

Design and Validation of a System of Categories to Assess the Written Discourse of L2 English Learners

SILVIA CORRAL-ROBLES

Universidad de Granada (España)

MICAELA SÁNCHEZ-MARTÍN

Universidad de Murcia (España)

GRACIA GÓNZALEZ-GIJÓN

Universidad de Granada (España)

Received: 21 September 2020 / Accepted: 4 June 2021

DOI: 10.30827/portalin.v0i36.15944

ISSN paper edition: 1697-7467, ISSN digital edition: 2695-8244

ABSTRACT: This study is part of a broader research focused on written discourse analysis and its application in English language teaching. This paper aims to design and evaluate the validity of the content of an instrument (a system of categories) built to analyse the written discourse in L2 (English) of 112 Spanish students in Upper Secondary Education. This study was carried out from a qualitative approach using the technique of content analysis and the Delphi technique with five experts in the field. The results show that the category system has content validity as it meets the criteria to ensure the scientific rigour of qualitative methodology: credibility, transferability, dependence and confirmability. Moreover, there is a high score in terms of interrater reliability on the criteria contrasted with Kendall's W test ($>.07$), as well as on its application in the coding process which was confirmed with Cohen's Kappa test (.658). It can be concluded that this instrument is a valid tool to help ELF teachers and secondary school students in the teaching and learning process of the written discourse in English as a foreign language.

Key words: Content analysis, Delphi technique, system of categories, written discourse, English as a foreign language.

Diseño y validación de un sistema de categorías para evaluar el discurso escrito de estudiantes de inglés como segunda lengua.

RESUMEN: Este estudio forma parte de una investigación más amplia centrada en el análisis del discurso escrito y su aplicación en la enseñanza del inglés. El objetivo principal de este trabajo es diseñar y evaluar la validez de contenido de un instrumento (un sistema de categorías) construido para analizar el discurso escrito en L2 (inglés) de 112 estudiantes españoles de Educación Secundaria Obligatoria (4ºESO). Este estudio se ha realizado desde un enfoque cualitativo, utilizando la técnica de análisis de contenido y la técnica Delphi con cinco jueces expertos. Los resultados muestran que el sistema de categorías tiene validez de contenido, ya que cumple con los criterios que aseguran el rigor científico de la metodología cualitativa: credibilidad, transferibilidad, dependencia y confirmabilidad. Además, existe un alto índice de acuerdo interjueces en los criterios contrastados con la prueba W de Kendall ($>.07$), así como con la aplicación del proceso de codificación, siendo este confirmado con la

prueba Kappa de Cohen (.658). A modo de conclusión, se puede indicar que este instrumento es una herramienta válida para ayudar al profesorado y a los estudiantes de secundaria en el proceso de enseñanza y aprendizaje del discurso escrito en inglés como segunda lengua.

Palabras clave: Análisis de contenido, técnica Delphi, sistema de categorías, discurso escrito, inglés como lengua extranjera.

1. INTRODUCTION

This study is part of wider research focused on written discourse analysis and its application in English language teaching. This more comprehensive research attempted to compare the written proficiency of a group of Spanish secondary school students who followed a CLIL (Content and Language Integrated Learning) programme with a similar group who followed an EFL (English as a Foreign Language) programme. The written production of these groups was analysed and assessed, taking into consideration a system of categories based on four main categories: linguistic competence, discourse competence, sociolinguistic competence and strategic competence.

CLIL is an approach that has become increasingly widespread in the different countries of Europe and beyond (Eurydice, 2017). Numerous studies (Mehisto, Marsh & Frigols, 2008; Pérez-Cañado, 2018) have undertaken this topic in an attempt to assess the effectiveness of this approach. Nonetheless, the study of this approach from the perspective of L2 written discourse is still inconclusive (Dalton-Puffer, Nikula, & Smit, 2010b; Whittaker, Llinares & McCabe, 2011). There is a gap in the literature, as most of this research relies on oral competence (Pérez-Cañado & Lancaster, 2017; Martínez, 2019).

A few decades ago, some linguists defined the written skill as the Cinderella of communicative skills, as it was considered a secondary issue. “Written language has no existence in its own right but is a shadow cast by speech” (Cook, 2004, p. 31). The origin of this issue may lie in the difficulty this skill entails. Writing is considered an arduous task when achieving a good command of it. Moreover, if we add this difficulty to the complexity of writing in a second language, the task seems to be challenging (Nunan, 1991; Alcaráz Varó, 2000; Palmer Silveira, 2001).

McLaren, Madrid and Bueno (2005) state that the difficulties learners face when writing in a second language are due to the complexity of converting ideas into written language, as well as adapting the type of writing to the grammar, lexis, structure and organisation specific to this second language, among others. However, far from claiming that these difficulties can diminish the importance of writing, Camps (1992a) considers that technological progress has given the written language a crucial role in society. Nowadays, it has become a fundamental instrument of social impact. For example, social media can be a tool where students are encouraged to use this skill. In light of this fact, Galera (2001) rightly states that “not mastering it is one of the causes of school failure and certain social marginalisation” (p. 59).

Additionally, Harris (2001) points out, this area has become a very important focus of attention in the study of a second language (L2). In fact, authors such as Raimés (1983) and Porte (1996) indicate that writing skills consolidate and reinforce learned oral language and grammatical structures. As Manchon *et al* argue (2005, p.382), “the language learning potential of writing enables the student to become not only more proficient in regard to

knowing more about the language but also becoming more able at using the language”. This is also set out in the official Spanish curriculum for second language teaching.

This skill is a necessity in today’s world, and it makes sense that more attention should be paid to how students learn to write and how teachers learn to teach writing. Learning how to write cannot be left to chance. Thus, the role of explicit teaching in raising foreign language writing achievement is a key factor.

Employing an instrument such as the system of categories that we present here for the analysis of the written production in English is of great help. Therefore, the main aim of this paper is to describe the process required to design and validate this system of categories by using content analysis, a qualitative research technique.

2. METHODOLOGICAL ASPECTS

In the light of the objective indicated above, this part describes the different steps taken to design the system of categories in addition to the actions taken to validate this system. Thus, the following sections are intended to describe in-depth the research design, the participants who took part in the study and those who helped to validate the system of categories, the instrument and the procedure implemented to collect the data, as well as the data analysis process aimed at exposing the credibility, transferability, dependency and confirmability of the system of categories (Kynğäs & Kääriäinen, 2020).

2.1. Research design

Formulating a pertinent research question, as well as defining the design of the research are considered two crucial steps for the research process. Therefore, in this section, we attempt to describe the research design concerning the two main issues of this paper: the design of the system of categories and its validation.

Regarding the *design of the system of categories*, the written productions of the students are the main source of information for the creation of the system. Thus, in order to discern the meaning of certain significant features of the texts, the qualitative content analysis technique was employed (Kvale, 2011; Flick, 2012). Content analysis is a technique for the analysis of trends in communication contents (Souza, Ferreira, & Gomes 2017). It is characterized by:

- Objectivity: it is defined by the possibility to reach the same results in different studies carried out by different persons.
- Consistency: it refers to the categories that include the content of the text.
- Generalization: it allows to extract conclusions that can be directly extrapolated.
- Quantification: it refers to the quantification of absences and presences of certain aspects in the study.

In relation to the *content or construct validation* of the system of categories, it does not make sense to evaluate the validity of a qualitative research with the traditional criteria used in the quantitative paradigm (Ruiz-Olabuénaga, 2012). Although at present there is still

no unanimous terminology, qualitative research has moved away from the use of positivist terms and guarantees the reliability of its results according to criteria such as credibility, instead of internal validity. It looks at the “truth” value of the research, transferability, instead of external validity. It considers the applicability of the results to other contexts; it also places emphasis on dependency, instead of reliability, which studies the consistency of the data, and confirmation, instead of objectivity, which focuses on neutrality (Kynğäs & Kääriäinen, 2020).

Thus, the validation process was carried out from a qualitative approach, using the technique of content analysis through the Delphi technique of expert judgement. Three rounds of experts were carried out until an acceptable rate of agreement was reached. It was contrasted with Kendall’s W test, and the application of the coding process was confirmed with Cohen’s Kappa test.

2.2. Participants

In this section, we present two groups of participants: those who were involved in the design process of the system of categories, that is, the 112 secondary school students and those involved in the validation process of the system of categories.

2.3.1. Participants involved in the design process of the system of categories

The participants who took part in the wider research mentioned above consisted of 112 students who were enrolled in two public schools in Granada; one of which is located in the city of Granada and the other in a town of the province. They were selected by means of an intentional or deliberate non-probabilistic sample, by criteria of accessibility, personal interest, relevance and adequacy (Tójar, 2006). The written production of these students was the main source of information used for the creation of the system of categories.

2.3.2. Participants involved in the validation process of the system of categories

The validation process of the system of categories was carried out by applying a Delphi-based expert judgment method. Thus, five expert judges from various Spanish universities were involved to assess the content of the system of categories: two specialists in Language and Literature Teaching, one in Pedagogy and two in Research Methods, whose teaching and research experience ranges from 9 to 22 years (table 1).

Table 1. Personal and professional profile of the expert judges

SEX	AGE	DEPARTMENT	UNIVERSITY	PROFESSIONAL CATEGORY	YEARS OF EXPERIENCE
Male	60	Language and Literature Teaching	Granada	Full-time professor	22
Male	49	Language and Literature Teaching	Valencia	Full-time professor	14
Female	44	Pedagogy	Granada	Associate professor	11
Female	45	Research Methods	Murcia	Associate professor	9
Male	43	Research Methods	Murcia	Full-time professor	12

2.3. Instrument to collect the written production of the students

As previously mentioned, the students' written production was the main source of information for the creation of the system of categories. Therefore, the instrument selected to collect this information was a written test previously designed by Madrid and Hughes (2011). It consists of three parts that correspond to three different types of writing: 1) a short email, 2) a story about an accident the student had, and 3) the students' opinion about the school uniform for a school magazine. After carrying out an extensive review of the literature, this test was selected due to its flexibility in helping to assess the written discourse of the students. Moreover, using a previously employed instrument can create a coherent dialogue and a more comprehensive analysis of the phenomenon questioned (Berry, Poortinga, Segall & Dasen, 2002).

2.4. Data collection procedure

In this section, we present two different data collection procedures. The first one is related to the data collection procedure of the written production of the students who participated in the design process of the system of categories. The second part is related to the data collection procedure of the experts' feedback (Delphi technique).

2.4.1. Gathering the written production of the students

This section corresponds to the description of the data collection procedure carried out to gather the written production of the students. Following the references of Souza, Ferreira and Gomes (2017), due to practical purposes, the data collection procedure was divided into three phases: the exploratory phase, the fieldwork phase and the treatment of the information.

The exploratory phase is the preliminary research to clarify the procedure of the investigation in order to be constituted for the fieldwork. In this phase, the problem and the

objectives of the research are delimited, and the study is developed at a theoretical and methodological level. The instrument is chosen and described, the timeframe is set out, as well as the space and the participants of the research. The second phase is based entirely on the available fieldwork. It consists of putting into practice the theoretical construction elaborated in the previous stage. The analysis and treatment of the material, which is the third stage, refers to the set of procedures for assessing, understanding and interpreting the empirical data, articulated with the theory that the project was based on.

In the exploratory phase, the instrument was selected. The fieldwork started in the second phase. For this purpose, an action plan was carried out in order to obtain access to the participating schools. Obtaining the necessary permissions was a laborious task, as parental authorization and voluntary participation were required. We guaranteed their anonymity since the ethical code in any research is crucial. Once this process was finished, the test was distributed to a total of 112 students by one of the researchers of this study. Two days were needed to collect all the data. It was agreed that the teachers and managers responsible for the two different programmes would receive a report with the results.

2.4.2 Gathering feedback from the experts (Delphi technique)

The validation process of the system of categories was carried out by applying a Delphi-based expert judgment method. The Delphi technique is a process used to survey and collect the opinions of experts on a particular subject. Linstone and Turoff (1975) provides a basic definition of the Delphi technique: “Delphi may be characterized as a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem” (p.3).

The process implemented to collect the opinions of the five experts was carried out in three rounds, where the different experts shared their opinions and perspectives on the system of categories. The general tone was favourable, with a positive assessment, alluding to the coherence presented in the system. However, various doubts and weaker points were clarified, such as the exclusivity of certain traits.

These suggestions allowed us to delimit and define the system of categories in a more exhaustive way. This improvement guaranteed us, as we mentioned at the beginning, a useful and valid instrument for the research and for its application in the study of written discourse in a second language. The final result of the category system is presented in table 4.

In the following figure, the process, which is described above, is presented for better clarification:

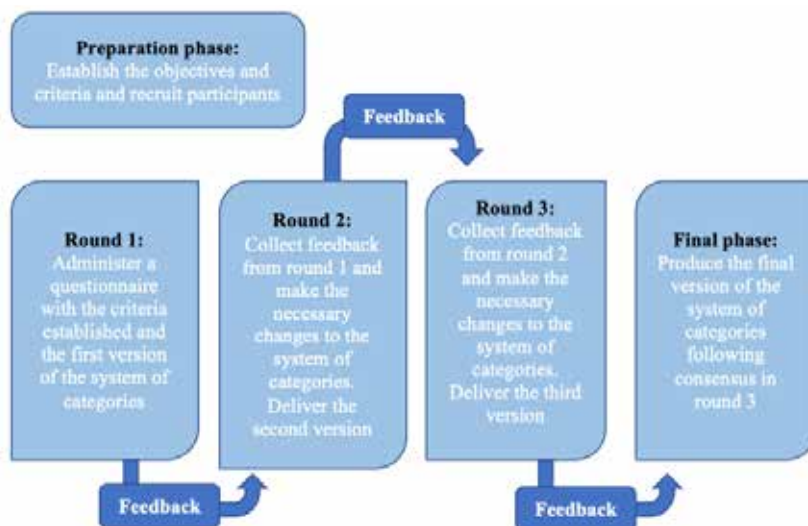


Figure 1. Process to collect feedback from the experts (Delphi technique)

2.5. Data analysis process

The design of the system of categories was carried out through the qualitative technique called content analysis. Qualitative content analysis is not merely a classification of informant opinion, it is actually the discovery of their social codes from stories, symbols and observations. It is the search for understanding and interpretation that provides a unique and contextualised contribution of the researcher (Souza, Ferreira & Gomes, 2017).

This part corresponds to the third phase called the treatment of the information (mentioned in the above section) in which we proceed to the ordering, classification and analysis of the data for the creation of the system of categories. It started once the corresponding phase to the fieldwork had finished.

When the information was collected from the 112 students, the next step was to register the data using the N-vivo11 qualitative data analysis software. This qualitative software allows researchers to deal with a large amount of data and improves the validity of the qualitative research. Moreover, it simplifies the coding and categorising task. The qualitative content technique requires a reflective process to extract the most relevant information from the text by subsuming it into categories and storing it separately for later processing (Navarro & Díaz, 1995).

This is a continuous process of categorising and codifying the information and then return to the raw data, reflect about the analysis and start the process once again (Rodríguez, Gil & García 1996, Bardin, 1996). In this process, reading and rereading is a key aspect as it is the only way to get a sense of the whole. This process shows how the analysis of the raw data progressed toward the identification of the themes that emerge from the performance feedback of the participants. These themes have to be divided into smaller units, known as meaning

units (Erlingsson & Brysiewicz, 2017). The following step is to condense these meaning units into codes. Codifying is a process that allows segregating, grouping and regrouping similar data into categories. Coding meaning in a text into categories makes it possible to quantify how often specific topics are addressed in a text, and the frequency of topics can be compared and correlated with other measurements (Kvale, 2011). The data collection is finished when redundancy of information was obtained, and the categories were saturated; this provides more credibility and transferability to the results (Ruiz-Olabuénaga, 2012).

The process of elaboration of the system of categories integrates different procedures amongst which three approaches are central: *deductive category development*, *inductive category application*, and *validation of the system of categories*. These three approaches are described in the results section.

3. RESULTS

Regarding the *deductive approach*, the analysis starts with relevant findings that serve as a guide for the initial process of codification and categorisation. In this respect, it is worth mentioning that the main theoretical source of the system of categories is the approach of Canale and Swain (1980) based on Communicative Competence and also the classification of errors from Vázquez (1991), Bueno, Carini and Linde (1992). By using and adapting the information derived from these authors, we were able to establish the main categories and codes of our study, which are shown in the following figure:

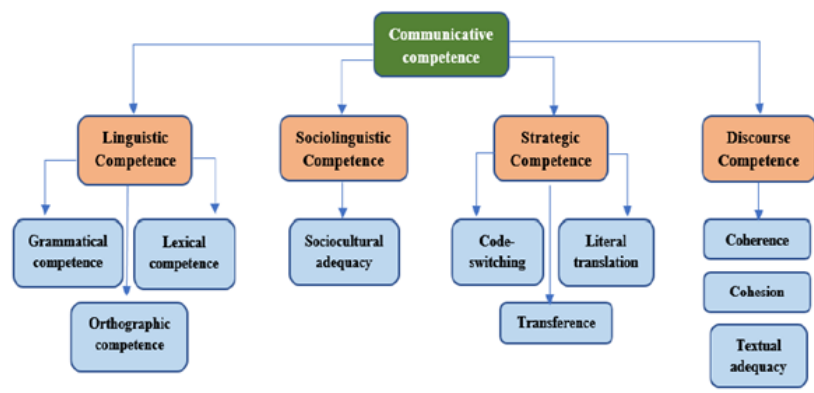


Figure 2: Main categories and subcategories derived from the deductive approach

These four categories: Linguistic Competence (LC), Sociolinguistic Competence (SC), Strategic Competence (SC) and Discourse Competence (DC) and their respective subcategories attempted to cover the different written discourse aspects. Nonetheless, this was just the first step taken in the process of elaboration of the system of categories.

The next step was to apply these initial categories to the written productions in order to observe if they were adjusted to the nature of the research. To do this, we started the categorisation and codification process from an *inductive approach*. This process is based on two rules: formulating the definition of the categories and codes. They have to be defined in advance and can be changed in the elaboration process before the validation of the system of categories.

These definitions serve to establish a selection criterion; therefore, in the reading and rereading process all the information that fitted into the category definitions was accepted, the rest of the information was ignored. This procedure allowed us to create new categories, modify them or even eliminate them from the initial system of categories. As mentioned above, this process was completed using a qualitative software analysis, N-vivo11.

Furthermore, it is worth mentioning the difficulty in obtaining the exclusivity of the categories as each reading unit should fit into only one category on a given scoring dimension. Despite the difficulty, this concern has been solved satisfactorily.

As previously, the process of elaboration of the system of categories integrates different procedures. The deductive category development and the inductive category application have already been explained; thus, we proceed to describe the *validation process of the system of categories*.

Once the category system was created, it was submitted to a content validation process, which refers to how well an instrument measures all facets of a given construct (Oluwatayo, 2012) by applying the Delphi technique of expert judgement. The validation process of our system of categories was carried out by five experts of the field in a discussion group. Their profile is described in Table 1. To carry out this process, the objectives of the study were explained to the experts and they were given a questionnaire in which the categories and codes with the definitions and examples taken from the written productions were included.

In order to establish the internal reliability of the category system, the five expert judges were asked to evaluate the categories and subcategories using a Likert scale (from 0 to 5), according to the criteria set out:

- Homogeneity: each category is obtained from the same principles used for the whole categorisation.
- Exhaustiveness: they give an account of all the material analysed.
- Exclusivity: the content of the material analysed cannot be classified in more than one category.
- Concreteness: they are easily understood, not expressed in abstract terms that admit different meanings.
- Appropriateness: the categories and the objective to be reached are adapted to the content.

The experts carried out three rounds to make the modifications they considered relevant in order to improve the instrument. They were also asked to give improvement proposals for the worst-rated items. The development of the process can be seen in the section of data collection process (figure 1).

For the validation of the category system, the analysis of means and degree of inter-rater reliability was carried out by applying Kendal's W test for ordinal data. These data

were obtained from the expert judgement using descriptive statistics (mean and standard deviation). This process was repeated for the three rounds implemented; however, Table 2 shows the data of the last round due to space and word limitations. Analysing the results following the consensus given in round three, it could be observed that the mean scores obtained for each category were higher than 3, between 3.20 and 4.20 (on a Likert type scale with five response levels).

Table 2. Descriptive statistics

Codes	HOMOGENEITY		EXHAUSTIVENESS		EXCLUSIVITY		CONCRETENESS		ADEQUACY	
	\bar{X}	σ	\bar{X}	σ	\bar{X}	σ	\bar{X}	σ	\bar{X}	σ
1. CLLA	4.00	.70	3.60	1.14	3.60	1.14	3.80	.83	3.60	.89
2. CLGGT	4.00	.70	3.80	.83	4.00	.70	4.00	.70	4.20	.44
3. CLGGF	3.80	.83	3.60	1.14	3.60	1.14	4.00	.70	4.20	.44
4. CLGGC	4.00	.70	3.80	.83	4.00	.70	4.00	.70	3.80	1.09
5. CLGGO	3.80	.83	4.00	.70	3.60	1.14	4.00	.70	4.20	.44
6. CLGEP	4.00	.70	3.60	1.14	3.60	1.14	4.00	.70	3.60	1.51
7. GLGES	3.80	.83	3.60	.89	3.80	.83	4.00	.70	3.80	1.09
8. GLGMP	3.80	.83	3.80	.83	4.00	.70	4.00	.70	3.20	1.64
9. CLGMB	4.00	.70	3.60	1.14	4.00	.70	4.00	.70	3.40	1.51
10. CLGMD	4.00	.70	3.80	.83	3.40	1.51	4.00	.70	4.20	.44
11. CLOPP	3.80	.83	3.60	1.14	4.00	.70	4.00	.70	3.80	1.09
12. CLOPC	4.00	.70	3.80	.83	3.60	1.14	3.80	.83	4.20	.44
13. CLOPE	3.80	.83	4.00	.70	3.40	1.51	4.00	.70	3.60	1.51
14. CLOO	3.80	.83	3.60	1.14	3.80	.83	4.00	.70	3.80	1.09

The results of Kendall's W test with a significance index $p \leq .05$ led us to reject the null hypothesis, that is, there is agreement among experts, with the results shown in table 3.

Table 3. Interrater reliability

DIMENSIONS	N	W DE KENDALL	χ^2	GL.	P
Homogeneity	14	.912	51.059	4	.000
Exhaustiveness	14	.756	42.361	4	.000
Exclusivity	14	.767	34.441	4	.000
Concreteness	14	.950	53.193	4	.000
Adequacy	14	.766	42.907	4	.000

Therefore, all the criteria present a high concordance ($W > .05$), which indicates that the category system has a good internal reliability. In light of these results, the definitive category system was established as shown in Table 4.

Table 4. Definite System of Categories

Linguistic competence (CL)	Lexical competence (CLL)	Lexical adequacy (CLLA)	It refers to the misuse of words or phrases that frequently are employed together with another word or phrase.	
	Grammatical competence (CLG)	Grammatical aspects (CLGG)	Verb Tenses (CLGGT)	It refers to the inadequate use of verb tenses, that is, for instance the confusion between the present perfect tense and the past simple tense.
			Verb Forms (CLGGF)	It refers to the inadequate use of verb forms together with a personal pronoun. For instance: She are*
			Grammatical Agreement (CLGGC)	It refers to the lack of the agreement between the subject and the verb used.
			Verbal Omission (CLGGO)	It refers to the lack of the main or auxiliary verb in a sentence.
		Structure (CLGE)	Standard Position (CLGEP)	It refers to the inadequate use of the structure in the target language.
			Omission of Subject (CLGES)	It refers to the lack of the subject in a sentence.
		Morphosyntactic aspects (CLGM)	Preposition (CLGMP)	It refers to the incorrect use of prepositions in the target language.
			Pronoun (CLGMB)	It refers to the incorrect use of pronouns in the target language.
			Determiners (CLGMD)	It refers to the incorrect use of determiners in the target language.

	Orthographical competence (CLO)	Punctuation (CLOP)	Period (CLOPP)	It refers to the lack or inadequate addition of a period in a sentence.
			Comma (CLOPC)	It refers to the lack or inadequate addition of a comma in a sentence.
			Exclamation mark (CLOPE)	It refers to the lack or inadequate addition of an exclamation mark in a sentence.
		Orthography (CLOO)	It refers to the lack of knowledge of a set of conventions for writing in the target language. It can be caused by interlingual or intralingual transferences.	

Strategic competence (CE)	Code-switching (CEAC)	It refers to the use of words in L1 when writing in L2.
	Transference (CEAE)	It refers to the use of non-existing words in L2 caused by negative transference from L1.
	Literal translation (CEAT)	It refers to the use of expressions whose source, the mother tongue, is causing a deviation in the target language.

Discourse competence (CD)	Coherence (CDH)	It refers to the lack of ability of the user or learner to arrange coherent stretches of language.
	Cohesion (CDC)	It refers to the lack of logic relations when it comes to produce a sentence. It can be identified by the lack of cohesive devices.
	Textual adequacy (CDA)	It refers to the lack of adaptation in a communicative situation.

Sociolinguistic competence (CS)	Sociolinguistic adequacy (CSA)	It refers to the knowledge of sociolinguistic rules and cultural patterns that allow the user or learner to formulate adequate linguistic interventions.
--	--------------------------------	--

When the content validity of the category system was carried out and the interrater reliability verified, we proceeded to check the agreement between judges in the application of this category system. This means that whenever we evaluate the same text with the same category system (the instrument), we obtain the same results. To do that, the Cohen's Kappa coefficient for nominal data was used with the help of the statistical program SPSS v25.

For this purpose, two expert judges were given the same text together with the category system to code it. Once codified, the number of agreements and the number of disagreements were registered in a matrix of the statistical programme for the analysis of quantitative data SPSS V25 and the Kappa coefficient of Cohen was calculated, establishing the following hypotheses:

H_0 : The degree of agreement is 0, i.e., there is no agreement.

H_1 : There is a significant agreement among evaluators, i.e., $K > 0$.

Table 5. Cross Tabulation Judge_2*Judge_1

		JUDGE_1		
		.00	1.00	Total
Judge_2	.00	1	1	2
	1.00	0	48	48
Total		1	49	50

Table 6. Symmetrical measurements

		VALUE	ASYMPTOTIC STANDARD ERROR	T APPROXI- MATED	SIGNIFICANCE APPROXIMATED
Measure of agree- ment	Kappa	.658	.319	4.949	.000
Number of valid cases		50			

The null hypothesis is not assumed.

Use of the asymmetric standard error assumed by the null hypothesis.

As shown in the results of Tables 5 and 6, the H_0 is rejected as the degree of significance is lower than 0.05. Thus, it is concluded that there is agreement among evaluators, with a satisfactory strength of association ($K=.65$).

4. DISCUSSION AND CONCLUSIONS

Writing is one of the least practised skills, due to the difficulty it entails for both teaching and learning (Nunan, 1991; Alcaráz Varó, 2000; Palmer Silveira, 2001). It can be considered as a tool of recognised value that consolidates and reinforces other skills and constitutes an essential element in the learning of a language, as Marsh (2000), Porte (1996) and McLaren, Madrid and Bueno (2005) indicate.

Numerous studies have pointed out the effectiveness of teaching writing explicitly (Manchón & Roca de Larios, 2007) and the fact of using a system of categories related to the different aspects of the written discourse can be highly recommended. As it is proved to be a great learning strategy for students, cognitively attractive and allow them to know more about their strengths and weaknesses in the writing skill (Cassany, 2005).

The results from the reliability and validity testing suggest that the system of categories demonstrates adequate reliability and validity for use by EFL teachers and students to assess the different aspects that comprise written discourse in English language.

As it has been shown throughout the process, different reliability techniques have been used, so the results allow us to consider that the study meets the following criteria, introduced above:

- With regard to credibility: triangulation with pre-existing scientific literature, consultation of various types of documentation (in order to be able to contextualise the data obtained), triangulation of instruments for collecting textual information (e-mail, story and opinions on a student's school subject), review of the information obtained, and analyses carried out on different occasions and by different experts of the field.
- Regarding transferability: the entire research process has been described in detail, listing and indicating all the data. In short, all the elements that have allowed us to describe or interpret the data. The steps in the process of data analysis have been presented.
- With regard to dependency and confirmability: the control process has been carried out by expert judges who are not related to the research.

Nonetheless, this study related to written aspects must be considered in light of two main limitations. First, the categories and subcategories of the system were created in an inductive-deductive process whose main source were the texts provided by students of a high level in high school, 4^o grade. This issue can potentially limit the transferability of the instrument to lower levels. Although, the transferability to similar contexts is indeed guaranteed. Second, there is a research gap in literature related to the process of content validation for systems of categories. Hence, it was a challenging task to find out the best way to do it.

Nonetheless, despite these limitations, the results indicate that the system of categories is a valid, useful and valuable instrument as it has been shown good internal reliability and a satisfactory agreement among experts. Moreover, this study has contributed by providing relevant information for all those English teachers and professors interested in how to treat the different aspects of written discourse in a second language. In addition, it has introduced a better understanding of the designing and validation process of an instrument such a system of categories.

In short, this study presents the design and validation process of an instrument for the analysis of the written discourse of L2 English learners. This instrument is a system of categories whose main objective is to provide EFL teachers with a linguistic and discursive resource that will help them to clarify and solve some everyday problems that they encounter when trying to promote the learning of the written language. It is not only a valid instrument for teachers but also for L2 learners, since analysing their own written production through this system of categories allows them to observe the different shortcomings they have in written discourse.

5. REFERENCES

- Alcaráz Varó, E. (2000). *El inglés profesional y académico*. Alianza Editorial.
- Bardin, L. (1996). *Análisis de contenido*. Akal.
- Berry, J., Poortinga, Y., Segall, M., & Dasen, P. (2002). *Cross-cultural Psychology: Research and Applications*. University Cambridge Press.
- Bueno, A., Carini, C. & Linde, A. (1992). *Análisis de Errores en inglés: Tres Casos Prácticos. Grupo de Investigación Lingüística. Estilística y Didáctica de la Lengua Inglesa*. Editorial Universidad de Granada.
- Camps, A. (1992a). *Hacia una renovación de la enseñanza de la composición escrita en la escuela*. Aula, 2.
- Canale, M., & Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1(1), 1–47.
- Cassany, D. (2005). *Expresión escrita en L2/ELE*. Arco.
- Cook, V.J. (2004) *The English Writing System*. Arnold.
- Dalton-Puffer, C., Nikula, T. and Smit U. (2010b). Language use and language learning in CLIL: Current findings and contentious issues. In C., Dalton-Puffer, T., Nikula and U., Smit. *Language use and language learning in CLIL classrooms*, pp.279-91. (Eds.). John Benjamins.
- Erlingsson, C. & Brysiewicz, P. (2017). A hands-on guide to doing content analysis. *African Journal of Emergency Medicine*, 7(3), 93-99.
- Eurydice. (2017). *New Eurydice report: Citizenship Education at School in Europe*. Eurydice.
- Flick, U. (2012). *Introducción a la investigación cualitativa*. Morata.
- Galera, E. (2001). Aspectos didácticos de la lectoescritura, Granada, GEU. In Camps, A. (1992a). *Hacia una renovación de la enseñanza de la composición escrita en la escuela*, pp.125-136. Aula, 2.
- Harris, T. (2001). L2 Writing: The search for approach. In E. García Sánchez, (ed.) *Present and Future Trends in TEFL*, 235-262. Universidad de Almería,
- Kvale, S. (2011). *La entrevista en investigación cualitativa*. Morata.
- Kyngäs H., & Kääriäinen M., Elo S. (2020). The Trustworthiness of Content Analysis. In: Kyngäs H., Mikkonen K., Kääriäinen M. (eds) *The Application of Content Analysis in Nursing Science Research*, pp.145-162. Springer. https://doi.org/10.1007/978-3-030-30199-6_5
- Linstone, H. A., & Turoff, M. (1975). *The Delphi method: Techniques and applications*. Addison-Wesley Publishing Company.
- Madrid, D., & Hughes, S. (2011). *Studies in Bilingual Education*. Peter Lang.
- Manchón, R.M. & Roca de Larios, J. (2007). On the temporal nature of planning in L1 and L2 composing. *Language Learning* 57, 549-593.

- Manchón, R.M., Murphy, L., Roca de Larios, J. & Aguado, P. (2005). Learning and Teaching Writing in the EFL Classroom. In McLaren, N., Madrid, D. & Bueno, A. *TEFL in Secondary Education*, pp. 377-407. Editorial Universidad de Granada.
- Marsh, D. (2000). *Using languages to learn and learning to use languages*. University of Jyväskylä.
- Martínez, J.D. (2019). Which instructional programme (EFL or CLIL) results in better oral communicative competence? Updated empirical evidence from a monolingual context. *Linguistics and Education*, 51: 69-78.
- McLaren, N., Madrid, D. & Bueno, A. (2005). *TEFL in Secondary Education*. Editorial Universidad de Granada.
- Mehisto, P., D. Marsh & M. J. Frigols. (2008). *Uncovering CLIL. Content and language integrated learning in bilingual and multilingual education*. Macmillan.
- Navarro, P. & Díaz, C. (1995). Análisis de contenido. In Delgado, J. & Gutiérrez, J. *Métodos y técnicas cualitativas de investigación en Ciencias Sociales*, pp.58-67. Síntesis Psicológica.
- Nunan, D. (1991). *Language teaching methodology*. London: Prentice-Hall.
- Oluwatayo, J. A. (2012). Validity and Reliability Issues in Educational Research. *Journal of Educational and Social Research*, 2(2), 391.
- Palmer Silveira, J.C. (2001). Different approaches to academic writing and their applications to the teaching of ESP summarising. In S. Posteguillo Gómez, I. Fortanet Gómez, & J.C. Palmer Silveira (Eds.), *Methodology and new technologies in languages for specific purposes*, pp. 61-75. Universitat Jaume I.
- Pérez Cañado, M.L (2018). CLIL and Educational Level: A Longitudinal Study on the Impact of CLIL on Language Outcomes. *PortaLinguarum*, 29, 51-70.
- Pérez Cañado, M.L & Lancaster, N. (2017). The effects of CLIL on oral comprehension production: a longitudinal case study. *Language, Culture and Curriculum* 30(3), 300-316.
- Porte, G. (1996). The forgotten function of the pen: second language writing as an awareness-raising activity. GRETA. *Revista para Profesores de Inglés*, 44-50.
- Raimes, A. (1983). *Techniques in Teaching Writing*. Oxford University Press.
- Rodríguez, G., Gil, J. & Rodríguez, E. (1996). *Metodología de la investigación cualitativa*. Ediciones Aljibe.
- Ruiz-Olabuénaga, J.I. (2012). *Teoría y práctica de la investigación cualitativa*. Universidad de Deusto.
- Souza, M., Ferreira, C. & Gomes, R. (2017). *Investigación Social. Teoría, método y creatividad*. Lugar Editorial.
- Tójar, J.C. (2006). *Investigación cualitativa. Comprender y Actuar*. La muralla.
- Vázquez, G. (1991). *Análisis de Errores y Aprendizaje de Español/Lengua Extranjera*. Peter Lang.
- Whittaker, R., Llinares, A. & McCabe, A. (2011). Written discourse development in CLIL at secondary school. *Language teaching research*, 15(3), 343-362.