Evaluation of the Index of Similarity Detected by Turnitin® in Research Projects of a Master's Degree in Higher Education

Dennis Arias-Chávez[1] [0000-0003-1500-8366], Teresa Ramos-Quispe[2] [0000-0003-4607-4745], Alberto Patricio Lanchipa-Ale[3] [0000-0002-4873-1123], Elmer Benito Rivera-Mansilla[3] [0000-0002-6107-4164], Juan Enrique Quiroz Vela[4] [0000-0002-3836-0197]

[1]Universidad Continental, Arequipa, Perú
darias@continental.edu.pe
[2]Universidad Nacional de San Agustín de Arequipa, Perú
tramosq@unsa.edu.pe
[3]Universidad Nacional Jorge Basadre Grohmann, Tacna, Perú
{alanchipaa, eriveram}@unjbg.edu.pe
[4]Universidad del Pacífico
je.quirozve@up.edu.pe

Abstract. The objective of this study is to compile samples of errors extracted from research papers that illustrate the different conflicts in the theoretical sections after the application of the Turnitin® software. To do this, 28 theoretical sections were selected, drawn from research projects written by students of a master's degree in Higher Education at a private university in southern Peru in 2020. The design is non-experimental, cross-sectional with a quantitative approach. Descriptive statistics were used to analyze the data. The results show that with regard to the level of similarity found the documents analyzed are at level IV (50-74%), while the types of plagiarism with the greatest presence are "copy and paste" and "search and replace". Among the sources with a high coincidence index are theses and articles from scientific journals, while the section with the highest percentage of similarity and the highest number of cases of plagiarism is the theoretical framework. In this sense, the Turnitin software® is suitable for reporting the degree of similarity in research papers as it helps to detect signs of possible cases of plagiarism.

Keywords: Plagiarism, academic writing, originality, Turnitin**.**

1. Introduction

Intellectual work is an activity that must be done honestly and fairly, accepting the consequences of our actions at all times. One of its pillars is respect for intellectual property through compliance with rules and principles that encourage this action. However, acts that go against this principle have been on the increase, endangering the intellectual activity embodied in research works such as theses and scientific articles [1]. It is common that both teachers and students, regardless of the level or degree of studies, voluntarily or involuntarily omit to give credit to the authors from whom

information is taken, which hinders the intellectual and professional development of whoever commits it [2].

The above describes what is understood as plagiarism, an act that is not always assumed as dishonest behavior but as a normal activity in the university [3]. The internet has contributed to reinforcing this idea since it offers the possibility of accessing a large amount of information, thus making thin the line that separates the legal from the illegal. This means that in the field of "downloadable", anything goes, there are no limits and everything can benefit whoever uses it. The original goes into the background and the field of reproduction is entered without giving credit to those who deserve it. This is a problem whose solution is not only to punish or punish whoever commits it, but also to educate and correct the offenders. This responsibility falls not only on teachers but also on the institutions affected by plagiarism in their prestige and the quality of their processes.

Beyond understanding plagiarism as an infringement of copyright, its practice is a reflection of the absence of academic writing skills that are not always promoted by universities. Writing at the university allows the acquisition and communication of the contents that are studied in the career and thus account for the learning and passing of courses [4], that is, it fulfills an epistemic, rhetorical and enabling function [5].

The use of tools to detect cases of plagiarism has become an ally in the fight against academic malpractices at all educational levels. Turnitin is a commercial product that was launched in 1997 to compare files uploaded by users with those on the Internet, with student documents that are stored in its repository and with indexed databases. Its objective is to ensure academic integrity, and since its creation it has been implementing various services that respond to the academic needs of the institutions that acquire it. Its use helps to detect and prevent possible cases of plagiarism by performing a match search and establishing a similarity index [6] [7] [8]. However, it is not enough to detect--universities need to educate. Detecting cases of dishonesty in time will help to detect the problem and correct it for the future.

1.1 Literature review

Over the years, the need to identify cases of plagiarism in writing processes in the university environment [9] has been increasing. Likewise, the thesis genre is one of those that has aroused the greatest interest among researchers in the areas of linguistics and education, given its role in the academic training of students. This document (thesis) allows, at the most advanced level, to take solid steps to enter the scientific community [10, 11]. This concern has led various authors to focus on proposing preventive and educational measures [12]. Of these, those that focus on the use of tools to detect cases of plagiarism as well as to verify the author's contribution on the subject (originality) stand out. Although there are various systems in the environment that allow evaluating the level of originality of the work carried out by students, Turnitin is one of the most popular and well-known given the alternatives it offers both students and teachers to maintain academic integrity in their scholarly activities [13].

By the same token, the studies that focus on analyzing the effectiveness of Turnitin as a resource to detect plagiarism [14], its impact on laboratory reports and argumentative

essays [15, 16], its use as a resource to support academic integrity to a higher level [17] [18], and its use to evaluate the practices of graduate students in the handling of sources [19]. Likewise, among the studies that have focused on analyzing plagiarism in undergraduate and graduate students, the one by Marsden et al. [20] on dishonest academic behaviors in Australian students, the study by Duff et al. [21] in students of a master's degree in engineering, and the study by Gilmore et al. [22] who conducted a comparative study between master's and doctoral students stand out. Lack of originality and plagiarism are the result of the lack of skills in academic writing, a problem that is sometimes not confronted by universities given the fact that few activities that promote the proper use of authorship recognition are organized. For this reason, software like Turnitin is used rather as a punitive mechanism than a training tool.

2. Method

A non-experimental, cross-sectional study with a quantitative approach was carried out. 28 research projects written by students of a master's degree in Higher Education from a private university in southern Peru in 2020 were selected. The projects were developed in the Research Workshop I course. For the analysis, only the theoretical sections of the projects (Statement of the problem, Justification and Theoretical framework) were analyzed since these allow to clearly distinguish the works of others with their own research [11], as well as to recognize and assess the writing skills (summarize and criticize) applied by the students to build the makeup of their work [23]. The projects were presented at the end of the course. It should be noted that, among the contents developed are those related to the essential aspects of the search and systematization of information as well as the process for the construction of the theoretical and methodological bases of the research, the process to reference the sources of consultation, and the importance of avoiding plagiarism and promoting academic integrity.

The Initial Similarity Index (ISI) report issued by the Turnitin software was used to determine the level of similarity and the type of plagiarism. The ISI can be found on the originality report issued by the software once the documents are uploaded. This report suggests an overall percentage of the student's text that matches the sources in their database and indicates the level of match with a colored (see Fig. 1). To determine the type of plagiarism, the researchers carried out a qualitative evaluation of the document in order to find the cases in which the student did not carry out the corresponding citation process. The cases in which the student correctly applied the process of registration and credit of the source (citation and reference) were excluded from the study. Once the cases of plagiarism had been detected, their quantification and classification ensue with the help of a matrix prepared in Excel and with registration and analysis sheets created for the study on the basis of the categories established by the software (see Table 1).
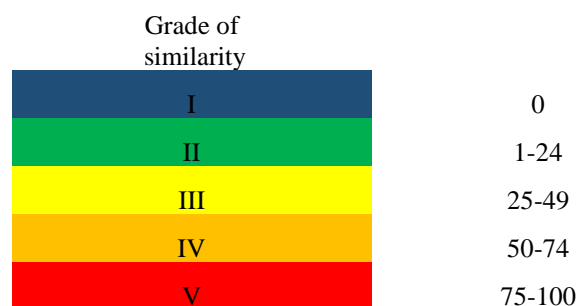
Grade of
similarity

I	0
II	1-24
III	25-49
IV	50-74
V	75-100

Fig. 1. Degrees of similarity used by the Turnitin software.

Table 1. Modalities of plagiarism

Modality	Form in which it is presented
Cloning	Presenting someone else's work as his/her own, reproduced verbatim.
Copy and paste	Including ample text passages from a single source without modifying them.
Search and replace	Changing keywords and expressions without altering the essential content of the sources.
Remix	Mixing paraphrased text extracted from multiple sources.
Recycling	Taking ample passages from a previous work of one's own without proper quotation.
Hybrid	Combining perfectly quoted sources with unquoted fragments or passages.
Mosaic	Material copied from multiple sources that fit well.
Error 404	Citing non-existent sources or including inaccurate information on sources.
Source RSS	Correct quoting of sources, but almost without including paragraphs created by the autor.
Reuse	Correct quoting of sources making too much use of the original text structure and/or words.

Note: Information taken from "10 unoriginal work modalities", by Turnitin, 2020. Retrieved from https://www.turnitin.com/static/plagiarism-spectrum/

3. Results

Percentage of similarity

The similarity index allows the reviewer to determine the student's contribution. Although logic leads to consider that the similarity index must be below the originality index (the latter understood as the content that the student contributes to the theoretical context of the study), various specialists have ventured to determine a permissible

standard percentage of similarity that can be between 15 and 20% without further technical support. The truth is that, whatever the percentage of similarity shown by the software, a qualitative analysis must be carried out to determine whether or not this percentage corresponds to duly cited information. Indicating a standard percentage can lead to create in the student a "false sense of security" [17] that in the long run can generate justifications for committing plagiarism or tolerance towards it. Regarding the level of similarity obtained by the works analyzed in this study, of the 28 projects evaluated, a maximum of 90% and a minimum of 7% were obtained as descriptive value.

The average value was 56.11% with a standard deviation of 22.63%. These results partially coincide with what was found by Tran et al. [13] who determined an average of 42.6% similarity in works carried out by Vietnamese university students, also with the results obtained by Bautista et al. [24] who, after analyzing the master's theses in social sciences, determined a level of plagiarism of 62%, and with the results obtained by Saldaña et al. [25] who investigated plagiarism in theses and in their advisers from the medical career of a public university in Peru, determining that, of 33 theses, 27 (82%) had evidence of plagiarism. Table 2 shows the degrees of similarity obtained.

Table 2. Degrees of similarity

Categories	fi	%
I (0%)	0	0
II (less than 25%)	0	0
III (25% to 49%)	8	28,6
IV (50% to 74%)	14	50,0
V (more than 74%)	6	21,4
Total	28	100,0

As can be seen, 50% of the jobs are at Level IV, a fact that is worrying. Although these results are already indications of possible plagiarism, when the students were notified, many of them argued that what had been presented "was a first version and that this would be corrected already in the thesis." Others, for their part, indicated that given the urgency to present the works "they had taken information from the sources without stopping to cite or reference the sources." These arguments support the idea that the theoretical component is complex since it implies the search and systematization of information, an action that few students claimed to have done.

Types of plagiarism

Regarding the types of plagiarism, these can be seen in Table 3. Plagiarism by "copy and paste" is presented as the type with the highest frequency. This result is repeated at other levels as in the case of high school students [12]. Likewise, the "search and replace" type (27.08%) is the second most frequent. These results are striking given that they are evidence of how little effort the students had put into creating their own theoretical content. The fact of copying entire fragments (including the quote) shows a rush to complete the task, and if searching and replacing information could mean a

slight effort to adapt the information, both types show an intention to plagiarize. When consulting the students about these cases, the arguments they provided were on the side of including the citation, that is, that it is not plagiarism because it was cited, despite the fact that the similarity report indicates that the entire component (text and abbreviated reference) had been taken from sources that the software accurately recognized. Although the software shows a possible case of plagiarism, it is important to point out that the differences between direct and indirect quotes were addressed during the sessions. One of the indications given in classes was that, if a text fragment is included, it is necessary for the author of the work to attach the fragment and add to the abbreviated reference, in addition to author and year, the page or pages, as this would provide a solution to the problem in a formal way.

Table 3. Types of plagiarism found

Type	Fi	%
Cloning	1	2.08
Copy and paste	25	52.08
Search and replace	13	27.08
Remix	2	4.17
Recycling	0	0.00
Hybrid	0	0.00
Mosaic	7	14.58
Error 404	0	0.00
Source RSS	0	0.00
Reuse	0	0.00
Total	48	100.00

Sources with the highest percentage of similarity

One of the options offered by Turnitin is to show the source whose information matches the one that appears in the theoretical section of the work. Turnitin compares the information with three databases or repositories: resources taken from the internet, student work stored in the system repository, and indexed academic scientific content. This option is relevant since it allows the institution to verify the origin of the content taken by the student and go, if need be, to the source itself. Another important aspect is that, as part of the contents of the Research Workshop I course, contents such as the search for information in academic databases and activities are included in which students must present a preliminary list of the scientific literature that will serve them as the basis for writing his/her thesis project. When evaluating which of the sources of origin of the information taken has the greatest presence, theses are placed first as one of the genres most consulted by students, followed by scientific journals (see Fig. 2).

Likewise, web pages and Wikipedia rank last with lower percentages. These data are positive, especially if one takes into account that the use of Internet resources is common not only in students in the training process but also in professionals from various specialties who prefer these sources because they are of immediate consultation, especially in the case from Wikipedia [26, 27, 28] and others such as Blogspot, Prezi and Scribd [24].
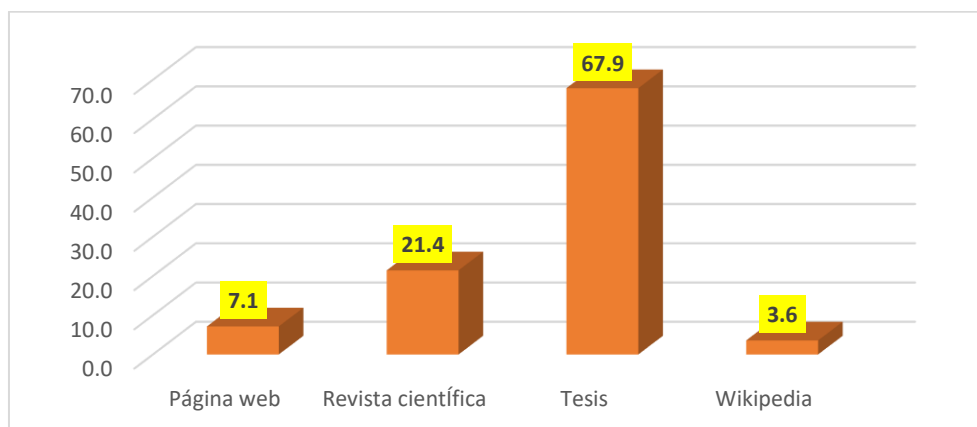

Fig. 2. Source type with higher percentage of similarity.

Theoretical sections

The purpose of the theoretical section of a thesis is to present the theoretical panorama of the chosen topic as well as the information gaps and justification for carrying out the study [29]. Authors such as Paltridge and Starfield [23] highlight how complicated it is to elaborate this section since its development requires, in addition to searching for information, reading it in order to prepare summaries, comments and critical analysis of the contents. Of the three components that were analyzed in the present study, the one that presented the greatest complications was the Theoretical framework, since this requires the use of one's own sources and external sources of information, which, in the words of Phillips and Pugh [30], allow the reader to recognize that the author has a command of the subject and the knowledge of the discipline in which the study belongs. In that sense, this section is the most vulnerable to plagiarism. This fact is demonstrated in the present study, where the theoretical frameworks concentrate the highest percentage of similarity (86.7%), as can be seen in Table 4.

Table 4. Sections with the greatest presence of similarity

Section	fi	%
Problem exposition	3	10.0
Justification	1	3.3
Theoretical framework	26	86.7

4. Conclusions

Any proposal that seeks to counteract dishonest behavior in the university must focus, above all, on understanding the phenomenon, which is not an easy task. Understanding the nature and facets of dishonest acts will help make better decisions at the institution level [20]. A cultural component underlies the way in which people perceive acts of academic dishonesty such as plagiarism, that is, the promotion and practice of certain habits that are dragged from high school to university and even at higher levels such as postgraduate [twenty-one]. Although it could be considered that, due to the years of student life in university, graduate students may have greater awareness of the norms and conventions for the avoidance of acts such as plagiarism, this is not always true, which leads teachers to insist on influencing the procedures for writing and referencing consultation sources [22].

The objective of this study was to collect samples of errors extracted from research papers that allow to illustrate the different conflicts in the theoretical sections after the application of the Turnitin software among posgraduate students. Regarding the level of similarity found, the analyzed documents are at level IV (50-74%), a result that shows possible cases of plagiarism as well as barely a small contribution from the student to the theoretical construction of the project. Regarding the type of plagiarism, "copy and paste" and "search and replace" are presented as the types most frequently resorted to. These results could be overcome by applying a correct handling and interpretation of the similarity index: reducing the tendency to easy things and developing sufficient skills for the construction of the theoretical section of the thesis project [31]. Among the sources with the highest rate of coincidence are theses (first place) and articles from scientific journals (second place), whereas sources such as web pages and Wikipedia are in the last place. The section that concentrates the highest percentage of similarity and the highest number of cases of plagiarism are the thesis' theoretical bases, a foreseeable section since it is in this section where the author must combine his/her speech with that of others in order to give theoretical support to his/her proposal.

The evidence allows us to conclude that there are serious shortfalls in the management of sources, in the citation process and in the management of strategies and skills in the writing of academic texts by students. Although the process of searching and processing

scientific information is part of the course content, it is necessary to implement as a rule the feedback on the use of the software and on the implications of committing acts of plagiarism, which will lead students to apply and foster the principles of academic honesty, which will help promote originality in academic writing. Finally, it is recommended, to avoid cases of academic fraud such as plagiarism, to establish mechanisms for regulation, monitoring, detection and training from the early years in university. These mechanisms must be included in the regulations and academic guidelines as part of the academic integrity policies of the university. Likewise, it is advisable, as a preventive measure, to promote and regulate the use of plagiarism detection software, developing regulations for use consistent with the practice of academic writing with the help of digital resources such as plagiarism detection software, also training teachers, researchers and consultants of thesis on the use of these tools not only in the classroom but also in research activities. Taking this into consideration, the results of this study are important since they allow us to have evidence of a phenomenon that is not exclusive to the academic field but also to others, such as politics, advertising and communications.

References

1. Comas, R., Sureda, J.: Academic plagiarism: explanatory factors from student's perspective. Journal of Academic Ethics, (2010) 8(3), 217-232. DOI: https://doi.org/10.1007/s10805-010-9121-0
2. Lugo-Machado, J., Jacobo-Pineli, R.: (2020). El Plagio en Pregrado y Posgrado: Revisión Narrativa Artículo de Introspección. Revista de Medicina Clínica (2020) 4(2), 68-72. DOI: https://doi.org/10.5281/zenodo.3873508
3. Ramos Quispe, T., Damian Nuñez, E., Inga Arias, M., Arias Chávez, D., Caurcel Cara, M.: Actitudes hacia el plagio en estudiantes de Administración de Empresas de dos universidades privadas en Arequipa. Propósitos y Representaciones (2019) 7(1), 33-58. DOI: http://dx.doi.org/10.20511/pyr2019.v7n1.264
4. Gardner, S., Nesh, H.: A classification of genre families in university student writing. Applied Linguistics (2013) 34(1), 25-52. DOI: https://doi.org/10.1093/applin/ams024
5. Navarro, F.: Más allá de la alfabetización académica: las funciones de la escritura en educación superior. *In:* M.A. Alves, V. Iensen Bortoluzzi (eds.): Formação de Professores. Ensino, Linguagens e Tecnologías, Porto Alegre (2018) 13-49. Available at; https://www.researchgate.net/publication/326377982_Mas_alla_de_la_alfabetizacion_aca demica_las_funciones_de_la_escritura_en_educacion_superior
6. Bruton, S., Childers, D.: The ethics and politics of policing plagiarism: A qualitative study of faculty views on student plagiarism and Turnitin®. Assessment & Evaluation in Higher Education *(2016) 41*(2), 316-330. DOI: https://doi.org/10.1080/02602938.2015.1008981
7. Henderson, P.: Electronic grading and marking: A note on Turnitin's GradeMark function. History Australia (2008) 15(1), 11.1-11.2. DOI: https://doi.org/10.2104/ha080011
8. Sutherland-Smith, W., Carr, R.: Turnitin.com: Teachers' perspectives of anti-plagiarism software in raising issues of educational integrity. Journal of University Teaching & Learning Practice (2005) 2(3), 95-101. Available at:https://ro.uow.edu.au/jutlp/vol2/iss3/10

9. Ochoa L., Cueva Lobelle, A.: El plagio y su relación con los procesos de escritura académica. Forma y Función (2014) 27(2), 95-113. DOI: https://doi.org/10.15446/fyf.v27n2.47667

10. Meza, P., da Cunha, I.: Comunicación del conocimiento propio y relaciones discursivas en el género tesis. Sintagma (2018) 31, 103-120. Available at: http://www.sintagma.udl.cat/export/sites/Sintagma/documents/articles_31/Sintagma-31_7.pdf

11. Meza, P., Rivera, B.: The communication of author's own knowledge within thesis: variation among academic degrees in the theoretical framework section. Revista de Lingüística Teórica y Aplicada (2018) 56(1), 115-138. DOI: http://dx.doi.org/10.4067/S0718-48832018000100115

12. Dias, W., Eisenberg, Z.: Vozes diluídas no plágio: a (des)construção autoral entre alunos de licenciaturas. Pro-Posições, Campinas (2015) 26(1), 179-197. DOI: https://doi.org/10.1590/0103-7307201507602

13. Tran, U.T., Huynh, T., Nguyen, H.T.T.: Academic Integrity in Higher Education: The Case of Plagiarism of Graduation Reports by Undergraduate Seniors in Vietnam. J Acad Ethics (2018) 16, 61–69. Available in https://link.springer.com/article/10.1007/s10805-017-9279-9

14. Balbay, S., Kilis, S.: Perceived Effectiveness of Turnitin® in Detecting Plagiarism in Presentation Slides. Contemporary Educational Technology (2019) 10, 25-36. DOI: https://doi.org/10.30935/cet.512522

15. Bertram Gallant, T., Picciotto, M., Bozinovic, G., Tour, E.: Plagiarism or not? investigation of Turnitin®-detected similarity hits in biology laboratory reports. Biochem Mol Biol Educ (2019) 47, 370-379. DOI: https://doi.org/10.1002/bmb.21236

16. Villar-Mayuntupa, G.: Using Turnitin to boost the originality in the elaboration of argumentative essays among Peruvian university students. *In: 2020 IEEE World Conference on Engineering Education (EDUNINE)*, Bogota (2020), 1-4. DOI: 10.1109/EDUNINE48860.2020.9149492

17. Kaktiņš, L.: Does Turnitin support the development of international students' academic integrity? Ethics and Education (2019) 14(4), 430-448. DOI: https://doi.org/10.1080/17449642.2019.1660946

18. Mphahlele, A., McKenna, S.: The use of turnitin in the higher education sector: Decoding the myth. Assessment & Evaluation in Higher Education (2019) 44(7), 1079-1089. DOI: https://doi.org/10.1080/02602938.2019.1573971

19. Yu, S.: Giving genre-based peer feedback in academic writing: sources of knowledge and skills, difficulties and challenges. Assessment & Evaluation in Higher Education (2020), 1-8. DOI: https://doi.org/10.1080/02602938.2020.1742872

20. Marsden, H., Carroll, M., Neill, J.: Who cheats at university? A self-report study of dishonest academic behaviours in a sample of Australian university students. Australian Journal of Psychology (2005) 57(1), 1-10. DOI: 10.1080/00049530412331283426

21. Duff, A., Rogers, D., Harris, M.: International engineering students—avoiding plagiarism through understanding the Western academic context of scholarship. European Journal of Engineering Education (2006) 31(6), 673-681. DOI: https://doi.org/10.1080/03043790600911753

22. Gilmore, J., Strickland, D., Timmerman, B., Maher, M., Feldon, D.: Weeds in the flower garden: An exploration of plagiarism in graduate students' research proposals and its connection to enculturation, ESL, and contextual factors. International Journal for Educational Integrity (2010) 6(1). Available at: https://scholarcommons.sc.edu/biol_facpub/2/

23. Paltridge, B., Starfield, S.: *Thesis and dissertation writing in a second language: A handbook for supervisors*. Routledge England (2007).

24. Bautista, F., Sánchez Escobedo, P. A., Canto Herrera, P. J.: Plagio en los posgrados de ciencias sociales en una universidad estatal de México. Revista Educación y Ciencia (2017) 6(47), 82- 97. Available at: http://www.educacionyciencia.org/index.php/educacionyciencia/article/view/406/pdf_43

25. Saldaña, J., Quezada, C., Peña, A.: Alta frecuencia de plagio en tesis de medicina de una universidad pública peruana. Revista Peruana de Medicina experimental y salud pública (2010) 27(1). Availabre at: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1726-46342010000100011&lng=es&tlng=es

26. Head, A. J., Eisenberg M. B.: How todaýs college students use Wikipedia for courserelated-research. First Monday (2010) 15(3). DOI: https://doi.org/10.5210/fm.v15i3.2830

27. Selwyn, N., Gorard, S.: Studentś use of Wikipedia as an academic resource – Patterns ofuse and perceptions of usefulness. Internet and Higher Education (2016) 28, 28-34. DOI: 10.1016/j.iheduc.2015.08.004

28. Soler J., Pavlovic D., Freixa P.: Wikipedia en la Universidad: Cambios en la percepciónde valor con la creación de contenidos. Comunicar (2018) 54, 39-48. DOI: https://doi.org/10.3916/C54-2018-04

29. Rüger, S.: How to write a good PhD thesis and survive the viva (2011). Available at: http://people.kmi.open.ac.uk/stefan/thesis-writing.pdf

30. Phillips, E., Pugh, D.: How to get a PhD. A handbook for students and their supervisor. Open University Press Maidenhead, England (2005).

31. Arias-Chávez, D., Ramos-Quispe, T., Villalba-Condori, K. y Postigo-Zumarán, J.: The Characteristics of Academic Plagiarism in Four Universities in the City of Arequipa: A Comparative Study Conducted on Male and Female Students. International Journal of Innovation, Creativity and Change (2019) 11(10), 358-373.