The Effectiveness of Turnitin Feedback Studio

1.0 Introduction

This analysis examines student groups' average similarity scores across the first five papers submitted to Turnitin Feedback Studio. The findings reveal that the levels of unoriginal content among students whose first submissions had high rates of similarity dropped significantly by their second paper. Students whose initial submissions did not have enough similarity (i.e., they did not include an adequate amount of quotations and/or references) seem to correct their practices by the second submission, as well. This study found that these effects are long-lasting, occur both in secondary and higher education institutions, and appear across the globe regardless of the country in which the students were studying.

2.0 General Methodology

This study follows the trends of average similarity scores among students within education institutions for the first five papers they submitted to Turnitin Feedback Studio since 2015.

The authors of this study examined average similarity scores at the student level (as opposed to the institution level) because this approach has a greater likelihood of producing more accurate results. When averaging similarity scores for an institution at any point in time, there are many variables that cannot be controlled or accounted for and could affect outcomes: the score data within institutions will come from students at different education levels, students with different levels of proficiency skills in proper citation and attribution, and students whose instructors have different degrees of familiarity with Turnitin Feedback Studio.

Therefore, it was determined that the most accurate way to demonstrate the effectiveness of Turnitin Feedback Studio was to take a specific group of students (2.6 million in total in higher education and 2.15 million in secondary), culled according to certain criteria, and to track how the average of their similarity scores changed across their first five submissions.

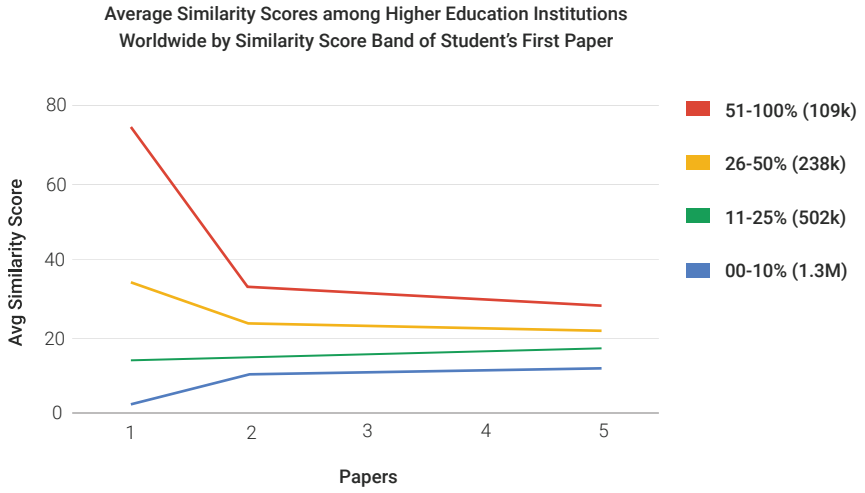
Students may enter any stage of their academic careers with levels of proper citation and attribution skills that differ greatly from those of their peers. Averaging all students' scores, regardless of their initial skill level, would prevent us from having a more accurate picture of their progress. Therefore, we segmented the data into bands based on the similarity score of each student's first submission (0-10%, 11-25%, 26-50%, and 51-100%).

All student data were anonymized and aggregated to preserve privacy. For more detail on the criteria and methodology employed by the study, please refer to this paper's Appendix.

3.0 Findings

3.1 Worldwide, the average similarity scores among higher education students with high initial similarity scores dropped significantly and immediately.

This study's sample of students in higher education from across the globe included 2.6 million students. The majority (just over 2 million) received a similarity score of 25% or less on the first paper they submitted. Nearly 500 thousand received scores that were above 25%. Looking at the chart below, it is clear that each average score band starts out in a very different place, as a result of data segmentation. However, it can be noted that each band seems to head toward a similar end-point,which is not determined by the segmentation. Additionally, one can easily see that the most change in the average similarity scores for each band happens between the first and second papers.



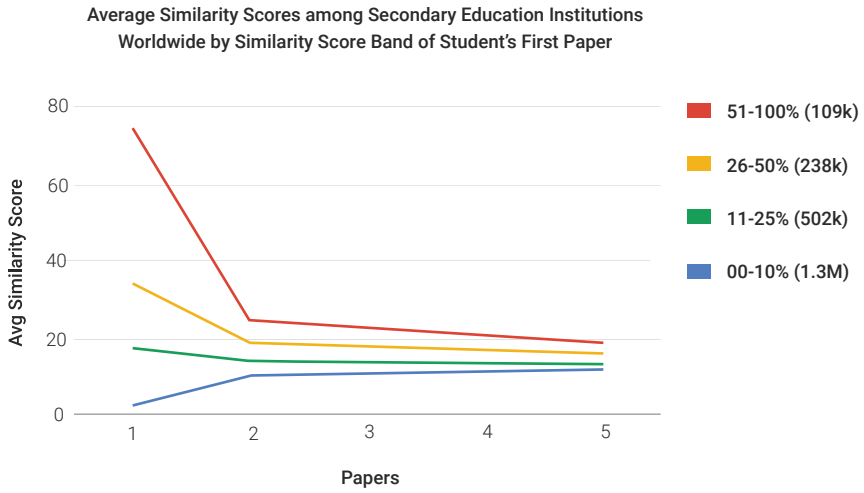
Students whose first paper had a low similarity score (10% or below) saw an increase in their similarity scores between Paper 1 (3.1% similarity) and Paper 2 (10.8%). By Paper 5, this group's average score had increased to 13.5%. This trend indicates that students in this group actually may have produced work that was too original—in other words, they had not used enough quotes, bibliographies, or other devices of reference. One might hypothesize that the immediate jump in similarity scores by Paper 2 could be the result of instructors offering feedback to the students and teaching them how and why to incorporate more sources into their writing.

The next similarity score band, which includes students whose first paper was 11-25% similar, appears almost completely flat; though it does exhibit a very gradual increase in average scores from Paper 1 (16.9%) to Paper 5 (18.0%). This very slight change may signify that instructors, on average worldwide, are expecting their students' written work to fall into this range of similarity.

This conclusion receives further evidence in the significant drop one observes in the final two bands. Papers from students who fall into the 26-50% band have an average similarity score of 34.8% on Paper 1. This score falls down to 24.3% by Paper 2 and then to 22.7% by Paper 5. Even more significantly, the scores from students in the highest similarity band decrease from 74.4% on Paper 1 to 33.9% on Paper 2, and then decline further to 27.3% by the fifth paper. We believe these large drops in similarity scores indicate that, after seeing the reports in Turnitin Feedback Studio, instructors helped students to make a swift correction to their writing and attribution practices.

3.2 The initial average similarity scores among secondary students resemble those in higher education, but decrease more steeply after Paper 1. These lower scores may reflect the differences in institutional structures and goals between secondary and higher education.

There were 2.15 million students worldwide in the secondary education sample set. Overall, the patterns one sees here are almost the same as those found in the higher education data.



In fact, the patterns resemble one another so much that one can easily see that the average similarity scores for each of the bands on Paper 1 start out almost at the same exact points:

Band	Average Similarity Score for Paper 1	
	Secondary	Higher Ed
00% - 10%	2.5%	3.1%
11% - 25%	17.0%	16.9%
26-50%	34.5%	34.8%
51 - 100%	75.4%	74.4%

However, there are some significant differences: first off, the secondary students' average similarity scores within each band all end up at lower points than those in higher education by both the second and fifth papers.

Band	Average Similarity Scores			
	Paper 2		Paper 5	
	Secondary	Higher Ed	Secondary	Higher Ed
00% - 10%	10.4%	10.8%	11.9%	13.5%
11% - 25%	14.4%	17.3%	13.9%	18.0%
26-50%	19.1%	24.3%	16.4%	22.7%
51 - 100%	24.9%	33.9%	19.4%	27.3%

One should take particular note of the movement of average similarity scores among secondary students in the 11-25% band. In higher education, the scores for students in the same band increased very slightly from 16.9% on Paper 1 to 18.0% by Paper 5. Among secondary students the scores instead decreased: from 17.0% on Paper 1 to 13.9% by Paper 5.

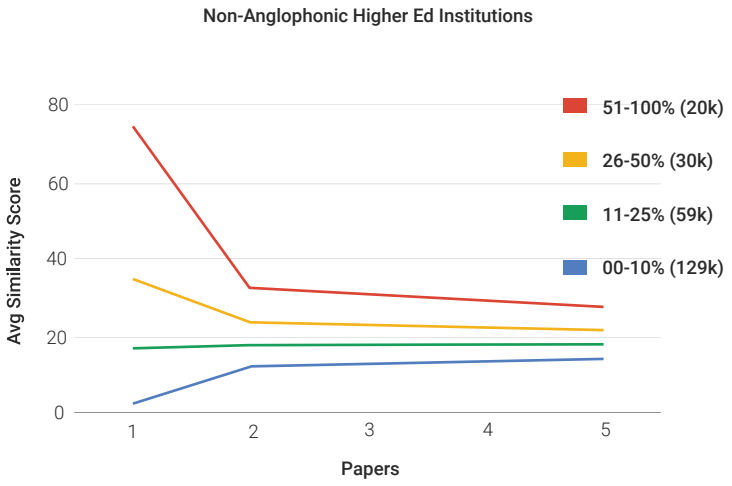
Considering only these numbers, there are several conjectures that could explain these results. In general, assignments in secondary education more than likely use fewer sources of reference and are shorter than assignments in higher education and therefore, might present fewer opportunities for misattribution or copying-and-pasting. The problem with this argument, however, is that it does not explain why the average similarity scores on the first paper are so alike to one another.

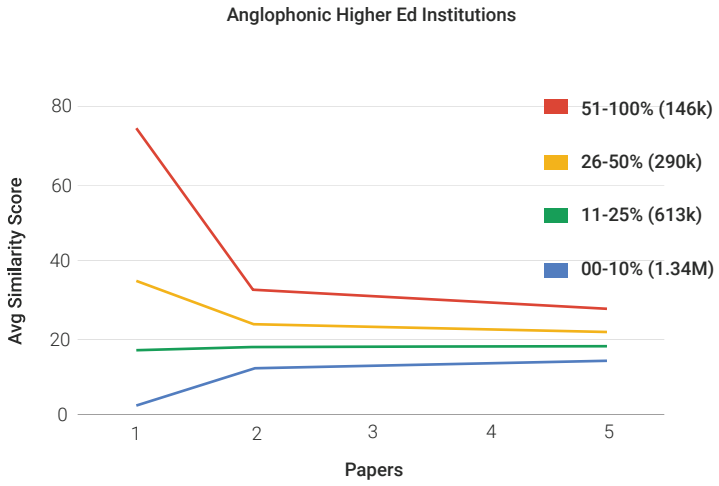
Instead, one might consider that the causes for these differences between secondary and higher education students are more nuanced and dependent upon how these two types of institutions 1) are structured and 2) use Turnitin Feedback Studio. Higher education institutions undoubtedly use the system for instructional purposes, but they are also more prone than secondary institutions to employ it primarily for the benefits of quality assurance and ensuring the integrity of student work. Class sizes in higher education institutions are also, on average, larger than those in secondary schools, which can make it more challenging to accurately assess the development of student attribution skills. In addition, many higher education institutions and/or courses may operate under the assumption that students have already acquired adequate citation and writing skills in their secondary careers.

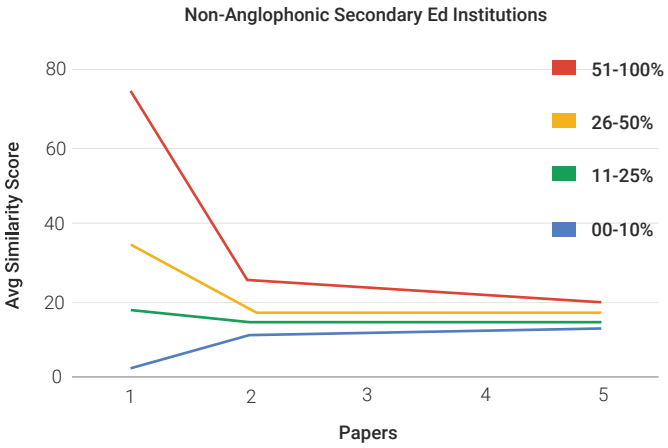
Secondary schools that have implemented Turnitin Feedback Studio, on the other hand, use it more explicitly with the intention of teaching pupils the skills they need to produce original work. Their aim is more instruction oriented, with a goal to ensure the integrity of student writing and not necessarily arising from a need to protect their school's reputation. All of this combined, it could be understood that secondary students show lower average similarity scores by their second and fifth papers than their higher education peers because they are receiving specific feedback in the classroom on how to produce original writing.

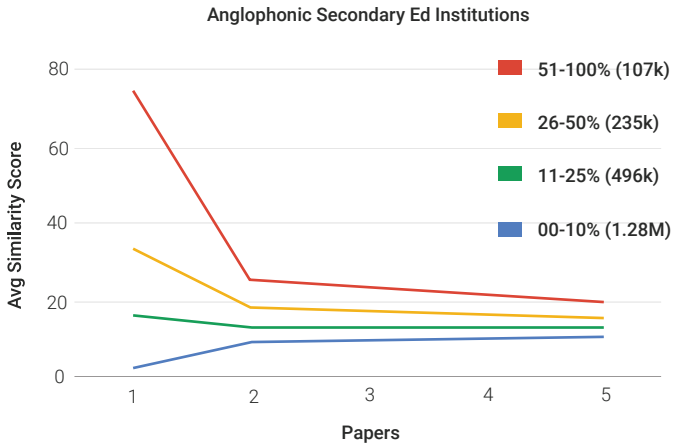
3.3 Similarity score trends are almost exactly the same in Anglophonic and non-Anglophonic countries, demonstrating that all institutions (regardless of the country in which they operate) find value in Turnitin Feedback Studio.

Recognizing the fact that institutions from Anglophonic countries predominate Turnitin Feedback Studio's base of customers, one would be right in questioning whether these data are skewed to more reflect the students' behaviors from the United States, Canada, the United Kingdom, Ireland, Australia, and New Zealand than from anywhere else. Segmenting the data along these lines clearly demonstrates that this is not the case.





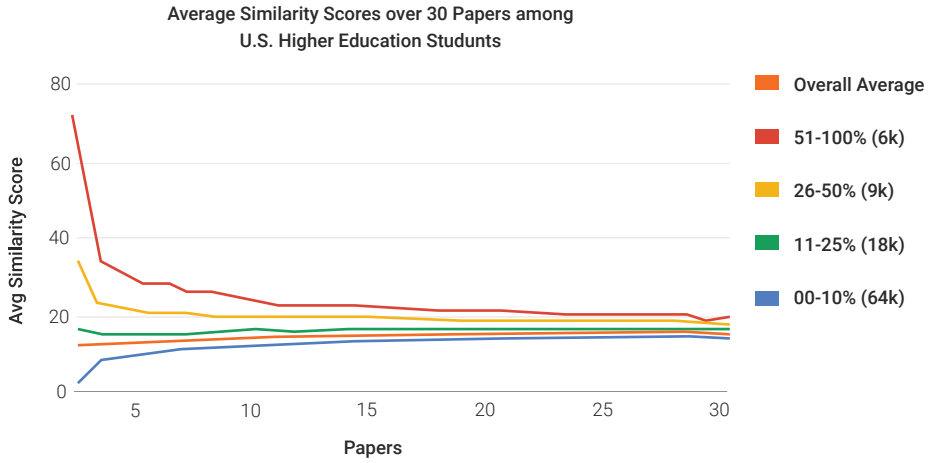




Though there are some slight differences, higher education students in countries where English is the predominant language and those where it is not both demonstrate almost identical reductions in their similarity scores, as do secondary students in each segment. These data points lend much credence to the claim that Turnitin Feedback Studio assists instructors in teaching their students how to produce original writing, no matter the culture or country in which they are using it.

3.4 Data from the United States show that the average similarity scores among students who initially submitted highly unoriginal papers continue to decline beyond five papers.

This study of average worldwide similarity scores limited its analysis to students' first five papers because there is not consistently robust datasets available across all countries beyond that. However, if higher education students in the United States who have submitted at least thirty papers to Turnitin Feedback Studio since 2014 are examined, then there is a large enough sample to get an idea of how long these effects last.



As one might expect, after having seen our previous data, the scores from students in the two highest bands show an immediate and significant drop by the second paper and continuously decline over time. Scores from students in the 11-25% band consistently hover between 15% and 17% across all thirty papers. The lowest band's average scores rise to 8.9% by Paper 2 and very gradually increase from there to 15% by the twenty-fourth paper, where they hover up until where our data end.

The overall average of all students within this sample increases slightly over the first ten papers from 12.3% to 15.0%. This average then increases more gradually to 16.0% by Paper 20, where the scores plateau. This makes sense since the largest group of students submitted initial papers with similarity scores of 10% or less. The overall average trend is important because it demonstrates the general pattern towards convergence for all of the score bands.

Conclusion

One must note that "original" here may not mean that instructors are hoping that each sentence of a work is completely unique. The trend lines observed do not head towards 0. Instead, instructors in both secondary and higher education expect to see some similarity (around 15-20%) in their students' work in the form of direct quotations, proper citations, and reference lists. Not only does Turnitin Feedback Studio help students avoid plagiarism, but it also supports instructors, across the globe, in teaching students how to include the appropriate amount of accurate references in their work.

The authors of this study believe that these trends are the results of instructors' and students' diligent use of Turnitin Feedback Studio. Instructors employ Turnitin's Similarity Reports to help students make immediate corrections to their attribution practices and use of references in their work. The students learn very quickly by their second paper how to adjust much of their approach. Over time, instructors work with students to develop their writing and hone their citation skills. Turnitin Feedback Studio aids in this process by highlighting where students can improve and produce more original writing.

Appendix: Student Data Privacy, Applied Criteria, and Student Characteristics

The aggregated student data for this analysis included no personally identifiable information. Although this approach ensured student privacy, it limited the study's ability to segment the data with strong precision. Thus, exact education levels or ages of the students are not included in the datasets. Only the similarity scores of the papers that students submitted, when those submissions occurred, whether they were secondary or higher education students, and the institution for which they submitted the papers are known.

However, it was determined that it would be most useful if the analysis focused as much as possible on the "typical" student who was early in either his or her secondary or higher education career. "Typical" is defined here as students pursuing a standard path of completion at their institution. The study's authors therefore applied filters against the dataset so that it would include only those students who had studied at their institution for at least three years. These filters limited the dataset to students whose first submission occurred in 2015 and whose last submission happened in either 2017 or 2018. One criterion was modified slightly, using a beginning year of 2014 instead, for the analysis of U.S. students' first thirty submissions.

The study also excluded some students or assignments that fell outside the normal parameters of submission activity. Students who submitted over forty papers between 2015 and 2017 or 2018 were filtered out, as that is unusually productive activity and may indicate an institution's usage of a student account for training purposes of Turnitin Feedback Studio or the activity of an instructor submitting on behalf of their students. Assignments that were submitted more than thirty days before or after an assignment deadline were also removed, as this activity could contribute more noise to the resultant data.

Since students in the dataset submitted work to the system for at least three years, it can be assumed that most students in these cohorts did not have much experience using Turnitin Feedback Studio when they turned in their first five papers. Although most students were most likely early in their education careers, their levels of skill in using proper attribution could not be assumed. Therefore, students were segmented into several bands based upon the similarity score of the first paper they submitted. The data then tracked the average scores of the students in each of these bands from their first paper to their fifth.