

Traces of plagiarism in academic essays in students of an online doctorate

DOI: 10.46932/sfjdv3n4-085

Received in: April 14th, 2022

Accepted in: June 30th, 2022

José Luis Soto Ortiz

Doctor

Institution: Universidad Pedagógica Veracruzana

Address: Av. Justino Sarmiento, Jose Cardel, 91130, Xalapa-Enríquez -Veracruz, México

E-mail: joseluis.sotoortiz@gmail.com

Juan Carlos Zamudio Hermida

Master

Institution: Instituto Universitario Veracruzano (IUV)

Address: Km 148+090, Carr. Xalapa Perote, Centro, 91300, Banderilla -Veracruz, México

E-mail: juancarlos.zamudio@iuv.edu.mx

Carlos Arturo Torres Gastelú

Doctor

Institution: Universidad Veracruzana

Address: Lomas del Estadio, 91090, Xalapa - Veracruz, México

E-mail: ctorres@uv.mx

ABSTRACT

In this work, an exploratory study was carried out to analyze, through an anti-plagiarism program, the essays prepared by a group of postgraduate students. The focus of the study is quantitative, taking as a starting point the similarity index obtained with the iThenticate application, contrasting the essays in an artisanal way for an assessment of plagiarism. In a first analysis, a similarity index of textual copying of up to 90% was detected in some works, giving feedback to each student to correct their texts. Finally, the similarity index in the final delivery of the essays was on average 20.3%. In conclusion, using iThenticate allowed a rapid analysis of the essays, and that when collating manually, the plagiarism committed was evidenced; however, providing feedback to the student about the mistakes made was conducive to improving their writing and citation.

Keywords: academic writing, plagiarism, learning, education, digital tools.

1 INTRODUCCIÓN

The Veracruzano University Institute (IUV) teaches higher level and postgraduate studies under the mixed and virtual modality, among which are the online Doctorate in Administrative Sciences, promoting the professionalization and research of its students in their study plans. As part of its institutional values, there are ethical aspects, among which the originality of the documents made in order to avoid the conduct of academic plagiarism stands out.

In this regard, various studies indicate that academic plagiarism has been increasing due to the availability of the Internet as a means of accessing information, which has caused its frequency at all

educational levels (Hernández, Pozos, Ortiz and Vergara, 2015) other authors point out that the copy and paste effect indicates a reading deficiency in students (Cortez et al., 2021; Pérez-Calero et al., 2021). Likewise, some researchers emphasize that the bad practice of plagiarism affects the academic performance of the student (Greenberger, Holbeck, Steele and Dyer, 2016).

This bad practice is also observed in the different types of learning environments (face-to-face, mixed or virtual) where students, in an effort to fulfill their obligations, deliver their documents in a dishonest manner (Hernández, et al, 2015). For the specific case of this work, the optics consists of mentioning in the virtual context, in this way the deliveries of the works are made in digital format, and whose review of citations or references is not carried out, either due to lack of knowledge of the teacher in the management of computer programs in the detection of plagiarism or simply because of the time it takes between analyzing the tasks and giving feedback to the students.

In higher education, plagiarism is also latent, when presumably, at this level, students have already achieved the informational skills necessary for argumentation in the development of their work. Research in this regard indicates that the main indicators are the lack of training in the treatment and selection of information (Rebollo-Quintela, et al., 2017) as well as the misunderstanding of university regulations on plagiarism (Cebían-Lopez et al. ., 2018) being indistinct from the type of career, be it technical, humanities, administrative or health (Campus and Solano, 2012).

For their part, Comas et al. (2011) found in a sample of 1,025 undergraduate students assigned to the University of the Balearic Islands that a good part of the students did not directly cite the sources of consultation, or else, they copied text fragments without citing. The reason for this, according to the results presented, was that approximately more than 50% did not manage time well, leaving the completion of tasks to the last moment.

Based on the above, this paper analyzes the signs of plagiarism detected in the essays carried out by a group of students of the General Theory of Administration subject belonging to the period 2018 under the virtual modality.

2 METODOLOGY

2.1 DIGITAL TOOLS FOR PLAGIARISM DETECTION

With the growth of information and communication technologies, a variety of digital tools have been developed that allow the detection of plagiarism for the evaluation of the academic honesty of the documents, there are anti-plagiarism computer programs, which can be of the type commercial or free. Commercial applications include: turnitin, urkund, grammarly. iThenticate, among others. As for the free ones, they are: Duplichecker, Small Seo Tools, Plagscan, to name a few.

The operation of anti-plagiarism programs consists of loading the document to be analyzed or simply placing the text in the space designated for it, as the case may be. Subsequently, the program performs a comparison searching for equivalent texts in its database or on Internet sites in order to find similarities of the matching paragraphs, and then presents the result of the detected percentage, called degree of similarity. The higher the percentage presented, the greater the degree of similarity, and to facilitate the location the text appears highlighted in those similar statements, likewise the anti-plagiarism applications, show a list with the electronic addresses referring to the possible original sources from where it was information extracted. However, these applications have certain disadvantages, among which are: literary translations from English to Spanish, modified wording with synonyms, as well as the generation of false positives when the source was cited correctly.

2.2 GOAL

To analyze the essays prepared by a group of virtual doctorate students with an anti-plagiarism computer application.

2.3 TYPE OF STUDY

The methodology of this research was quantitative. To do this, 23 essays were validated using the "iThenticate" anti-plagiarism tool, which allowed contrasting the results of those sentences copied verbatim against the original sources detected in the similarity index report; the results were processed in the statistical program SPSS version 20 in order to obtain the type of plagiarism committed.

2.4 POPULATION

The sample criterion was defined as non-probabilistic and intentional sampling as it was considered pertinent and accessibility for reviewing the documentary works being relevant factors for the study (Hernández, Fernández and Baptista, 2010). This sample consisted of 23 students.

2.5 INSTRUMENTS

For the evaluation of the academic essays, the iThenticate tool of the company Turnitin was used, which analyzes the texts, resulting in a report with the similarity index found and the electronic address of each of the possible sources of information. After that, the essays were compared with the sources of information indicated in the iThenticate report to corroborate plagiarism.

To collect the quantitative assessments, an adaptation was made to the Balbuena and Lamela (2015) instrument, adapting the questionnaire to ten items with a five-level Likert scale. Within the variables of the survey, objective questions were included about the concept of plagiarism, attitudes

towards plagiarism and about the correct citation.

2.6 EDUCATIONAL ACTIVITY

The scenario in which this research was developed includes the General Theory of Administration course corresponding to the Doctorate in Administrative Sciences under the online modality. The final objective of the subject is to provide knowledge regarding the general theoretical aspects of administration that encompass the different approaches to administrative thought and changes at the epistemological level of administrative sciences. For this, the thematic content of the course consists of eight topics distributed in three units, with individual and collaborative instructions, through communication in virtual forums.

For the final evaluation of the course, the participants had to prepare an academic essay taking as reference the doctoral protocol or elements of their research (theoretical framework or state of the art). In this sense, from the beginning of the course, the students were given a recommendation about the elements that would contain their final work in accordance with the following points:

- The topic will be the research problem.
- Bibliographic citations based on APA standards.
- The argument must consider the presentation, content, body of the document, conclusions and references that coincide with the citations.
- Maximum 15 pages.

The textual analysis of the content of the essay was carried out in two moments:

First moment:

The activity of the course focused on the interaction in the discussion forums, within which the participants made contributions regarding their theoretical framework, exposing the main authors, as well as the sources of information consulted. In this regard, they were told that to cite these sources they should use the APA format in its sixth edition. During the academic development of the third unit, the student delivered a first version of the essay, each of these works were entered into the iThenticate platform for the validation of the similarity index (IS), to later send the results to the students in the incidents committed with respect to the writing and citation. Each feedback work carried with it the IS indicating those paragraphs or segments copied verbatim, as well as the electronic address of the original source of information.

Second moment:

In this part, the students had to solve the observations made previously for the delivery of the final work. To do this, each work was resubmitted to the iThenticate tool to validate the authenticity of the

document using the similarity index. Also, this stage, an evaluation rubric was applied with the following points:

1. Concludes the drafting of the referential framework, explaining what and how the TGA contributes to the identified problem (references and perspectives), in a well-founded manner.
2. The final document considers the research problem as its theme; arguments (presentation, content, body of the document and conclusions) and sources consulted.
3. Integrate the activity based on what was requested and with a maximum of 15 pages. References are noted with APA attachment, excellent writing, and no spelling errors.

It should be noted that both phases, the works were classified according to the type of plagiarism, considering the following aspects:

- Total and/or partial copying without modification as it appears textually in the source(s) of information referred to in the Similarity Index report of the iThenticate tool.
- The incorrect use of paraphrases, where the student changed the words to synonyms, or made the arrangement of sentences presenting them as their own authorship without giving credit to the original source.
- Lack of the use of citation and referencing to differentiate between the text of own authorship and the one copied from other authors.
- Correct citation, where the student cites either directly or indirectly the source of information.

The results of the Similarity Index of the iThenticate tool, and of the survey were processed in the SPSS version 20 tool. From the comparative textual analysis in the writing of the texts by the students, plagiarism was classified.

3 RESULTS

3.1 PLAGIARISM SIMILARITY INDEX DETECTED

Based on the results, the gender of the students is 56.3% female and 43.5% male, who have their professional training in different areas, predominantly Medicine (39.13%), Administration (17.39%) and Nursing (13.04%). which shows that the essays come from different areas of knowledge. In this sense, 23 essays whose length ranged between 1,442 and 4,984 words, with an average of 3,774.74 words, were analyzed.

Regarding the similarity index in the first delivery of essays, the results found with iThenticate show that plagiarism fluctuated between 11% and 90%, this last value represents a work copied almost entirely from the Internet. In this regard, a total of 23,408 plagiarized words were detected in the analyzed essays, see Table 1.

Table 1. General results of the tests. Own elaboration.

<i>Analysis of trials n=23</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Media</i>	<i>Standard deviation</i>
<i>essay words</i>	1442	4984	3378	1022
<i>Percentage of plagiarism detected with iThenticate</i>	11.00%	90.00%	49.521%	25.38229%
<i>Words with possible plagiarism</i>	22	4045	1018	1139

When manually comparing the corpus of essays with the sources of information obtained in the iThenticate reports, the types of plagiarism were validated, cataloging as follows:

Table 2. Results of the types of plagiarism detected in the first review of the essays. Own elaboration.

<i>Type of plagiarism</i>	<i>Words count</i>	<i>Average type of plagiarism</i>	<i>Standard deviation</i>
<i>Full verbatim copy, n=16</i>	19656	854.61	1071.33
<i>Incorrect use of paraphrase, n=11</i>	1916	83.30	112.60
<i>Lack of citation, n=11</i>	547	23.78	33.43
<i>Correct citation, n=5</i>	678	29.48	78.57
<i>False positive, n=18</i>	611	26.57	19.22

Regarding the second installment of the essays, when they were submitted for a review with the iThenticate tool, Table 3 presents the comparison of the results between the first and second analysis, the similarity index went from 90% to 37% and from 4045 to 193 words detected as possible plagiarism. In this sense, when carrying out the manual comparison, it was detected that in the second analysis false positives and correct citations were found in the resubmitted papers, so they were not considered as plagiarism. In such a way that, the feedback given to the students by the teacher, allowed them to solve the textual aspects indicated as plagiarism, likewise, the writing of the essay regarding the citation style was corrected.

Table 3. Comparison between the first and second analysis. Own elaboration.

<i>Analysis of trials n=23</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Media</i>	<i>Standard deviation</i>
<i>First analysis</i>	11.00%	90.00%	49.52%	25.38%
<i>Words detected as plagiarism 1st. analysis</i>	22	4045	1018	1139
<i>Second analysis</i>	9.0%	37%	20.30%	6.75%
<i>Palabras detectadas como plagio 2do. análisis</i>	2	595	193	188

4 CONCLUSIONS

Digital tools for plagiarism analysis facilitate validation of the authenticity of academic texts (Díaz, 2017). From this, the similarity index was obtained, representing the degree of possible plagiarism detected in the reviewed essays. It should be noted that the Similarity Index prepared by iThenticate does

not show the type of plagiarism, so the comparison was made manually from said reports, this allowed presenting a more accurate report in the way of objective feedback, but above all that the student is aware of the mistakes made in their written work.

Among the faults detected were the complete copying of the texts, the incorrect paraphrasing and the lack of citation. In this sense, the feedback given in these aspects allowed the students to improve the writing of their texts. Another point to highlight was that, when contrasting with the survey responses, differences were found in the sense that the students know the concept of plagiarism well, but at the time of the first installment the similarity rates were high, which confirms that plagiarism was committed; In the case of incorrect paraphrasing, the indications found were due to the change of words in paragraphs and that they follow the logic of the author of the original source and not of the student himself.

So, currently there are endless sources of information, they are susceptible to being copied, therefore, it is important for the teacher to review the tasks so that the guidance they receive allows the student not only to identify the reliable sources of information, but also cite the sources used according to the standards (Egaña, 2012).

It is necessary to specify that some works were detected that contained correct citations causing false positives, this shows that the technological tools do not carry out an exhaustive analysis. The foregoing was confirmed when comparing the reports issued by iThenticate with the texts of the original information sources, in which some words are indicated as plagiarism and the opposite is confirmed in the review. This is because some words that are concepts have definitions that cannot be changed for synonyms and that correspond to universal knowledge, which does not correspond to plagiarism.

RECOMMENDATIONS

As a recommendation, there is still a lot of work to be done in terms of detecting plagiarism in learning environments, especially in university degrees where the student must have the necessary research and argumentative knowledge in the preparation of academic writings. Similarly, teachers must review the content of the writings in order to identify when students plagiarize giving feedback on the matter, whether preventive, corrective and, where appropriate, subject to any sanction according to the regulations of each institution.

REFERENCES

- Balbuena, E. y Lamela R. (2015). Prevalence, Motives and Views of Academic Dishonesty in Higher Education. *Asia-Pacific Journal of Multidisciplinary Research*, vol. 3, no. 2. Recuperado de: <https://eric.ed.gov/?id=ED575015>.
- Cebrían-Robles, V.; Raposo-Rivas, M.; Cevbrián de la Serna, M. y Sarmiento-Campos (2018). Percepciones sobre el plagio académico de estudiantes universitarios españoles. Vol. 21, Núm. 2. Recuperado de: <http://revistas.uned.es/index.php/educacionXX1/article/view/20062>
- Comas-Forgas, R. y Sureda-Negre, J. (2016). Prevalencia y capacidad de reconocimiento del plagio académico entre el alumnado del área de economía. *El profesionalismo de la información*, vol. 25, no. 4.
- Cortés, K. R., Cruz, J. J. R., & Anaya, C. R.. (2021). De la Pedagogía conceptual y los niveles de lectura de un grupo de la LEIyP, UPN-México. *South Florida Journal of Development*, 2(1), 566–576. <https://doi.org/10.46932/sfjdv2n1-040>
- Díaz, D. (2017). Evaluación del desempeño de tres herramientas antiplagio gratuitas en la detección de diferentes formas de copy-paste procedentes de Internet. *EDUTEC, Revista Electrónica de Tecnología Educativa*, Núm. 59, marzo 2017.
- Egaña, T. (2012). Uso de bibliografía y plagio académico entre los estudiantes universitarios. *Revista de Universidad y Sociedad del Conocimiento*, vol. 9, núm 2, 2012. Barcelona España.
- Hernández, M.; Ponce, A.; Ortiz, V. y Vergara, M. (2015). Estrategias docentes emergentes para prevenir el plagio académico en alumnos de pregrado. Un estudio fenomenográfico con profesores experimentados. *Revista Consensus*, vol. 20, núm. 2.
- Greenberger, S.; Holbeck, R.; Steele, J y Dyer, T. (2016). Plagiarism Due to Misunderstanding: Online Instructor Perceptions. *Journal of the Scholarship of Teaching and Learning*, Vol. 16, No. 6. Recuperado de: <https://scholarworks.iu.edu/journals/index.php/josotl/issue/view/1506>.
- Pérez-Calero, L., Aguilar, R., & Larrañeta, B. (2021). Composition of working groups: How does the functional and social diversity affect their performance?. *South Florida Journal of Development*, 2(2), 2408–2414. <https://doi.org/10.46932/sfjdv2n2-100>
- Rebollo-Quintela, N; Espiñera-Bellón, E. y Muñoz-Canterio, J. (2017). Atribuciones causales en el plagio académico por parte de los estudiantes universitarios. *Revista de Estudios e Investigación en Psicología y Educación*. Vol. Extr. No. 06. Recuperado de: <http://revistas.udc.es/index.php/reipe/article/view/reipe.2017.0.06.2453>