Plagiarism, Internet and Academic Success at the University

Juan Carlos Torres-Díaz1*, Josep M. Duart2, Mónica Hinojosa-Becerra3
1Departamento de Ciencias de la Computación, Universidad Técnica Particular de Loja {jctorres@utpl.edu.ec}
2ELearn Center, Universitat Oberta de Catalunya, España {jduart@uoc.edu}
3Departamento de Comunicación, Universidad Nacional de Loja, Ecuador {monica.hinojosa@unl.edu.ec}

Received on 23 May 2018; revised on 30 May 2018; accepted on 5 June 2018; published on 15 July 2018
DOI: 10.7821/naer.2018.7.324

ABSTRACT
In this work, we determined, the level of incidence of the use of technologies on academic success and the incidence of interaction and experience on the level of plagiarism of university students. A sample of 10,952 students from 31 face-to-face universities in Ecuador was created. Students were classified based on their experience level, level of interaction with teachers and classmates, and the way they use technology for academic activities. The results showed that the level of experience does not affect academic success, but does have an incidence on plagiarism levels that increase as this experience increases. Plagiarism reaches higher levels when level of experience, family income and hours of connection per day increases. Academic performance depends on the number of hours that students seek information and the number of academic videos they watch. Also, plagiarism tends to decrease as the student makes better use of technology for their academic activities.

KEYWORDS: PLAGIARISM, ACADEMIC ACHIEVEMENT, INTERNET.

1 INTRODUCTION
1.1 Determinants of the digital gap
Education is an area that does not escape to dynamic of information society and today is one of the areas that experience more innovation. The dilemma of having or not having a connection that defined the digital gap in its beginnings has changed and focuses on how technology is used to take better advantage of it (Hargittai, 2002). In 2001, the skills and abilities in the use of technology were considered as one of the digital gap levels to be considered (DiMaggio & Hargittai, 2001). Today differences are fundamental and often have nothing to do with the lack of knowledge but with the culture of using computer tools. The differences that may exist between users regarding their use of the Internet requires a classification based on the most important and common activities using technology (Scheerder, Van Deursen, & Van Dijk, 2017).

Regarding the incidence of socio-economic variables on the use given to technology, the contributions are available (Di-Maggio, Hargittai, Celeste, & Shafer, 2004; Van Dijk, 2005) coinciding with the theory of gaps in knowledge (Tichenor, Donohue, & Olien, 1970) that the socioeconomic status that best assimilate information are the highest. Recent studies, point out to the prevalence of socioeconomic variables; in the case of Ecuador, Tirado-Morueta, Mendoza-Zambrano, Aguaded-Gómez and Marin-Gutiérrez (2017) conclude that the effect of belonging to a family of high socio-economic level loses strength as the level of Internet access increases. Scheerder, van Deursen and van Dijk (2017) point out that sociodemographic variables have the greatest impact on digital uses and skills. However, this relationship tends to decrease and at present, despite being significant, it has a low incidence (Jara et al., 2015; Torres-Díaz, Duart, Gómez-Alvarado, Marin-Gutiérrez, & Segarra-Faggioni, 2016; Wainer, Vieira, & Melguizo, 2015). This make us to rethink the research scenario, since the explanation for differences in use is determined by a different set of variables.

1.2 Technology and academic performance
There is evidence suggesting that the time spent studying online does not necessarily imply better results (Castano-Muñoz, Duart, & Sancho-Vinuesa, 2014). Positive effects can be found in the levels of learning and in the results obtained by students (Lee & Lec, 2012; López-Pérez, Pérez-López, Rodríguez-Ariza, & Argente-Linares, 2013; Markovic & Jovanovic, 2012; Mohd & Maat, 2013). There are also studies where no incidence of technology has been found on learning outcomes (Wittwer & Senkeil, 2008) or negative effects when problems of addiction are present that constitute a social problem (Chou, Condron, & Belland, 2005; Kim, LaRose, & Peng, 2009), having as a consequence the decrease of the student’s academic performance (Junco & Cotten, 2011; Kubey, Lavin, & Barrows, 2001). The inclusion of technology in education showed different effects on the academic performance of students, similar studies present contradictory results (Torres-Díaz et al., 2016).

If the measure of the effect of technology is academic performance, it is necessary to consider that it is a multidimensional concept (Fullana, 1992) whose evaluation requires an equally multidimensional approach. In this paper, the academic success associated with approving or not approving the subjects or credits that the student has taken is evaluated. Recent work in this area indicates that having a computer and Internet connection
has less and less impact on a better academic performance, the even greater effect, although decreasing, occurs in households with higher socioeconomic status (Wainer et al., 2015). The use of technology in the classroom requires an approach and an adequate design that makes the most of both resources and learning opportunities. The opposite supposes to have effects not necessarily expected. Technology can, in addition to contributing to improve learning, cause some problems related to the distraction that may involve not assimilating content and a consequent low qualification (Ravizza, Hambrick, & Fenn, 2014).

1.3 Plagiarism

The writing of academic documents is based on the current state of knowledge through the incorporation of ideas from different authors (Pecorari & Petric, 2014). This is a process that is governed by the habit that an author has to comply, with disciplinary established practices and shared to get away from accusations of plagiarism (Pecorari, 2008). In recent decades, the arrival of the Internet has led to a growing wealth of readily available and easy to plagiarize sources (Hu & Lei, 2012). The incidence of plagiarism is increasing and has generated great concerns in the academic world.

There are suggestions about how cultures differ in their understanding and acceptance of plagiarism (Sowden, 2005). Other researchers (Pecorari, 2008; Rinnert & Kobayashi, 2005) have the idea that difficulties of a second language authors, are probably the lack of an adequate command of the language, which can cause insecurity about their use of language and therefore make depend too much on the original texts.

In the Anglo-American academic world, it is considered that source documentation and paraphrasing are two important strategies to avoid plagiarism (Park, 2003). While the first one is reasonably simple and can be performed with adequate training, the second one implies high demand in the knowledge and linguistic competence (Keck, 2010). Researchers and academic leaders differ greatly in the standards on what is considered a paraphrase. While some ones believe that to get away from plagiarism, they should not have be any trace of the textual copy, even of strings of a few words since the original work (Benos et al., 2005; Roig, 2001), others adopt softer norms allowing the inclusion of more sources in a paraphrase (Keck, 2006; Pecorari, 2008).

To solve the problem of plagiarism, it is necessary to look beyond the symptoms and reach the root of it (Macdonald & Carroll, 2006). All the factors that contribute to plagiarism are reduced to a certain deficiency in plagiarists who lack the academic integrity, good-will and knowledge necessary to use the sources adequately (Pecorari, 2008). Teachers have a role to play detecting and responding to plagiarism and educating to avoid this happens. Researchers have paid increasing attention to teachers’ perceptions of plagiarism. Previous studies found that teachers differ among themselves in what constitutes plagiarism (Borg, 2009; Flint, Clegg, & Macdonald, 2006; Pickard, 2006), and that many had little knowledge about the institutional definitions of plagiarism and, that this was not effectively taught (Eriksson & Sullivan, 2008).

1.4 Higher education and plagiarism

Plagiarism has become a widespread problem at all levels, and it is easy to find cases of plagiarism at higher educational levels in the media. For example, in recent years we have witnessed the resignation of two German ministers accused of plagiarism in their doctoral thesis (Eddy, 2013). The tools to search and manage information and the information available on the Internet have opened the doors to a phenomenon that is not new, plagiarism has increased in the educational field and at the same time tools for control it have emerged. Despite this, it is impossible to prevent students from having the possibility of using information that does not belong to them by assigning their name. In the educational activity the students work with different activities, exercises, tests or solutions to problems and all the information is just a few clicks away (Atkins & Nelson, 2001; DeVoss & Rosatti, 2002; Moore Howard, 2007). The Internet has been divided into all aspects of our lives, including university classrooms, and has opened up new ways of finding solutions for class assignments. Many students seek the quickest solution to tasks, regardless of the validity of the sources or without respecting the work of others (Sureda, Comas and Oliver, 2015).

We are focused on the higher education field, where the increase in plagiarism has been a constant concern in recent years (Heckler & Forde, 2015), with teachers increasingly concerned about the frequency and apparent lack of awareness, into the students, of the moral implications (Perry, 2010). The results of a survey conducted by the companies Six Degrés and Le Sphinx Développement (2008) showed some significant behaviors of students and teachers, identifying the Internet as the main source of documentation and more than 40% of students admitted that they had never cited their sources. These results are consistent with another experiment carried out among 1,025 students (Comas, Sureda, & Oliver, 2011), which states that 70% of university students admitted to copying texts or fragments of texts for the development of their academic activities, presenting them as their own. It is interesting to observe that the greater the weight of the activity in the final grade, the lower the percentage of plagiarism, which suggests an interesting link between the perceived importance of the task and the tendency to plagiarize (Gómez, Salazar, & Vargas, 2013).

About the factors that conduct students to plagiarize, even before the Internet era, Ashworth, Bannister and Thorne (1997) identified four fundamental issues: 1) the students’ lack of awareness of whether they are plagiarizing or not; 2) the low probability of being detected; 3) the pressure derived from the level of demand and the deadlines established for deliveries; and 4) the actual writing of the activities provided by the teachers. These factors are still relevant: a more recent study by Eret and Ok (2014) observed that the tendency to plagiarize was increased with the arrival of the Internet, and points out as main reasons for plagiarism the limitations of time, excessive workloads and a high difficulty of the proposed tasks. Another recent study by Hussein, Rusdi and Mohamad (2016) found that students were fully aware of what plagiarism was and that it is not appropriate to use it. A study by Kaufman and Young (2015) analyzed how easy the access to copy and paste tools are, and the presentation of tasks, influenced attitudes towards plagiarism.

Most studies agree that: access to information has become so immediate that it is perceived by some as a “common knowledge” available for everyone to reproduce (Walker, 2010). These generalized problems have led to an academic / technological response in search of new ways and tools to detect plagiarism.

2 DATA AND METHODS

2.1 Data

This study includes a sample of 10,952 face-to-face universities students, from 31 universities in Ecuador, which 51.7% correspond to women and the remaining 48.3% correspond to men.
A physical survey was applied in each institution, collecting a significant sample in each case. The calculation of the sample was obtained using a confidence level of 95% (Z=1.96) and a sample error of 5%.

The instrument used was based on the questionnaire of The Internet Catalonia project (UOC, 2003) and on questionnaires of the Digital Literacy in Higher Education project (DLINHE, 2011). The questionnaire collects general information about the preferences of students regarding their connection patterns; it also collects information about the use of technology from academic activities and collects information about academic success.

2.2 Methods

The classifications were developed applying the non-hierarchical K-media algorithm, useful when working with large samples (Díaz De Rada, 2002). Classifications were built based on the use of technological tools into the learning process, and based on the students’ willingness to use technology. Different options were tested with several groups and the easier to interpret classifications was chosen in all cases. The classifications based on experience levels was made using the following variables:

- Internet connection hours per day.
- Years as an Internet user.

The classification based on the interaction activities used the following variables:

- Number of queries to the teacher.
- Number of queries to colleagues.
- Chat hours on academic topics.
- Post number in social networks about academic topics.

The classification based on the academic activities used the following variables:

- Number of academic videos.
- Hours of research of academic information.

The academic success variable (number of failed subjects) was constructed using the following variables: number of subjects enrolled and number of subjects passed. The difference between the two was extracted and the number of failed subjects was obtained. To measure the level of plagiarism of the students, an ordinal variable was used in which the student points out a value between 1 and 10, which indicates the frequency with which he copies his academic tasks from the Internet. For the relations search, we used the Chi-square statistic, Pearson’s R, Linear and Tobit regression, in which the academic success and level of plagiarism were the dependent variable.

3 RESULTS

3.1 Classification of students

Applying cluster analysis, three classifications were created based on: year of experience, level of interaction and based on the use of technology for academic activities.

3.1.1 Classification based on years of experience

Cluster analysis was performed on the variables indicated in Figure 1, this categorization is based on years of experience and the students were divided into two groups.

![Figure 1. Classification based on experience levels](image1)

The lower experience group shows the lowest values in the two variables. That is, it connects fewer hours per day and fewer years as Internet users in relation to the remaining students; this group represents 68% of the sample. The higher experience group is the one with the highest values in the variables levels, representing 32% of the students. In this classification, the greatest differentiation between the groups is given by the number of hours the student spends searching for academic information.

3.1.2 Classification based on interaction

To build the groups based on the interaction activities, it should be noted that the variables used actually measure academic activities, however these are more oriented to the exchange of information and messages with other people. This classification refers to the interaction activities of the student and considers the variables that can be observed in Figure 2.

![Figure 2. Classification based on interaction](image2)

The resulting classification divides the students into two groups: in group 1, which is called low interaction, the students’ activity in terms of writing queries to professors and classmates, writing on social networks and chatting on academic subjects is low with respect to students in group 2. This group is composed with 80% of the students in the sample. In the second group, students’ activity is greater and the queries to classmates and chat hours on academic topics stand out. This group is composed with 20% of the students. The biggest difference is given by two variables: the number of hours the student chats about academic topics and the number of questions he asks his classmates, which is greater than the number of questions he asks his teachers.
3.1.3 Classification based on academic activities

A third classification analyses the student’s work with the traditional activities carried out by a student when studying using technologies. These activities are considered in the variables in figure 3.

In general, in this classification, the differences are given by the number of academic videos that are watched, and in a major part, by the number of hours spent searching for information. The academic work group with minimum values in the two variables represents 89.3% of the students; the medium academic work group represents 9.9%; and the higher academic work group represents only 0.8%.

The classifications: interaction and academic activities, are complementary. In both, variables that measure the use of technology in academic activities are used. This complementarity is reflected in a result that indicates this: students who have high interaction, are those who use technology in a better way for their academic work.

Table 1. Relationship between interaction and use of technology for academic work

<table>
<thead>
<tr>
<th></th>
<th>High academic work</th>
<th>Medium academic work</th>
<th>Low academic work</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Interaction</td>
<td>43</td>
<td>532</td>
<td>8802</td>
<td>8777</td>
</tr>
<tr>
<td>High Interaction</td>
<td>45</td>
<td>549</td>
<td>1581</td>
<td>2175</td>
</tr>
<tr>
<td>Total</td>
<td>88</td>
<td>1081</td>
<td>9783</td>
<td>10592</td>
</tr>
</tbody>
</table>

It can be observed in the previous chart: the high interaction is the majority in the high and medium academic work groups. And the opposite, it is a minority in the low academic work group. The relation is significant ($X^2 = 0.259$, $p<0.05$).

3.2 Relations

The levels of plagiarism are affected by different variables, the incidence is low but significant. The level of knowledge of the internet affects the level of plagiarism, this implies that the higher the level of knowledge, the higher the level of plagiarism; another variable that has a significant impact is the number of hours the student connects per day, as the number of hours increases, the level of student plagiarism also increases; the same behavior has the variable level of income, the higher the level of income, the higher the level of student plagiarism.

The use that students give to technology affects the level of plagiarism. Students who use technology in a better way for academic activities tend to plagiarize less; on the other hand, students who download less educational resources, watch fewer academic videos or invest less time searching for information, tend to have higher levels of plagiarism.

Table 2. Linear regression model Academic_activities and Plagiarism

<table>
<thead>
<tr>
<th>Model</th>
<th>B</th>
<th>Std. Error</th>
<th>Beta</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cons</td>
<td>2.655</td>
<td>0.24</td>
<td>11.057</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Academic activities</td>
<td>0.492</td>
<td>0.083</td>
<td>0.057</td>
<td>5.953</td>
</tr>
</tbody>
</table>

Searching for cause-effect relationships we did the relation between the different categories with the academic success achieved by the students. First, the experience was related to academic success and no significant relationship was found ($X^2 = 0.010$, $p>0.05$). Something similar happened with the classification based on the levels of interaction and academic performance and with the relationship between the levels of academic work and academic success.

Significant relationships were found between the number of queries the student makes to the teacher and academic success. The relationship determines that the greater number of queries, the student tends to fail less. This same effect has the number of hours the student spent seeking information related to their academic activities. A different result is the one that indicates: the greater the participation in academic forums, the probability of failing increases.

The student’s gender does not have a relationship with the interaction categories ($X^2 = 1.5$, $p>0.05$). It also has no relationship with the categories of academic work ($X^2 = 1.55$, $p>0.05$). And in what has to do with the levels of knowledge of technology, we found that the tendency of women to belong to the middle and high categories is greater ($X^2 = 25.86$, $p <0.05$).

Age is not related to the interaction categories ($X^2 = 34.268$, $p>0.05$). There is also no significant relationship ($X^2 = 360.996$, $p >0.05$) with experience levels; the same applies to the categories of academic work ($X^2 = 58.731$, $p>0.05$).

The experience in the management of technology, classified in two levels, implies that the students show works, that are not their authorship. If a student change their level of experience handling technology to a higher one, the probability that the level of plagiarism rise one unit, is of the 10.4%
Table 3. Correlation with academic success

<table>
<thead>
<tr>
<th></th>
<th>How many inquiries do you make to your teachers each month?</th>
<th>How many virtual forums do you participate in each month?</th>
<th>How many hours do you look for academic information on the Internet, each month?</th>
<th>Failed Subjects (academic success)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many inquiries do you make to your teachers each month?</td>
<td>1</td>
<td>.154**</td>
<td>.084**</td>
<td>-.035**</td>
</tr>
<tr>
<td>How many virtual forums do you participate in each month?</td>
<td>.154**</td>
<td>1</td>
<td>0.002</td>
<td>.024*</td>
</tr>
<tr>
<td>How many hours do you look for academic information on the Internet, each month?</td>
<td>.084**</td>
<td>0.002</td>
<td>1</td>
<td>-.027**</td>
</tr>
<tr>
<td>Failed subjects (academic success)</td>
<td>-.035**</td>
<td>.024*</td>
<td>-0.027</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: * p<.05; ** p<.01

Table 4. Regression model, experience and plagiarism

<table>
<thead>
<tr>
<th>Plagiarism</th>
<th>Coef.</th>
<th>Std. Error</th>
<th>t</th>
<th>P&gt;t</th>
<th>[95% Conf.]</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience</td>
<td>0.1041582</td>
<td>0.0361338</td>
<td>2.88</td>
<td>0.004</td>
<td>0.0333295</td>
<td>0.174987</td>
</tr>
<tr>
<td>_cons</td>
<td>3990812</td>
<td>0.0845864</td>
<td>47.18</td>
<td>0</td>
<td>3825007</td>
<td>4156616</td>
</tr>
<tr>
<td>/sigma</td>
<td>3251112</td>
<td>0.0238202</td>
<td></td>
<td></td>
<td>320442</td>
<td>3297804</td>
</tr>
</tbody>
</table>

The use of technology for academic activities also makes a variation in the levels of plagiarism.

Table 5. Regression model uses of technology and plagiarism

<table>
<thead>
<tr>
<th>Plagiarism</th>
<th>Coef.</th>
<th>Std. Error</th>
<th>t</th>
<th>P&gt;t</th>
<th>[95% Conf.]</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic uses</td>
<td>-0.5544231</td>
<td>0.0908543</td>
<td>-6.1</td>
<td>0</td>
<td>-0.732514</td>
<td>-0.376323</td>
</tr>
<tr>
<td>_cons</td>
<td>4835513</td>
<td>0.106924</td>
<td>45.58</td>
<td>0</td>
<td>4627553</td>
<td>5043473</td>
</tr>
<tr>
<td>/sigma</td>
<td>3246986</td>
<td>0.0237882</td>
<td></td>
<td></td>
<td>3200357</td>
<td>3293615</td>
</tr>
</tbody>
</table>

In this case, there is an inverse relationship: if a student raises his level of use of technology for academic activities, the probability that the level of plagiarism is reduced by one unit, increases by 55%.

4 DISCUSSION AND CONCLUSIONS

The time of connection and the years as an Internet user, allow us to define a classification that shows the level of disposition that a student can have to use technology. In this classification there is a group called “lower experience” (68%) that is the one that presents lower levels of experience and connection time. This group, is the group with the lowest income, which would explain their situation, this is aligned with the postulates of the digital gap theory. There is one additional group called “higher experience” characterized by having more connection time and more years as an Internet user. These groups shows a different behaviour in terms of gender, which places women as a significant majority in the high level, unlike what was presented in previous research (Torres-Diaz et al., 2016) where women tended to have less experience and connection time. Something important we found, is the determination of incidence of experience levels using technology in the level of plagiarism of academic papers that after are presented as homework (investigation). As this experience increases, the level of plagiarism also increases.

The interaction levels divided the students into two groups: low interaction and high interaction. When this classification was related to academic success, no significant relationship was found. However, relating the number of queries done by the students to the teacher, and the academic success, the greater the number of queries the student fails less. This have a particularity due to the variable number of queries to the teacher is part of the variables that make up the interaction classification.

It is strange that we found this: participation in academic forums has a negative impact on academic success, however it is necessary to contextualize in this regard, stating that, if the forums are not part of the learning activities or are participations outside the educational institution, they can be considered distractors and the result is logical. In the case of the forums that are part of the academic activities, the theory and studies on the subject suggest that in the best of cases there is no incidence (Wittwer & Senkbeil, 2008). The levels of interaction are not influenced by gender, unlike similar studies where it is evident that gender discriminates activity levels. An aspect that was originally evidenced by the concepts of the digital gap and knowledge gap theory is the one that shows that the higher the level of income, the greater the level of interaction.

A third classification putted together the activities that can be considered appropriate for a training process. In this classification, the variables that have a higher level of discrimination are...
the number of academic videos that the student watches and the number of hours that he seeks information. Because both, the classification that measures the interaction and the one that measures the academic work, have complementary activities, a direct relationship was observed, students who have the most interaction are those who have the most academic work, this is a logical and expected result. The interaction in this work affects the way in which the student uses the technology but it is not related to age or income levels of the student. The use of technology in academic activities also has an impact on the level of plagiarism of students, the greater the use, the tendency to plagiarism decreases. This finding is not necessarily deterministic and could be only an expression other than the appropriate use of technology.

Level of knowledge of Internet, hours of connection and income level affects level of plagiarism, the higher the level of this variables, the higher the level of student plagiarism.

REFERENCES


