WHITE PAPER What's Wrong with Wikipedia?

Evaluating the Sources Used by Students

turnitin®

Prevent Plagiarism.
Engage Students.

Table of Contents

1.0 Introduction

" The internet has changed the very meaning of 'research'. "

- Pew Internet and American Life Project

In a recent report entitled, "How Teens Do Research in the Digital World," the Pew Research Center confirms what many educators already know: student reliance on Internet-based search has replaced the more rigorous and traditional approaches to research. While the survey highlights the value teachers believe the Internet provides (99 percent agreement) for empowering student access to information, the survey of over 2,000 middle and high school teachers also found that 64 percent reported that digital technologies are more distracting than helpful from an academic standpoint. In short, what constitutes "research" for students today has come to mean "Googling." [1]

As a way to address this gap in student skills, Turnitin has developed a source evaluation rubric for educators to share with their students. The rubric, created by secondary and higher education instructors, is designed to help students evaluate sources that they use in their writing. Its intent is to help enhance student mastery of "21st century," information literacy skills critical for academic, professional and career success in the digital age.

This white paper begins by highlighting the problem by reviewing findings from The Pew Research Center and Turnitin's own study of student research practices. Next is an overview of the rubric that includes examples of how common source sites perform against the rubric. Finally, this study will offer guidance on how the rubric can be used by instructors and students to improve student research skills.

[1] Pew Research Center, CollegeBoard, and National Writing Project, *How Teens Do Research in the Digital World* (Washington DC: PIP, 2012) 3.

2.0 Student 'Research' in the Digital Age

How do students research in the digital age?

The Pew report shows that the ease with which information "appears" online allows students to avoid any of the questions that may surface concerning the quality and intent of information they "research." The Pew survey revealed that only one percent of those surveyed reported as "excellent" the ability of students "to recognize bias in online content." As for their "ability to assess the quality and accuracy of information they find online," only three percent reported that they found students to be "excellent." [2]

The Pew study also revealed that instructors were not rudderless in terms of addressing this issue, finding that teachers turn to providing instruction on how search engines work and teaching students how to evaluate sources to meet the research skills gap. [3] More specifically, the teachers surveyed called for the necessity of teaching digital literacy skills in schools (93 percent of teachers surveyed support this change). And, when asked to rank the value of skills that students may need in the future, teachers responded by ranking the ability to evaluate the quality of information and to write effectively at the top of the list. [4]

Turnitin's own research reinforces many of the findings in the Pew report while providing a greater level of detail in terms of which Internet sources students include in their writing. Turnitin analyzed over 37 million higher and secondary education student papers submitted to the service from July 2011 to June 2012 and categorized each source into one of six categories. In these papers, Turnitin identified 156 million matches between content in the paper and the Internet. The chart below highlights the breakout of matches by category.

Category	Definition	Examples	% of Matches*
Homework & Academic	Academic, educational and homework help sites and content to educate students.	Google Books, Project Gutenberg, Mayo Clinic, MedLibrary, Springer, Sage	33%
Social Networking & Content Sharing	Sites that rely on user-generated content rather than professionally-published content.	Answers.com, Facebook, Scribd, SlideShare, Yahoo Answers.	24%
Paper Mills & Cheat Sites	Sites that promote the sale or sharing of written work.	All Free Essays, Essay Mania, Frat Files, Other People's Papers	18%
Encyclopedias	Reference and summary sites	Wikipedia, Britannica, Enyclopedia.com, Internet Movie Database	13%
News & Portals	This category consists of professionally-published content.	Examiner, The Huffington Post, The New York Times	9%
Shopping	Sites that offer reviews and sell products	Amazon, Barnes & Noble, Buy.com	2%

[2] Pew Research Center, 6. [3] Pew Research Center, 2. [4] Pew Research Center, 46. *Total does not add up to 100% due to rounding.

This data supports the following insights into student research behavior, specifically:

- **Students appear to value immediacy over quality in online research**,
  The ease with which "the answer" may be found online places sites such as Wikipedia, homework help sites, answer sites, and other social and content sharing sites to the top in terms of source matches.

- **Students often use cheat sites and paper mills as sources**
  Less a research competency issue than a moral and ethical one, the significant number of sources that match to cheat sites and paper mills suggest that for students there is a bias towards immediate outcomes and results rather than towards concerted effort to meet assignment goals.

- **There is an over reliance on the "wisdom of the crowd"**
  Students appear to demonstrate a strong appetite for crowd-sourced content in their research. Though it is not immediately evident why students seek these sources out, the strong reliance on these types of sites indicate difficulty assessing the authority and legitimacy of the content these sources present.

- **Student "research" is synonymous with "search"**
  The frequent and uninhibited use of sites with limited educational value (as defined by the quality and authority of content) in student work underscores a preference for "searched," rather than "researched" content.

- **Existing student source choices warrant a need for better search skills**
  In addition to a preference for immediacy, the popularity of crowd-sourced content online indicates that a majority of students are engaging in cursory or shallow searches for content. At play may be an absence of awareness of how search engines work and how to effectively conduct searches to find appropriate content. What also appears to be absent is the use of criteria (whether internally—or externally—defined) to judge that content.

3.0 Evaluating Online Sources

The Source Educational Evaluation Rubric (SEER) represents the evolution of the critical approach that Turnitin has adopted and used to categorize websites in our analysis of student sources. The rubric was designed by academic experts and used by secondary and higher education educators who field-tested the rubric by using it to evaluate over 300 of the most popular student sources (which will be shared in a follow-up white paper.).

The rubric is built on five criteria:

- **Authority**: Is the site well regarded, cited, and written by experts in the field?
- **Educational Value**: Does the site content help advance educational goals?
- **Intent**: Is the site a well-respected source of content intended to inform users?
- **Originality**: Is the site a source of original content and viewpoints?
- **Quality**: Is the site highly vetted with good coverage of the topical area?

These criteria are evaluated along a numerical scale anchored by an explicit call out to "credibility," a move to make the scores more informative for students.

Instructors and students who use SEER can quickly arrive at an easy-to-interpret score based on the commonly used 4.0 grade point scale. By adding up all criteria values and dividing by five, users will generate a readily-understandable grade for sources. If so desired, the weighting of the criteria can also be adjusted to reflect varying evaluation-directed objectives (see the accompanying SEER Worksheet in the appendix).

- 3.0 - 4.0: highly credible, quality sources
- 2.0 – 3.0: credible sources
- 1.0 – 2.0: questionable sources
- 0.0 – 1.0: unacceptable or inappropriate sources

The rubric, in its entirety, appears below. Following the rubric are a few examples of sites and how instructors have scored them.

4.0 Source Educational Evaluation Rubric (SEER)

	Highly Credible		Credible		Discreditable
	4	3	2	1	0
AUTHORITATIVE	Highly-regarded site that is referenced and linked to by others. Information is well-referenced, cited, and written by authors with expertise in the content area	Well-regarded site that is well known in its category. Information is referenced and cited, with experts generating content	Regarded site that has reputable information that is referenced and cited	Site is slightly regarded--does not necessarily have information that is researched, referenced, or cited	Site is not regarded--information or resources provided are not researched, referenced, or cited
EDUCATIONAL VALUE	Site is a leading source of content that exceeds instructional goals	Site is a regarded source of content that meets instructional goals	Site content meets instructional goals	Site content does not meet instructional goals	Site content is inappropriate or actively subverts instructional goals
INTENT	Site is a highly-respected source of content provided to inform users	Site is a well-regarded source of information whose job is to inform	Site provides content to inform users	Site promotes content; the goal of the site is not to "inform" users	Site actively sells content
ORIGINALITY	Site is a primary source of original content and viewpoints	Site offers original content and viewpoints	Site offers some original content, combining it with repurposed content	Site aggregates or repurposes content	Site aggregates or repurposes content without appropriate citation or references
QUALITY	Site content is high quality (vetted, researched, and informed) and enough content is provided to establish good coverage	Site content is of good quality and there is a good content area coverage	Site content is of quality and there is some broader content coverage	Site content quality is lacking and good content is in limited quantity	Quality of content is questionable. Lack of quality information available.

Putting SEER into Practice

Following are three examples of SEER in practice. We evaluated three popular sites that appear in both secondary and higher education papers. Included in the analysis is the ranking by criterion, the number of matches this site produced against Turnitin's index of the web, and the overall score of the web site based on SEER.

Highly Credible

The New York Times
http://www.nytimes.com

# of Matches in Student Papers: 768,584

Total SEER Score: 3.6

	Highly Credible		Credible		Discreditable
	4	3	2	1	0
AUTHORITATIVE = 4.0	Highly-regarded site that is referenced and linked to by others. Information is well-referenced, cited, and written by authors with expertise in the content area	Well-regarded site that is well known in its category. Information is referenced and cited, with experts generating content	Regarded site that has reputable information that is referenced and cited	Site is slightly regarded--does not necessarily have information that is researched, referenced, or cited	Site is not regarded--information or resources provided are not researched, referenced, or cited
EDUCATIONAL VALUE = 3.8	Site is a leading source of content that exceeds instructional goals	Site is a regarded source of content that meets instructional goals	Site content meets instructional goals	Site content does not meet instructional goals	Site content is inappropriate or actively subverts instructional goals
INTENT = 3.3	Site is a highly-respected source of content provided to inform users	Site is a well-regarded source of information whose job is to inform	Site provides content to inform users	Site promotes content; the goal of the site is not to "inform" users	Site actively sells content
ORIGINALITY = 3.5	Site is a primary source of original content and viewpoints	Site offers original content and viewpoints	Site offers some original content, combining it with repurposed content	Site aggregates or repurposes content	Site aggregates or repurposes content without appropriate citation or references
QUALITY = 3.5	Site content is high quality (vetted, researched, and informed) and enough content is provided to establish good coverage	Site content is of good quality and there is a good content area coverage	Site content is of quality and there is some broader content coverage	Site content quality is lacking and good content is in limited quantity	Quality of content is questionable. Lack of quality information available.

Description:

The New York Times has an industry-leading reputation for accurate, timely, and unbiased reporting. Stories and content are original and bylined. Opinion-driven pieces are clearly demarcated as such. More significantly, the Times is the publication that other outlets refer to in their own reporting (Authoritative =4). As for "Intent," "Originality," and "Quality," the Times is primarily a vehicle for communicating the news and not a pure source of original content (like an encyclopedia, Britannica for example).
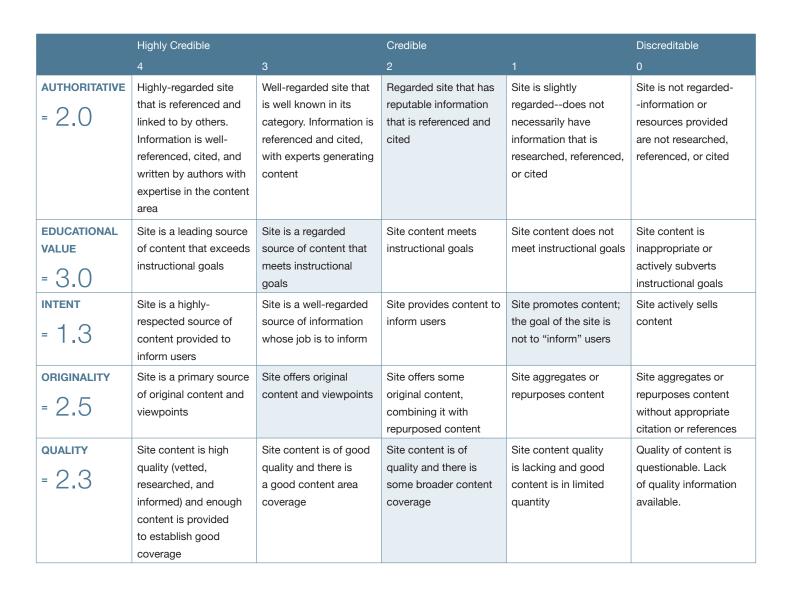
Credible

eNotes

http://www.enotes.com/

# of Matches in Student Papers: 1,440,129

Total SEER Score: 2.20

	Highly Credible		Credible		Discreditable
	4	3	2	1	0
AUTHORITATIVE = 2.0	Highly-regarded site that is referenced and linked to by others. Information is well-referenced, cited, and written by authors with expertise in the content area	Well-regarded site that is well known in its category. Information is referenced and cited, with experts generating content	Regarded site that has reputable information that is referenced and cited	Site is slightly regarded--does not necessarily have information that is researched, referenced, or cited	Site is not regarded--information or resources provided are not researched, referenced, or cited
EDUCATIONAL VALUE = 3.0	Site is a leading source of content that exceeds instructional goals	Site is a regarded source of content that meets instructional goals	Site content meets instructional goals	Site content does not meet instructional goals	Site content is inappropriate or actively subverts instructional goals
INTENT = 1.3	Site is a highly-respected source of content provided to inform users	Site is a well-regarded source of information whose job is to inform	Site provides content to inform users	Site promotes content; the goal of the site is not to "inform" users	Site actively sells content
ORIGINALITY = 2.5	Site is a primary source of original content and viewpoints	Site offers original content and viewpoints	Site offers some original content, combining it with repurposed content	Site aggregates or repurposes content	Site aggregates or repurposes content without appropriate citation or references
QUALITY = 2.3	Site content is high quality (vetted, researched, and informed) and enough content is provided to establish good coverage	Site content is of good quality and there is a good content area coverage	Site content is of quality and there is some broader content coverage	Site content quality is lacking and good content is in limited quantity	Quality of content is questionable. Lack of quality information available.

Description:

eNotes is a subscription-based site that offers authored content: literature study guides, lessons plans, literary criticism, and discussions for users. The content is original, authored and published by eNotes. In addition, there is a vibrant group of "eNotes Educators" (teachers) who contribute to discussions, answer questions, and moderate the site, thus ensuring quality of content and transparency in terms of content sources. Information on the site is also referenced, with instructions on how to properly cite eNotes content also provided. The site is credible in terms of its Authority, "Educational Value," "Originality," and "Quality." As for "Intent," the site clearly positions itself as a promoter of content.

Discreditable

123HelpMe.com
http://123helpme.com

# of Matches in Student Papers: 2,370,843

Total SEER Score: 0.46

	Highly Credible		Credible		Discreditable
	4	3	2	1	0
AUTHORITATIVE = 0.5	Highly-regarded site that is referenced and linked to by others. Information is well-referenced, cited, and written by authors with expertise in the content area	Well-regarded site that is well known in its category. Information is referenced and cited, with experts generating content	Regarded site that has reputable information that is referenced and cited	Site is slightly regarded--does not necessarily have information that is researched, referenced, or cited	Site is not regarded--information or resources provided are not researched, referenced, or cited
EDUCATIONAL VALUE = 0.8	Site is a leading source of content that exceeds instructional goals	Site is a regarded source of content that meets instructional goals	Site content meets instructional goals	Site content does not meet instructional goals	Site content is inappropriate or actively subverts instructional goals
INTENT = 0	Site is a highly-respected source of content provided to inform users	Site is a well-regarded source of information whose job is to inform	Site provides content to inform users	Site promotes content; the goal of the site is not to "inform" users	Site actively sells content
ORIGINALITY = 0.8	Site is a primary source of original content and viewpoints	Site offers original content and viewpoints	Site offers some original content, combining it with repurposed content	Site aggregates or repurposes content	Site aggregates or repurposes content without appropriate citation or references
QUALITY = 0.3	Site content is high quality (vetted, researched, and informed) and enough content is provided to establish good coverage	Site content is of good quality and there is a good content area coverage	Site content is of quality and there is some broader content coverage	Site content quality is lacking and good content is in limited quantity	Quality of content is questionable. Lack of quality information available.

Description:

There are a number of telling indicators on the 123HelpMe.com homepage that indicate that the site has questionable intentions. First, there is a call to search their "free directory" of "essays, research papers, term papers, and speeches." There is also the ability to browse papers by content categories. And, to further assist students' search for essays, there is an "Essay Color Key" to help students distinguish between "unrated" to "powerful" essays. A deeper look into the site reveals an archive of papers that may be used for free or for a fee. No author or byline information is provided for these essays, making it impossible to assess their authoritativeness. Because this site is a purveyor of essays, it lacks "Educational Value" and has a discreditable "Intent." In terms of "Originality" and "Quality," the site is wholly unoriginal and promotes the use of unoriginal content. And, again, because of the lack of author information or essay/article bylines, the quality of the content is questionable.

5.0 Recommendations

The Source Educational Evaluation Rubric is an effective tool for helping students to surface their own internal biases and assumptions around online search and source evaluation. At the same time, the rubric can be tremendously useful for introducing students to a critical approach to researching sources. To successfully use the rubric with students, we suggest the following approach:

1. **First, establish a baseline to help students identify their own biases and assumptions**
   Provide students with a list of common sites to review (a range of 10 across the credibility spectrum). Have individual students complete their own ratings of the sites and then share out the results to stimulate discussion.

2. **Establish a degree of inter-rater reliability**
   Via student discussion of site ratings, identify points of departure on scores and use criterion score differences as an opportunity to open dialogue about the search and research process. For example, ask students to consider why an "authoritative" source has greater merit as a source for information for a paper. Digging deeper, ask students to consider the problems with crowd-sourced content.

3. **Use this opportunity to share your ratings**
   There is tremendous value in having students understand instructors' perspectives on site credibility. This is an effective way to provide students with some clear-cut guidance on how to review sources. And, similar to any grading rubrics that you may employ, The Source Educational Evaluation Rubric provides instructors with an opportunity to share specific expectations for the approach students take to researching information for their written work.

6.0 Appendix: SEER Worksheet

	Highly Credible		Credible		Discreditable
	4	3	2	1	0
AUTHORITATIVE Criteria Weight	Highly-regarded site that is referenced and linked to by others. Information is well-referenced, cited, and written by authors with expertise in the content area	Well-regarded site that is well known in its category. Information is referenced and cited, with experts generating content	Regarded site that has reputable information that is referenced and cited	Site is slightly regarded--does not necessarily have information that is researched, referenced, or cited	Site is not regarded--information or resources provided are not researched, referenced, or cited
EDUCATIONAL VALUE Criteria Weight	Site is a leading source of content that exceeds instructional goals	Site is a regarded source of content that meets instructional goals	Site content meets instructional goals	Site content does not meet instructional goals	Site content is inappropriate or actively subverts instructional goals
INTENT Criteria Weight	Site is a highly-respected source of content provided to inform users	Site is a well-regarded source of information whose job is to inform	Site provides content to inform users	Site promotes content; the goal of the site is not to "inform" users	Site actively sells content
ORIGINALITY Criteria Weight	Site is a primary source of original content and viewpoints	Site offers original content and viewpoints	Site offers some original content, combining it with repurposed content	Site aggregates or repurposes content	Site aggregates or repurposes content without appropriate citation or references
QUALITY Criteria Weight	Site content is high quality (vetted, researched, and informed) and enough content is provided to establish good coverage	Site content is of good quality and there is good content area coverage	Site content is of quality and there is some broader content coverage	Site content quality is lacking and good content is in limited quantity	Quality of content is questionable. Lack of quality information available.

Rubric Score _____   Rubric Percentage _____   sales@turnitin.com | **turnitin.com**

Examples

Below are three examples of SEER in practice. We evaluated three popular sites that appear in both secondary and higher education papers. A total of 300 sites were reviewed as part of this study. Included in the analysis is the ranking by criterion, the number of matches this site produced against Turnitin's index of the Web, and the overall score of the site based on SEER.

HIGHLY CREDIBLE	CREDIBLE	DISCREDITABLE
The New York Times	**eNotes**	**123HelpMe.com**
http://www.nytimes.com	http://www.enotes.com/	http://123helpme.com
# of Matches in Student Papers: 768,584	# of Matches in Student Papers: 1,440,129	# of Matches in Student Papers: 2,370,843
Total SEER Score: 3.6	Total SEER Score: 2.20	Total SEER Score: 0.46

AUTHORITATIVE

4.0 Highly-regarded site that is referenced and linked to by others. Information is well-referenced, cited, and written by authors with expertise in the content area	**2.0** Regarded site that has reputable information that is referenced and cited	**0.5** Site is slightly regarded--does not necessarily have information that is researched, referenced, or cited

EDUCATIONAL VALUE

3.8 Site is a leading source of content that exceeds instructional goals	**3.0** Site is a regarded source of content that meets instructional goals	**0.8** Site content does not meet instructional goals

INTENT

3.3 Site is a highly-regarded source of information whose job is to inform	**1.3** Site promotes content; the goal of the site is not to "inform" users	**0** Site actively sells content

ORIGINALITY

3.5 Site is a primary source of original content and viewpoints	**2.5** Site offers original content and viewpoints	**0.8** Site aggregates or repurposes content

QUALITY

3.5 Site content is of good quality and there is a good content area coverage	**2.3** Site content is of quality and there is some broader content coverage	**0.3** Quality of content is questionable. Lack of quality information available.

The New York Times has an industry-leading reputation for accurate, timely, and unbiased reporting.	eNotes is a subscription-based site that offers authored content: literature study guides, lessons plans, literary criticism, and discussions for users. The content is original, authored and published by eNotes.	The intent of 123HelpMe.com is clearly to share and sell essays for student use. Because of the lack of author information or essay/article bylines, the quality of the content is questionable.

*Examples from secondary instructor SEER reviews of sources.

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

turnitin.com