
 **WHITE PAPER**

The Sources in Student Writing – Higher Education

Sources of Matched Content and Plagiarism in Student Writing



Prevent Plagiarism.
Engage Students.

www.turnitin.com

Table of Contents

1.0 Summary	3
2.0 About This Study	4
3.0 Categorizing Student Sources	5
4.0 The Most Popular Sources and Sites	6
5.0 The Educational Value of Student Sources	7
6.0 Paper Mills & Cheat Sites Continue to Thrive	8
7.0 Comparing Secondary and Higher Ed Sources	9
8.0 Wikipedia Reigns Supreme	10
9.0 Recommendations for Educators	11
About Turnitin	12

1.0 Summary

This 2012 study examines which Internet sources higher education students most frequently use in their written work and the implications of those choices. It is based on an analysis of over 112 million content matches from 28 million student papers submitted to Turnitin between July 2011 and June 2012.

KEY FINDINGS INCLUDE:

Only just over half of Internet sources in student writing come from legitimate educational resources. 57% of matches come from academic and homework sites, news and portal sites and encyclopedias.

Poor research practices lead students to a significant number of sites that are not authoritative. 43% of matches lead to sites that are academically suspect, including cheat sites and paper mills, shopping sites, and social and user-generated content.

More pointedly, 19% of content matches come from paper mills and cheat sites. In looking at the issue of plagiarism, it is safe to assume, at minimum, 19% of matched content comes from sources of academic disrepute.

Already the most popular student source, reliance on Wikipedia continues strong. Wikipedia remains the most popular source for unoriginal content in student writing.

Higher education students need further instruction on proper research habits. Educators should incorporate the teaching of proper research habits upfront in order to reduce the number of academically dubious sources that appear in student writing.

2.0 About This Study

This paper is an update to a report published in November 2011, entitled “Plagiarism and the Web: A Comparison of Web Sources for Higher and Secondary Education Students.”¹

Papers submitted to the Turnitin originality checking and online grading service are compared against 30 billion current and archived web pages, 120 million textbook and academic journals and 275 million student papers. For the purposes of this study, only matched content between student papers and Turnitin’s archive of the Web were considered.

For each month of the study, Turnitin documented the top 100 Internet sites (domains) that displayed matches between student papers and Turnitin’s database of 30 billion web pages. Each site was identified, reviewed and placed into one of six categories:

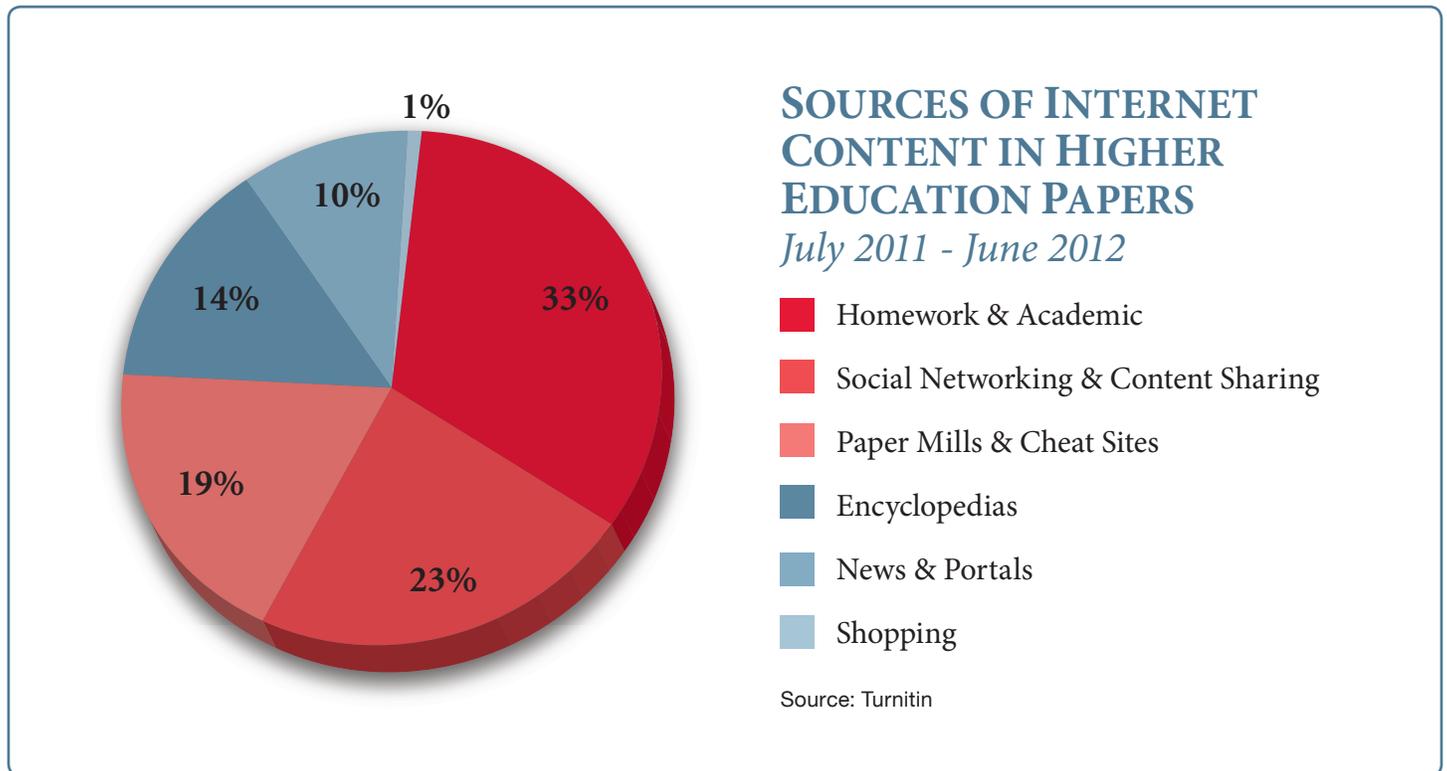
Category	Definition	Examples
Social Networking & Content Sharing	Sites that rely on user-generated content rather than professionally-published content.	Answers.com, Facebook, Scribd, SlideShare, Yahoo Answers
Homework & Academic	Academic, educational and homework help sites that offer a range of content to educate students.	Google Books, Project Gutenberg, Mayo Clinic, MedLibrary, Springer, Sage
News & Portals	This category consists of professionally-published content.	Associated Content, Examiner, The Huffington Post, The New York Times
Paper Mills & Cheat Sites	Sites that promote the sale or sharing of written work.	All Free Papers, All Free Essays, Essay Mania, Frat Files, Other People’s Papers
Encyclopedias	Reference and summary sites	Wikipedia, Britannica, Encyclopedia.com, Internet Movie Database
Shopping	Sites that offer reviews and sell products	Amazon, Barnes & Noble, Buy.com

Note: The matched texts identified for study in this report **may or may not have** been properly cited and attributed. Also, no school, student or other identifying information was viewed when conducting the analysis.

¹See: http://pages.turnitin.com/plagiarism_sec_vs_he.html

3.0 Categorizing Student Sources

The chart below shows the percentage of sources by category of student writing based on six categories.



Category	2011-2012	2010-2011	Change
Homework & Academic	33%	22%	+11
Social Networking & Content Sharing	23%	26%	-3
Paper Mills & Cheat Sites	19%	20%	-1
Encyclopedias	14%	12%	+2
News & Portals	10%	17%	-7
Shopping	1%	2%	-1

Comparing the results to the previous year reveals some significant changes. Homework & Academic sites showed the biggest increase (+11%), displacing the most popular category last year, Social Networking & Content Sharing sites (-3%). Paper Mills & Cheat Sites increased two percentage points as did Encyclopedias sites, with Wikipedia accounting for the majority of the gain. There was also a marked decrease in the number of News & Portals sites.

4.0 The Most Popular Sources and Sites

MOST POPULAR SITES.

The top 10 most popular Internet sources for student writing include four Social Networking & Content Sharing sites, three Paper Mills & Cheat Sites, two Homework & Academic sites and one Encyclopedias site.

2011-2012 Top Sites	%	Category	2010-2011 Top Sites	%	Category
wikipedia.org	11%	Encyclopedias	wikipedia.org	11%	Encyclopedias
oppapers.com	4%	Paper Mills	answers.yahoo.com	4%	Social
slideshare.net	4%	Social	slideshare.net	4%	Social
coursehero.com	4%	Homework	answers.com	4%	Social
scribd.com	3%	Social	oppapers.com	3%	Paper Mills
answers.yahoo.com	3%	Social	coursehero.com	3%	Homework
answers.com	3%	Social	scribd.com	3%	Social
medlibrary.org	3%	Academic	justanswer.com	2%	Paper Mills
bignerds.com	2%	Paper Mills	enotes.com	2%	Homework
papercamp.com	2%	Paper Mills	amazon.com	1%	Shopping

Seven of the top 10 sites appear in our most recent study as well as in the previous time period (July 2010 to June 2011). Three Paper Mills & Cheat Sites appear in the top 10 list of the most recent academic year compared to just one in the previous year.

In terms of potential for educational value, only three sites—Wikipedia, MedLibrary and Course Hero—could be considered as having academic merit. Three sites stood out as untrustworthy, for their practice of sharing and/or selling written work (Oppapers, BigNerds and PaperCamp). The four remaining sites all fall under the Social Networking & Content Sharing category: SlideShare, Scribd, Yahoo! Answers and Answers.com. The educational value of the content in these sites is quite varied, but instructors would most likely prefer not to see these types of sites included in a list of cited works.

5.0 The Educational Value of Student Sources

While results may vary by individual sites, Homework & Academic sites, Encyclopedias and News & Portal Sites all provide some degree of content with educational value. Given these criteria, roughly 57% of identified sources can be linked to sites that offer some content of educational value.

Category	2012
Homework & Academic	33%
Encyclopedias	14%
News & Portals	10%
Total	57%

Conversely, Paper Mills & Cheat Sites, Shopping sites and Social Networking & Content Sharing sites fail to meet the criteria of possessing educational value. Book reviews (from Amazon.com, for example, the most popular domain in Shopping) and anonymously submitted opinions on answer sites (Answers.com, for example) also do not pass muster as material that should be included in student writing. As such, 43% of sources that appear in student writing could be considered inappropriate for inclusion in academic work.

Category	2012
Social Networking & Content Sharing	23%
Paper Mills & Cheat Sites	19%
Shopping	1%
Total	43%

6.0 Paper Mills & Cheat Sites Continue to Thrive

The category of Paper Mills & Cheat Sites provide concrete examples of sources students should not use in their writing. Unfortunately, these sites continue to be a common source for content that appears in student papers, growing from 17-19% in the most recent academic year.

While educators can debate the educational value of using Wikipedia or SlideShare.net in research, it is clear that drawing content from sites such as Opppapers.com (i.e. Other People's Papers) or Fratfiles.com is not an acceptable practice.

Below is a list of the top 10 cheat sites and paper mills that appear in student work.

Top Cheat Sites HE 2012		# of Matches
1	oppapers.com	4,492,283
2	bignerds.com	2,633,248
3	papercamp.com	2,454,094
4	antiessays.com	2,453,562
5	enotes.com	2,390,593
6	allfreeessays.com	2,068,932
7	termpaperslab.com	1,392,284
8	cyberessays.com	883,942
9	ivythesis.typepad.com	782,726
10	essaymania.com	653,638

These sites, along with a number of other cheat sites and paper mills that appear further down on the list, represent a surprising 19% of all sources in student writing. In short, nearly 20 percent of all sources used in higher education papers are plagiarized.

7.0 Comparing Secondary and Higher Ed Sources

A comparison of higher and secondary sources shows some similarities but also stark differences.

Secondary Sources	%	Higher Ed Sources	%
Homework & Academic	33%	Homework & Academic	33%
Social Networking & Content Sharing	28%	Social Networking & Content Sharing	23%
Paper Mills & Cheat Sites	18%	Paper Mills & Cheat Sites	19%
Encyclopedias	11%	Encyclopedias	14%
News & Portals	6%	News & Portals	10%
Shopping	4%	Shopping	1%
Total 100%	100%	Total	100%

50% of the sources secondary students use in writing comes from legitimate academic resources (Homework & Academic, Encyclopedias and News & Portals), while higher education students relied on these sites 57% of the time. A closer look at the categories shows where secondary and higher education students differ.

- Both levels show the same reliance on Homework & Academic sites (33%), the most popular category.
- Secondary students depend more on Social Networking & Content Sharing sites, most likely because they have less experience assessing the value of student sources.
- Reliance on Paper Mills & Cheat Sites remains stubbornly high though roughly equal for both categories.
- Higher education students showed a great dependence on Encyclopedias sites (largely Wikipedia) and for News & Portal sites that present professionally-published content.
- Secondary students favor Shopping sites to a much great degree than higher education students, where sites such as Amazon.com and Barnes & Noble barely register. Once again, this is most likely a reflection of the difficulty younger students face in assessing and including academically-viable source material.

8.0 Wikipedia Reigns Supreme

Wikipedia has an outsized presence as a content source for student writing. In terms of the number of matches, Wikipedia appears almost three times more than the second most popular student source, Other People's Papers (oppapers.com).

Wikipedia is a public good that offers an ever-growing wealth of information on just about every imaginable topic. Even so, educators hold mixed feelings about Wikipedia, specifically because it is a) crowd-sourced b) not a primary source and c) its degree of accuracy, while high, is not substantiated by the content checks and review that traditional encyclopedic content undergo. As for how to guide students in the appropriate use of Wikipedia, the next section will share some advice.

9.0 Recommendations for Educators

There are four steps educators can take to ensure students use and cite the most appropriate sources on the Web.

1. DON'T ASSUME STUDENTS KNOW HOW TO CONDUCT PROPER RESEARCH

Based on this analysis, it is clear that many higher education students do not have a strong grasp of what it means to conduct proper academic research. Instructors at colleges and universities who assume that students know how to sift through the myriad sources on the Internet and find educationally-valuable information will most likely find papers with spurious sources on their desks at the end of the term. In writing intensive courses, instructors should dedicate their first class and assignments to educating students on what constitutes academically appropriate sites and share best practices for uncovering sources that have high educational value.

2. GOOGLE IS THE FIRST STEP, NOT THE LAST, IN RESEARCH

Students think they know how to use Google, but when it comes to academic research, it is clear that they do not. Google's algorithms are tuned for relevance and popularity, not academic authority. When it comes to finding academic sources, teach students that they cannot defer to Google results and instead must use their judgment to decide what sources are of academic value. For more information, please view the Turnitin webcast "*New Demands, New Approaches: Achieving Information Literacy and Competency in the 21st Century.*" http://pages.turnitin.com/HE_22_archive.html

3. WIKIPEDIA IS A STARTING POINT

Like Google, Wikipedia should be a starting point in a student's research process. Educate students that the true value of Wikipedia is to provide a curated summary on a topic, and that they should follow the sources and citations at the bottom of Wikipedia entries to verify the accuracy of the information and to uncover primary source material.

4. EMPLOY TOOLS LIKE TURNITIN

In this digital age, educators need to use educational technology to better assess the value of student sources while providing better feedback to students. Turnitin helps instructors to uncover improper sources in student papers and to give better feedback on student writing. Within the first year of usage, Turnitin users see a reduction in unoriginal content of 10-35%. By the third year, institutions see levels of unoriginality fall by 50%, with some seeing improvements up to 75%. Moreover, educators who use Turnitin report a 46% increase in student engagement, a 52% increase in the quality of feedback to students and a 31% reduction in time spent grading.

About Turnitin

Turnitin is the global leader in evaluating and improving student writing. The company's cloud-based service for originality checking, online grading and peer review saves instructors time and provides rich feedback to students. One of the most widely distributed educational applications in the world, Turnitin is used by more than 10,000 institutions in 126 countries to manage the submission, tracking and evaluation of student papers online. Turnitin also offers iThenticate, a plagiarism detection service for commercial markets, and WriteCheck, a suite of formative tools for writers.

Turnitin is backed by Warburg Pincus and is headquartered in Oakland, Calif., with an international office in Newcastle, U.K.

For more information, please visit <http://turnitin.com>.

Turnitin is a service of iParadigms, LLC
1111 Broadway, 3rd Floor
Oakland CA, 94607 USA

USA/Canada: 866-816-5046
International: +1 510-764-7600
sales@turnitin.com
www.turnitin.com

All products and services mentioned in this document are trademarks or registered trademarks of their respective owners.
©2013 iParadigms, LLC All rights reserved. Version 0113.