementation of an experimental syllabus and materials developed for the target group of learners at issue. In particular, it aimed to investigate the impact of the experimental teaching intervention by measuring the outcomes of the implementation of CLIL in terms of the student's performance concerning both language communicative skills and subject-specific content of their target disciplines. More specifically the study was conducted to test the hypothesis: that CLIL promotes better performance in terms of developing both

communicative skills and mastering subject-specific content of the target disciplines in L2.

## **2. Methods**

### **2.1 Participants**

The participants involved in the experimental implementation of the CLIL were English language learners (N=50) assigned to two research groups, the experimental (N=25) and the control group (N=25) based on alphabetical order, an arrangement considered random enough to suit the research purposes of the present study. The distinguishing feature between the two groups was that of the instructional context since the experimental group was presented with English as a vehicular language (CLIL) whereas the control group received teaching in English as a curricular subject (non-CLIL).

### **2.2 Procedures**

#### **2.2.1 The Observation Phase**

Before the experiment, there was a classroom observation during which an observational checklist was filled out to have an idea of how the English language was being taught at ILANI. It resulted that the teaching/learning process method did not consider the aspects of CLIL.

#### **2.2.2 The Experimental Phase**

The experiment was carried out, for a period of 12 weeks (three months), on 50 students of ILANI, located in the city of Buea, in the South West Region of Cameroon. This was based on a CLIL project inspired by Coyle (2005) to build essential qualities for communicative competence for academic and professional success in learners; to discover the essential knowledge and skill base required in their different fields and which can be applied to real-life situations, to verify the impact of the implementation of CLIL on their performance.

After using conventional methods and CLIL to teach the control and experimental groups (25 students in each group) respectively, the two groups' performances were compared. The experimental design comprised the following major steps:

1-Assigning control and experimental groups: 50 students of ILANI for an English language course of over nine months were selected and trained for five months up to the intermediate level. After this period, they were evaluated and later assigned to both the control and experimental groups, using a stratified random sampling technique. This technique ensured that both groups (of 25 students each) had similar characteristics such as gender, first language, nationality, level of education, and age following their test performance and a class list that had such information. The control group was labeled "Group A" and the experimental group "Group B".

2- Pre-testing: Within the first month of the experiment, both groups studied the English language under two tutors (guided by the researcher) who taught alternatively using the same conventional methods and content, from 8 am to 10 am and from 10 am to 12 noon, from Monday to Friday. The above situation was to ensure that both groups were exposed to the same conditions (time of the day, methods and contents, teachers, learning environment, mixed grouping in terms of age, gender, language background, etc.). At the close of the first month of teaching, a test was given to both groups on the same day and at the same time (8 am to 10 am) and the scores were recorded, to determine the degree to which the control and the experimental groups vary in terms of performance at the test. The logic of this step was that if the pre-treatment measures or scores (before introducing CLIL) were identical or nearly so, despite non-equivalence between the experimental (Group B) and control (Group A) groups, we would feel more comfortable using the control group as a control.

3- The treatment step: CLIL was introduced exclusively to Group B as a language teaching-learning approach for three months in the training of both groups, ensuring that all conditions that existed before this step were found in both groups throughout the experiment. This was to avoid other factors besides teaching methods influencing performance in both groups. The CLIL approach was applied to the experimental group during this third step.

### **2.3. The Research instruments**

#### **2.3.1. The CLIL test**

A test instrument was developed as part of the experimental research project to measure the effect of the CLIL program. The aim of the test, which was administered as a pre-and post-test, was to assess the learners' performance in content and language related to their academic and occupational pursuits as reflected in the objectives of the CLIL program. The test served the purpose of a typical achievement test, intended to collect data that revealed how much of the syllabus had been mastered. However, an attempt was made to overcome common approaches to developing a test instrument that solely functions as an assessment tool, measuring students' achievement in an end-of-term exam to assign them a passing or failing grade. The underlying principle for its construction was to assess the ability of test-takers who would need to use the TL in contexts where English is established as the major language for communication. The tasks employed in the test aimed to reflect the purpose for which communication would be effective for academic/professional purposes in the real world. The starting point for the design of the tasks was a consideration of how the text, written or spoken would be used by a learner. Thus, every effort was made so that the tasks involved the students in communication processing operations, which represented to the most possible degree the kind of communicative skills the learners needed either for study or professional purposes. These involved: understanding the overall message of a written or oral text; deciding on an appropriate course of action based on information in the text or a speech; recognizing the attitudes and emotions of the writer/speaker expressed implicitly or overtly in the text/speech; recording the development of an argument; locating and understanding specific information. The task formats comprised gap-filling, note completion, multiple-choice and multiple-matching tasks.

#### **2.3.2. The Students' Interview**

Insights into the learners' attitudes towards the TL and the implementation of CLIL were provided through the interviews (N=25), which researched the learners' attitudes in two instructional contexts, CLIL and non-CLIL upon the completion of the academic term.

The students were randomly selected based on their performance to present a representative sample of the target population.

#### **2.4. The Data Analysis Techniques**

The data emanating from the pre-and post-intervention CLIL test were processed statistically to measure the effectiveness of the CLIL program. The data analysis techniques employed were mean and standard deviation (SD) and ANOVA. The independent variables were the experimental and control group during the pre-and post-intervention. The dependent variables were the subjects' mean scores in pre-and post-tests. The verbal data of the semi-structured interviews, which were analyzed qualitatively, underwent the procedures of data reduction, first and second-level coding as well as pattern coding. Codes resulted in groups of categories, 'labeled' by a specific name (Miles & Huberman, 1994). Then, similar concepts with common characteristics were clustered into themes, to reduce the number of categories.

### **3. Findings and Discussion**

#### **3.1 Data collected during the diagnostic test (pre-test)**

The diagnostic test permitted obtaining the first view of students' communicative language abilities for written and spoken production, as well as their effectiveness in the use of other language functional skills.

##### **3.1.1 Diagnostic test in the control and experimental group**

This section presents the general results in each of the assessment activities of the test and the total scores obtained for the control and the experimental group. The results are presented per skill and participant. The results are later on summarized by calculating each group average for each of the skills and the overall average aggregating performance for all the skills. The overall average is calculated for individual participants and each of the entire group.

**3.1.1.1 Diagnostic test in the control**

The results of the diagnostic test in the control group per scale, skill, and per participant are compared in figure 1 below and portray a very high level of variability in scores among the participants.

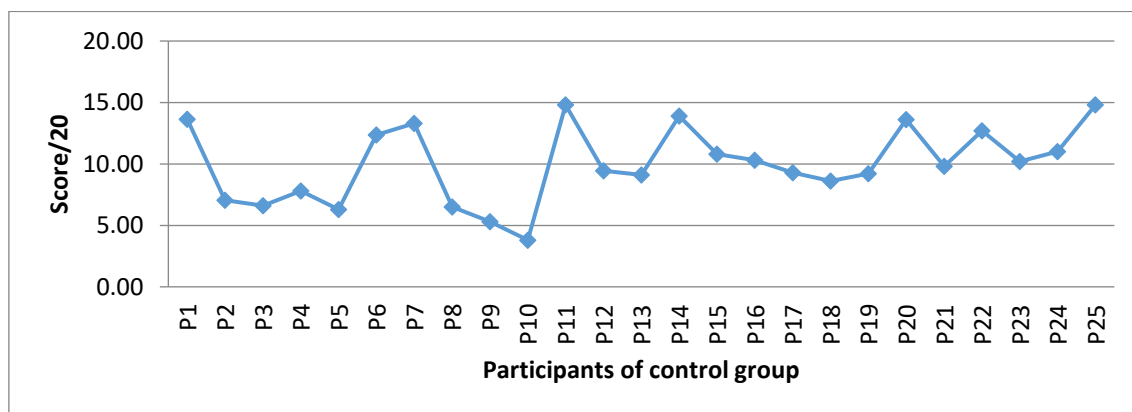


Figure 1: Comparing participants' scores at diagnostic tests in the control group.

The case summaries for the control group performance at the diagnostic test, as the record reveals the averages ranged from 1.10 (being the lowest for strategic competence) to 2.43 (being the highest).

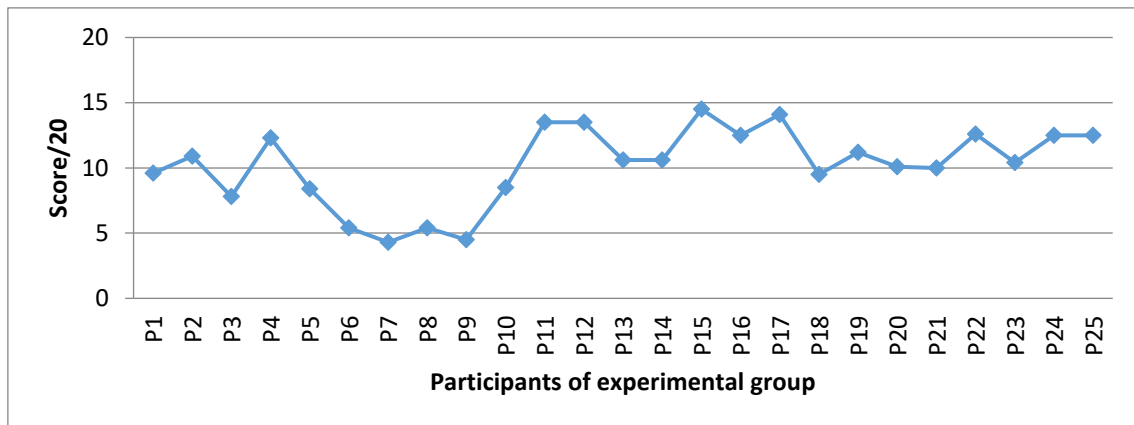
For the overall scale, the average was  $10.01 \pm 0.62$  on a scale of 20. The median was 9.80 thus, implying that half of the group performed below average.

As for the dispersion measurements, the range was 11.0, which was very high indicating therefore that the gap between the lowest and the highest score was too big. The lowest mark was 3.80 and the highest was 14.80. The standard deviation (SD) was equally high, with a value of 3.09 which indicates high variability in the performances of the participants.

**3.1.1.2 Diagnostic test in the experimental group**

The diagnostic test scores in the experimental group per scale, skill, and participant are compared in figure 2 below. One of the most evident tendencies is the high variability in the participants' scores.





**Figure 2: Comparing participants' scores at diagnostic test in the experimental group**

The case summaries for the experimental group performance at diagnostic test per scale indicate that the score of the participants on average ranged from 0.80, being the lowest for competence speaking, to 3.17 for identifying (6) being the highest.

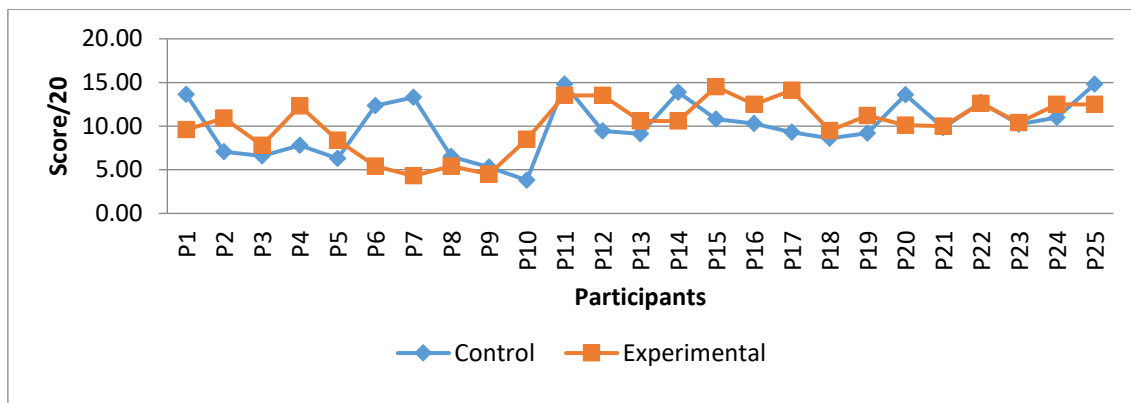
As for the overall scale, the average was  $10.21 \pm 0.59$  on a scale of 20. The median was 10.60. This implies that almost equal proportions had performed below and above average.

As for the dispersion measurements, the range was 10.20, which was very high, indicating that the gap between the lowest and the highest score was too big. The lowest mark was 4.30 and the highest was 14.50. The standard deviation (SD) was equally high, with a value of 2.95 which indicates high variability in the performances of the participants.

### ***3.1.2 Comparative analysis of diagnostic test results (control and experimental)***

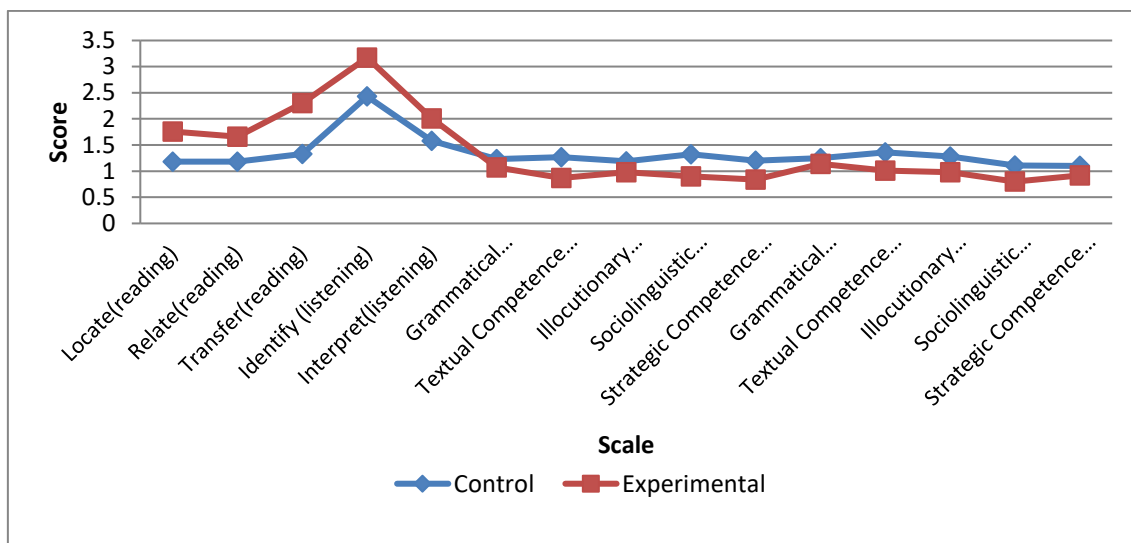
This section compares the control and experimental group results in the diagnostic test.

This comparison is done based on the various skills and the overall score. The overall score was diagrammatically compared for individual participants and then for the entire group.



**Figure 3: Comparing diagnostic tests for the overall score by individual between the control and the experimental group.**

At the diagnostic test, the score of the participants varied highly in both the experimental and the control groups with an SD of 3.09 and a range of 11.0 for the control group and an SD of 2.95 and a range of 10.20 for the experimental group.

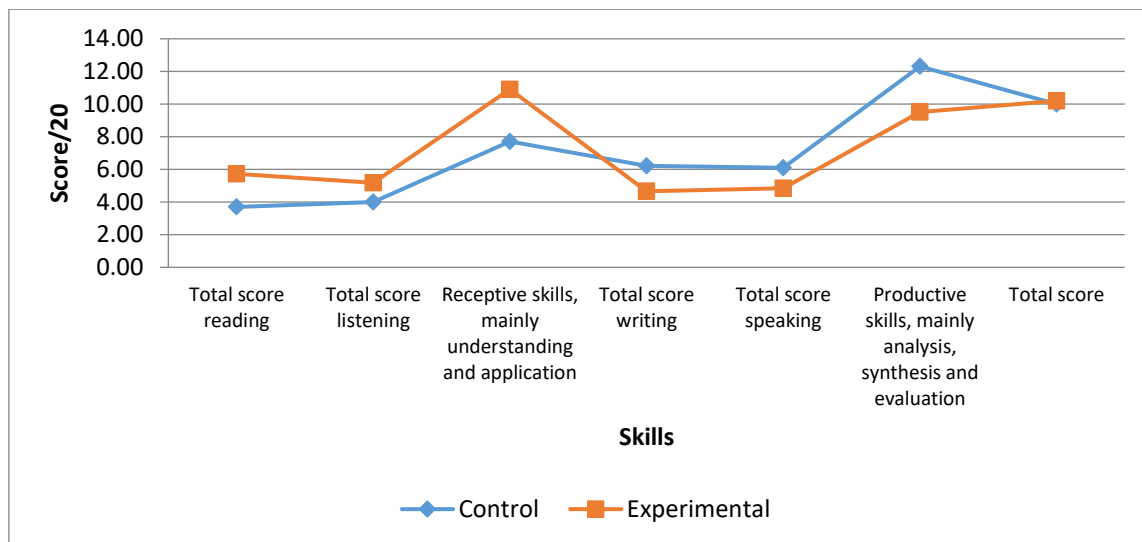


**Figure 4: Comparing diagnostic tests for the various skills between the control and the experimental group.**

It appears that both groups did better in receptive skills as compared to productive skills. The higher dispersion scores in both groups as indicated earlier show that each individual in these groups differs from their peers in the level of development of their language

skills, despite their belonging to the same group. When the comparison is based on the various skills, it appears as indicated on the chart that in both groups, participants performed the highest in identifying (6) with average scores of  $2.43 \pm 0.31$  for the control group and  $3.17 \pm 0.30$  for the experimental group.

Figure 5 gives a clearer representation of the case summaries comparing the control and experimental groups' performances at the diagnostic test for the major skill categories.



**Figure 5: Comparing performance based on the various categories of skills between the control and the experimental group at the diagnostic test.**

Comparing the various categories of skills between the control and the experimental group, it appeared at the diagnostic test that the overall performance in the control group gave the average score of  $10.00 \pm 0.62$  which was not significantly different from the  $10.20 \pm 0.59$  for the experimental group. To conclude, the difference was not significant between the two groups ( $P > 0.05$ ) at the diagnostic test thus, implying that they performed almost the same.

However, in both groups, as indicated earlier, the performances varied highly among the participants given the high ranges and SD obtained in both groups.

When comparing based on the major skill categories, the experimental group performed significantly ( $P < 0.05$ ) higher in the receptive skills with an average of  $10.90 \pm 0.619$  as compared to  $7.71 \pm 0.72$  for the control group.

The trend was the opposite for the productive skills whereby the control group performed significantly ( $P < 0.05$ ) higher with an average of  $12.31 \pm 0.75$  as compared to  $9.51 \pm 0.60$  for the experimental group.

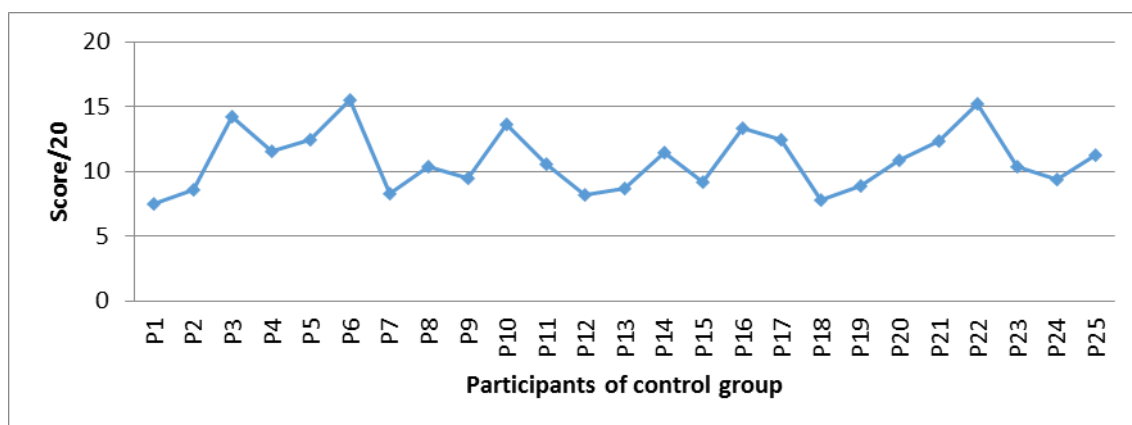
### **3.2 Data collected during the achievement test (post-test)**

Once the course was finished, the achievement test helped to obtain the necessary data to confirm whether or not an improvement in the use of language cognitive skills and communicative language abilities had taken place with the implementation of CLIL.

#### **3.2.1 Achievement test in the control and experimental group**

This section presents the general results in each of the assessment activities of the test and the total scores obtained for the control group and the experimental group. The results are presented per skill and participant. The results are later on summarized by calculating the group average for each of the skills and the overall average aggregating performance for all the skills. The overall average for the achievement test in the control and the experimental group is calculated for individual participants and the entire group.

##### **3.2.1.1 Achievement test in the control**



**Figure 6: Comparing participants' scores on achievement tests in the control group**

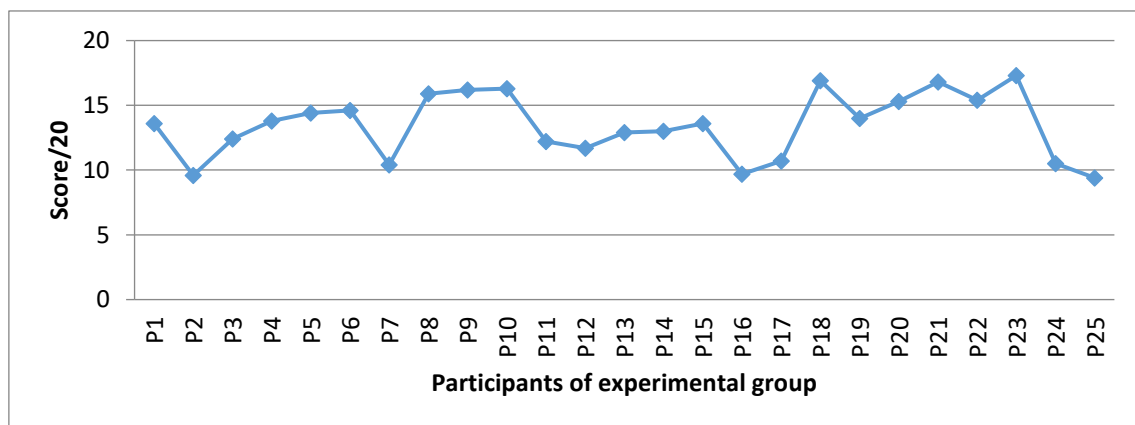
In the control group, at the achievement test, the score of the participants on average ranged from 0.98 being the lowest for Illocutionary Competence (writing) to 2.96 for identify (6) being the highest.

Concerning the overall score, the average was  $10.87 \pm 0.46$  on a scale of 20. The median was 10.60 thus, implying that almost equal proportions had performed below and above average.

As for the dispersion measurements, the range was 8.0 which was very high, indicating therefore that the gap between the lowest and the highest score was too big. The lowest average was 7.50 and the highest was 15.50. The standard deviation (SD) was equally high, with a value of 2.31 which indicates high variability in the performances of the participants.

### 3.2.1.2 Achievement test in the experimental group

The scores of the achievement test in the experimental group per scale, skill, and per participant are compared in figure 7 below.



**Figure7: Comparing participants’ scores at achievement tests in the experimental group**

In the experimental group, at the achievement test, the score of the participants on average ranged from 1.25, being the lowest for Textual Competence (writing), to 4.04 for identify (6) being the highest.

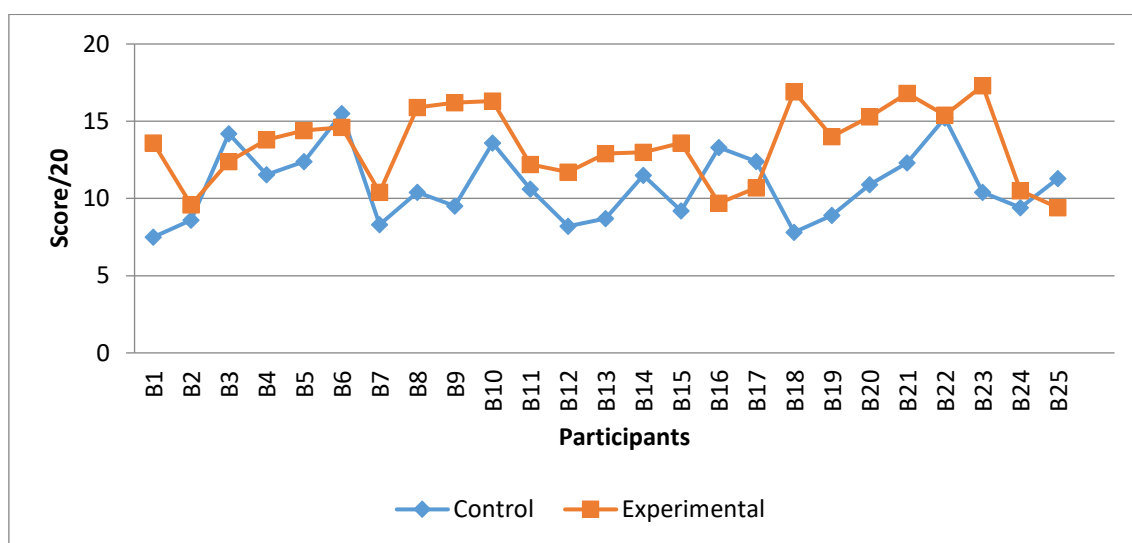
The average overall score was  $13.46 \pm 0.49$  on a scale of 20. The median was 13.60 thus, implying that the majority had performed below and above average.

The range for the dispersion measurements was 7.9, which was very high indicating therefore that the gap between the lowest and the highest score was too big. The lowest average was 9.40, and the highest was 17.30. The standard deviation (SD) was equally

high, with a value of 2.47, which indicates high variability in the performances of the participants on the achievement test as well.

**3.2.3 Comparative analysis of achievement test results (control and experimental)**

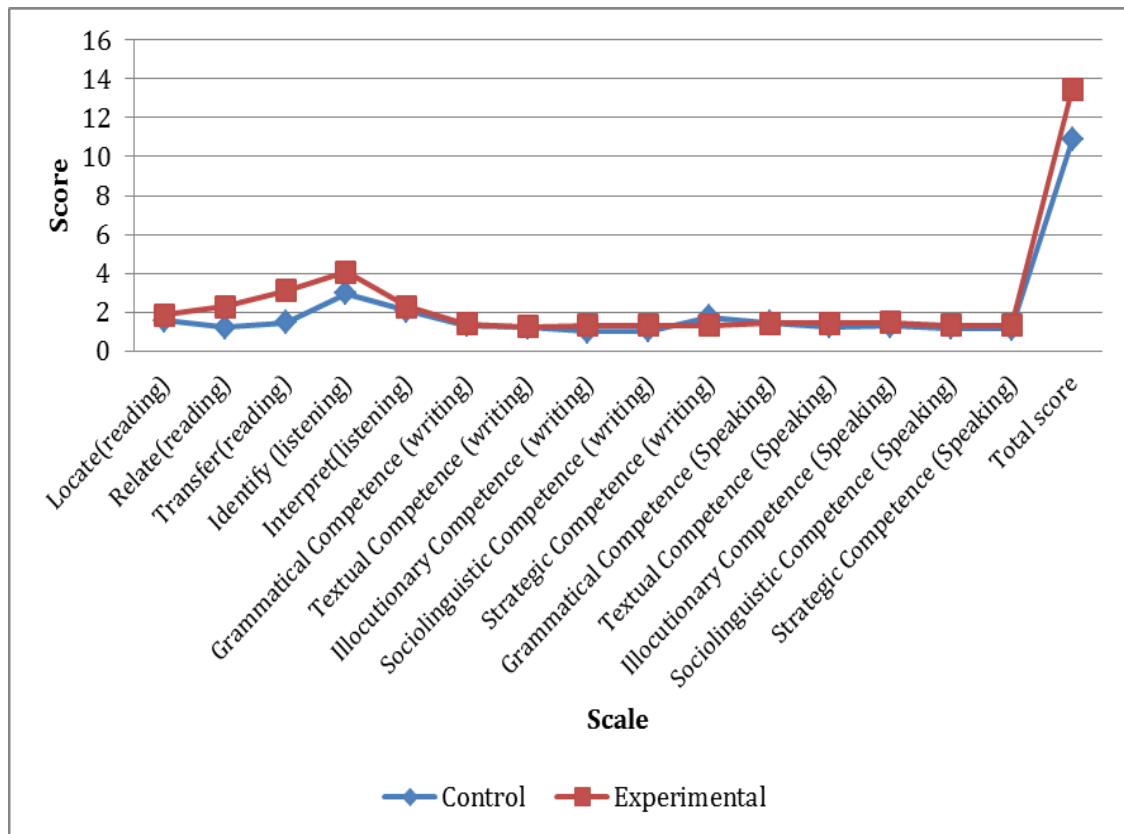
This section compares the results of the achievement test for the control and experimental groups. This comparison is done based on the various skills and the overall score. The overall score was diagrammatically compared for individual participants and then for the entire group.



**Figure 81: Comparing diagnostic tests for the overall score by individual between the control and the experimental group.**

At the diagnostic test, the score of the participants varied highly in both the experimental and the control groups with an SD of 2.31 and a range of 8.00 for the control group and an SD of 2.47 and a range of 7.9 for the experimental group. It is however important to realize that in both groups the variability or dispersion in performances had reduced at the achievement test.

The case summaries comparing the control and experimental groups’ performances at the achievement test per scale are represented in Figure 9 below.

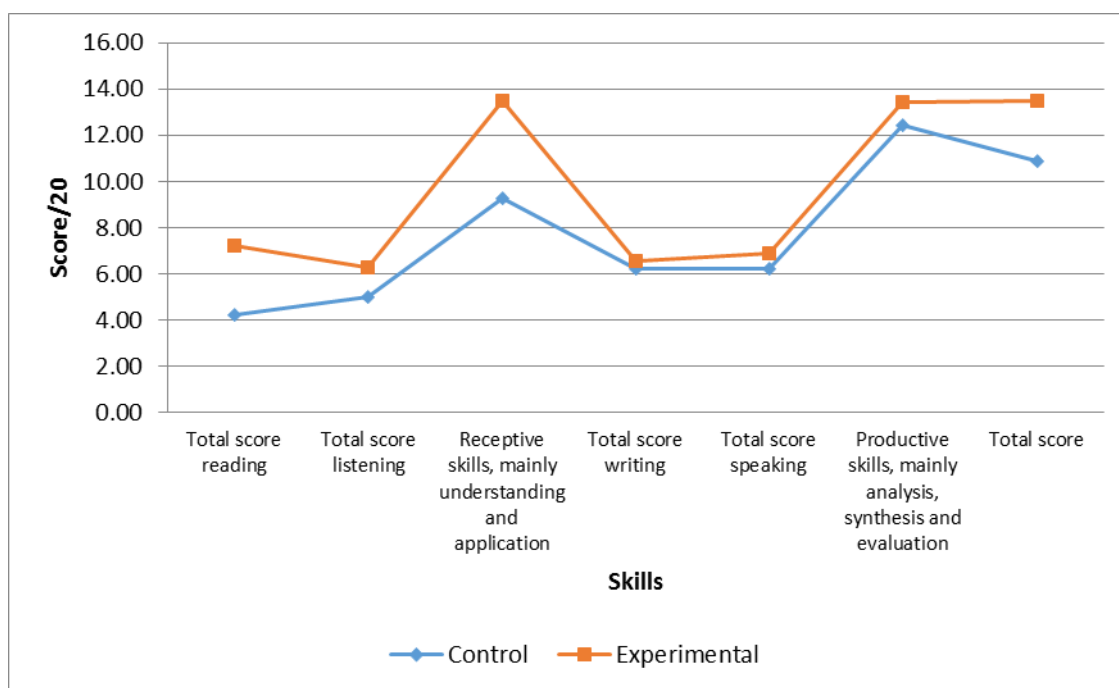


**Figure 9: Comparing performance based on the various categories of skills between the control and the experimental group at the achievement test.**

Comparing the various categories of skills between the control and the experimental group, it appeared at the achievement test that the overall performance in the control group gave the average score of  $10.87 \pm 0.46$ , significantly lower as compared to  $13.46 \pm 0.49$  for the experimental group. Thus, the difference was significant between the two groups ( $P < 0.05$ ) on the achievement test, implying that the experimental group performed significantly better.

However, in both groups, as indicated earlier, the performances varied significantly among the participants given the high ranges and SD obtained in both groups though a slight reduction was observed as compared to the diagnostic test.

**3.2.4 Integrative comparative analysis for the achievement test between groups**



**Figure 10: Comparing total scores for control and experimental groups’ participants at achievement test.**

When comparing based on the major skill categories, the experimental group performed significantly ( $P < 0.05$ ) higher in the receptive skills with an average of  $13.50 \pm 0.46$  as compared to  $9.28 \pm 0.48$  for the control group.

The trend was the same for the productive skills and contrasted with the diagnostic test as the experimental group had a higher average of  $13.43 \pm 0.60$  as compared to  $12.46 \pm 0.56$  for the control group though this difference was not significant ( $P > 0.05$ ).

**3.3 Progression score**

**3.3.1 Progression score within groups**

**Control group**

The progression was computed based on the difference in performance between the diagnostic and the achievement test. Progression is a major indicator to assess the effect of an intervention because it helps in appraising the level of improvement from one test level to another.



In the control group, for the overall performance, the average rose slightly from 10.01 at the diagnostic test to 10.87 at the achievement test for a progression of 8.6%, but this improvement or change in performance from pre-test to post-test was not significant ( $P>0.05$ ).

Assessing the progression based on the main skill categories, it appeared that the difference was not significant when comparing the score on the diagnostic test and the achievement test for the perceptive and productive skills ( $P>0.05$ ).

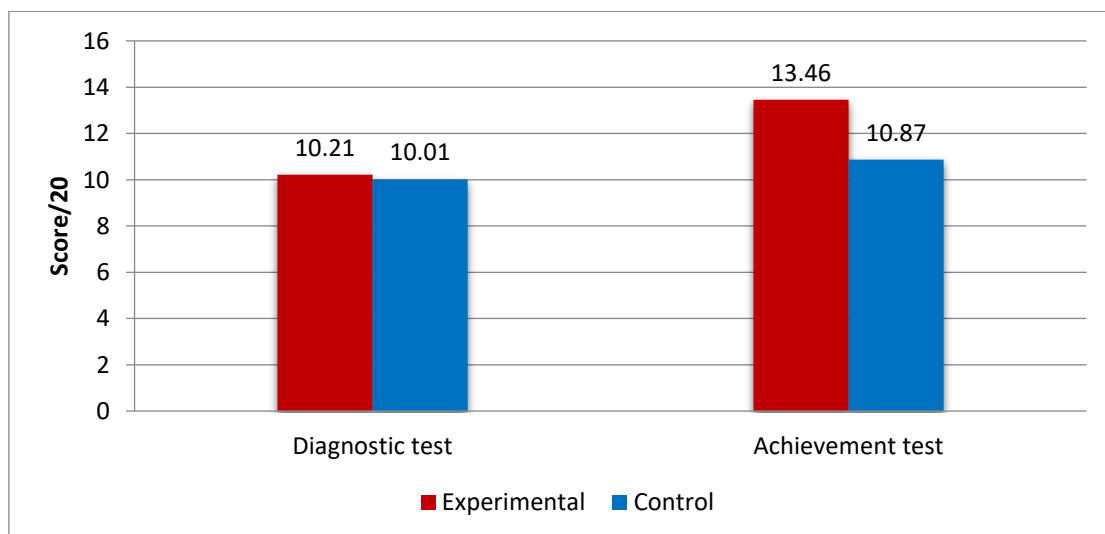
The slight progression here observed in the control group could be attributed to the Hawthorne effect where an initial exposure can influence subsequent appreciation even if no intervention was done, because of the natural recall potential of a human being (Nana, 2015).

### **Experimental group**

For the overall performance in the experimental group, the average rose significantly from 10.21 at the diagnostic test to 13.46 at the achievement test for a progression of 31.8%, and this improvement or change in performance from pre-test to post-test was very significant ( $P<0.05$ ).

When assessing the progression based on the main skill categories, it appeared that the difference was significant when comparing the score on the diagnostic test and the achievement test for perceptive and productive skills ( $P<0.05$ ).

### **3.3.2 Progression score between groups**

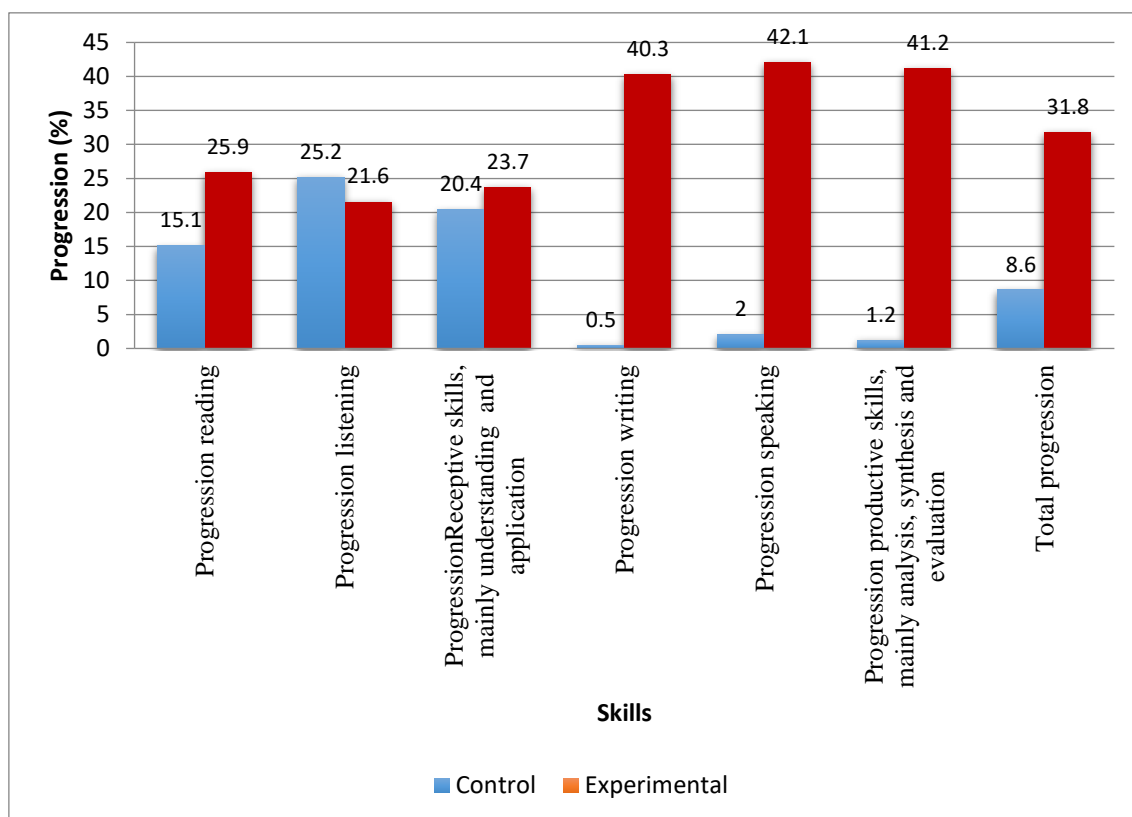


**Figure 11: Comparing performance between control and experimental groups at pre-test and post-test and between pre-test and post-test within the group**

This graph summarizes the overall experiment and presents comparatively the variation in the overall performance from the diagnostic to the achievement test between the control and experimental group.

At the pre-test or diagnostic test, the two groups performed almost the same ( $P > 0.05$ ) with averages of 10.01 and 10.21 for the control and experimental groups respectively.

At the post-test or achievement test, there was a clear gap between the control and the experimental groups with an average of 10.87 and 13.46 respectively ( $P < 0.05$ ).



**Figure 12: Comparing progression between control and experimental groups for the categories of skills**

The progression was higher in the experimental group for all the skills except listening. For reading, it was 25.9% in the experimental group as compared to 15.1% in the control group and this difference was significant ( $P < 0.05$ ) as indicated in the table above.

As for listening, the progression was slightly higher in the control group with a value of 25.2% as compared to 21.6% in the experimental group but this difference was not significant ( $P > 0.05$ ).

The progression for receptive skills was 23.7% for the experimental group and 20.4% for the control group, and this difference was significant ( $P < 0.05$ ).

Concerning writing, the progression was significantly higher in the experimental group with a value of 40.3% as compared to 0.5% for the control group.

The trend was almost the same for listening with values of 25.2% and 21.6% for the control and experimental groups respectively.

The overall progression in receptive skills did not differ significantly between the control and the experimental groups with values of 20.4% and 23.7% respectively ( $\chi^2 = 0.12$ ;  $df = 1$ ;  $P = 0.732$ ).

The overall progression for productive skills was 41.2% for the experimental group as against 1.2% for the control group and this gap was significant ( $\chi^2 = 12.50$ ;  $df = 1$ ;  $P = 0.000$ ).

The overall progression indicates a significant gap between the control and the experimental groups ( $P < 0.05$ ) with values of 8.6% and 31.8% respectively ( $\chi^2 = 4.50$ ;  $df = 1$ ;  $P = 0.033$ ).

## **Conclusion**

The CLIL approach had a more significant effect on the productive skills as compared to the receptive skills and the overall progression was significantly higher in the experimental group. The data provided backed the efficacy of implementing CLIL in a language center as the experimental group outperformed the control group in terms of

both communicative skills and content knowledge and showed a considerably higher positive attitude towards the CLIL course than their peers in the control group.

Based on both the differences in the mean and progression, the CLIL approach had a more significant impact on the performance of the experimental group making it clear that the experimental group had a significantly higher performance as compared to the control group, which suggests that the adoption of CLIL can lead to positive results in the teaching/learning process of English not only in a language center but also in other sections and levels of education. Arguably, the result is consistent with the hypothesis that CLIL promotes better performance in terms of developing both communicative skills and mastering subject-specific content of the target disciplines in L2. There is therefore the need for more experimental research to be done to discover more teaching and learning techniques that will enable to balance of the acquisition of both productive skills and receptive skills in a CLIL classroom.

## **References**

- Christie, F. (1985). Language and schooling. In S. Tschudi (Ed.), *Language, schooling, and society*(pp. 21-40). Boynton/Cook.
- Coyle, D. (2005). *CLIL: Planning Tools for Teachers*. University of Nottingham.
- Coyle, D. (2008). CLIL – A pedagogical approach from the European perspective. In N. Van Deusen-Scholl & N. H. Hornberger (Eds.), *Encyclopedia of language and education, 2nd edition* (pp. 97–111). Springer
- Coyle D., Hood P. & Marsh D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press.
- Crandall, J. (1998). Collaborate and cooperate: Teacher education for integrating language and content instruction. *Forum*, 36(1), 2-9. <http://exchanges.state.gov/englishteaching/forum/archives/1998/docs/98-36-1-a.pdf>
- Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In Leyba, F. C. (Ed.), *Schooling and language minority students: A theoretical framework* (pp. 3-49). Evaluation, Dissemination, and Assessment Center, California State University.
- Currie, P. (1993). Entering a disciplinary community: Conceptual activities required to write for one introductory university course. *Journal of second language writing*, 2, 101-117.
- Dalton-Puffer, C. (2007). *Discourse in content and language integrated learning (CLIL) classrooms*. John Benjamins.
- Dalton-Puffer C. & Smith, U. (2007). Introduction. In C. Dalton-Puffer & U. Smith (Eds), *Empirical perspectives on CLIL classroom discourse* (pp. 7-24). Peter Lang.

- Darn, S. (2006). Content and Language Integrated Learning (CLIL). *A European Overview* (pp.1-7). Izmir University of Economics, Izmir, Turkey: [https://www.researchgate.net/.../234652746\\_contentng](https://www.researchgate.net/.../234652746_contentng).
- García, O. (2008). *Bilingual education in the 21<sup>st</sup> century: A global perspective*. Wiley-Blackwell.
- Hadley, A.O. (2000). *Teaching language in context*(3rd ed.). Thomson-Heinle.
- Hammond, J. (Ed.) (2001). *Scaffolding. Teaching and Learning in Language Literacy Education*. Australia: PETA.
- Harvey, T. Tihinen, T., Määttä, K., & Uusiautti, S. (2013). Content and language integrated learning in practice: Comparison of three Cases. *International journal of academic research in progressive education and development*, 2(2), ISSN: 2226-6348
- Krashen, S, (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Lightbown, P. M., & Spada, N. (2006). *How Languages Are Learned* (3rd ed.). Oxford University.
- Marsh, D. and G. Langé, (Eds.). (1999). *Implementing content and language-integrated learning*. Finland: TIE-CLIL.
- Marsh, D. (2000). *Using languages to learn and learning to use languages*. Finland: University of Jyväskylä.
- Miles. M., & Huberman, M. (1994). *Qualitative data analysis*. Sage Publications.
- Mohan, B. A. (1986). *Language and content*. Addison-Wesley.
- Nana, C. (2015). *Research methods and applied statistics: Beginners and advanced learners*. Goodhead.
- Naves, T. (2009). Effective content and language-integrated learning (CLIL) programs. In Y. Ruiz de Zarobe, & R. Jimenez Catalan, *Content and language integrated learning: Evidence from research in Europe* (pp. 22-40). Multilingual matters.
- Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge University Press.
- Nunan, D. (1991). Communicative tasks and the language curriculum. *TESOL Quarterly*, 25(2), 279-295.
- Shih, M. (1986). Content-based approaches to teaching academic writing. *TESOL Quarterly*, 20, 617-648.
- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning*. Oxford University Press