Do College Anti-Plagiarism/ Cheating Policies Have Teeth in the Age of AI? Evidence from the United States

Rajeev K. Goel, Michael A. Nelson

Do College Anti-Plagiarism/Cheating Policies Have Teeth in the Age of AI?
Evidence from the United States

Abstract

The advent of the internet, and more recently of artificial intelligence (AI), has challenged academic and other institutions to ensure ethical practices and reward/promote true merit. The borderless and relatively anonymous nature of the internet creates policing challenges, leading to the abuse of established rules and standards. In the context of academia, this impacts the size and scope of resources to facilitate/check plagiarism and cheating, both from the demand and supply sides. Adding some formal insights into the current topic of fundamental importance to maintaining academic integrity, this paper examines the association of anti-plagiarism/anti-cheating policies with resources that facilitate such behavior (legal or otherwise). Using unique internet search indices of the policies and resources, we find that the two are positively associated – the associated resources ratchet up with the policies. This association is robust to different modeling formulations, including when the internet policies include course syllabi. The findings reinforce the view that policies to check plagiarism and cheating are likely to lack teeth and may be a step behind the resources that facilitate unethical behaviour.

JEL-Codes: A200, I230, L860.

Keywords: AI, artificial intelligence, plagiarism, cheating, internet, universities, colleges, United States.

Rajeev K. Goel*
Illinois State University
USA – Normal, IL 61790
rkgoel@ilstu.edu

Michael A. Nelson
University of Akron / OH / USA
nelson2@uakron.edu

*corresponding author

1. Introduction

The advent of the internet earlier and of artificial intelligence (AI) more recently have been nothing short of game changers in the production, delivery, and consumption of many goods and services (see Goel and Hsieh (2002)). Not all impacts of these technologies have been necessarily positive. One adverse outcome, coming especially to the fore with generative AI technologies is the issue of preserving and rewarding the integrity of true effort. This aspect, especially in relation to plagiarism and cheating in academic settings forms the focus of the present work.

The problem of plagiarism and cheating by students in academia has been longstanding, with recognition of the different types of plagiarism (Awasthi (2019), de Maio et al. (2019), Goffe and Sosin (2005), Kaposi and Dell (2012), Park (2003), Pecorari (2015), Walker and Townley (2012)). However, the recent developments related to the internet and more recently to generative AI (artificial intelligence) have arguably added a generational shift in the students' and other perpetrators' abilities to bypass the formal evaluation channels and fake signals of their true credentials (Cotton et al. (2023), Lo (2023), Padillah (2023)). The global reach of the internet has enabled internet abuse enablers to be located outside of the jurisdictions of individual nations. The issue of AI-aided abuse has drawn the attention of administrators and featured in the press in several places, although the technology is still evolving. See, for examples: https://calmatters.org/education/higher-education/2023/10/college-application-essays/; https://dbknews.com/2023/11/27/umd-chatgpt-academic-integrity-cases/; https://www.statepress.com/article/2023/07/ai-cheating-crackdown

Evaluators and instructors have been aware of the challenges posed by the incentives and abilities to cheat, and have devised different mechanisms and punishments to curb such undesirable behavior (see Tsertsvadze and Khurtsia (2020) for an example, and Awasthi (2019) for a broader review). However, the speed and scope of the technological changes related to AI have left monitors somewhat flatfooted in their abilities to control the use of new technologies that facilitate cheating (see Cotton et al. (2023), Lo (2023), Padillah (2023)). Yet, it is crucial to determine whether and to what extent the countermeasures to control cheating are effective and it is towards this that the present research is directed.

Recently, the European Union has agreed on some steps to regulate AI - https://www.bbc.com/news/world-europe-67668469. The effectiveness of these initiatives will only be known over time, especially when the underlying technologies are fast evolving.

Even before the AI developments, especially the ChatGPT software, there were questions about the efficacy of anti-plagiarism/anti-cheating initiatives in curbing abuse (Awasthi (2019). De Maio et al. (2019), Merkel (2021), Sutherland-Smith (2011)). A part of the drawback was identified as students not paying attention to the relevant policies (Brown and Howell (2001), Gullifer and Tyson (2014)). However, the recent technological developments have exacerbated the problem not only via the broad range of applications and quick delivery, but also via

promising relative anonymity and providing global reach. Stated differently, recent technological developments have lowered the transaction costs of exchanging information, keeping in mind that some of that information may be legitimate and the lower transaction costs might also apply to the enforcers of the rules.

The key research question that we address in this research is the following:

Are the internet resources facilitating academic cheating/plagiarism positively or negatively associated with internet resources that are aimed at preventing such behavior?

Specifically, this paper analyzes whether internet-based awareness about anti-cheating and anti-plagiarism policies are associated with reductions in the prevalence of information that facilitates cheating/plagiarism.[1] For this purpose, we create unique indices of internet searches of the costs and benefits/resources of term-paper writing aids across states in the United States. Is the information and resources about academic assistance on the web encouraged or dissuaded by the information about controlling plagiarism and cheating?

While the vast scope and the global nature of the internet, coupled with its multidimensional abilities, prevent a precise accounting of all the potential resources that might be available to someone looking to cheat or abuse the system, this research can be seen as among the first formal attempts to quantify the costs and benefits of potential plagiarism. A better understanding of the underlying incentives would aid in the formulation of more effective policies to maintain academic integrity and reward true talent.

While our focus is on the "demand side" of plagiarism resources for students, it is also important to recognize a "supply side" whereby the scope and magnitude of anti-plagiarism/anti-cheating policies put in place can be expected to be influenced by student use of plagiarism resources. With the advance of ChatGPT technologies, for example, many faculty and university administrators took proactive policy steps to educate students about the proper use of such technologies. These "supply-side" considerations can also be expected to lead to a positive association between plagiarism resources and policies that are considered in this paper. A full structural model, which would formally model the supply side of countermeasures to control student cheating, is beyond the scope of this initial inquiry into the effectiveness of college anti-plagiarism cheating policies.

---

[1] See, for examples, https://www.wired.com/story/how-to-tackle-ai-and-cheating-in-schools-classroom/; https://ohiocapitaljournal.com/2023/09/25/nearly-a-third-of-college-students-used-chatgpt-last-year-according-to-survey/; https://www.lifewire.com/how-this-college-plagiarism-case-shows-catching-ai-cheaters-is-tricky-7500954; https://thehill.com/opinion/education/4162766-ai-cheating-has-hopelessly-irreparably-corrupted-us-higher-education/; https://fortune.com/2023/08/10/chatpgt-cheating-plagarism-college-professors-full-on-crisis-mode/

The structure of the rest of the paper includes a discussion of the unique internet-based data generation on the costs and benefits of plagiarism in the next section, followed by the models, data, estimation, results, and conclusions.

2. Prevalence of internet information on the costs and benefits of plagiarism/copying

At the heart of the unique contribution of this work is the internet-based searches of the costs and benefits of cheating/plagiarism. As data on the true level of plagiarism is hard (or impossible) to come by, we generate data using the internet. This data generation strategy makes even more sense when we consider the internet-based nature of artificial intelligence-related technologies (e.g., ChatGPT).[2]

However, several fine search issues that to be kept in mind to ensure comparability. For example, the searches have to be done in one sitting (i.e., in the shortest possible time), given the tendency of internet search results to change momentarily.[3] Further, in some instances, contamination from other, similar sounding, terms should be avoided as much as possible. In particular, in the case of the United States, searches for the state of Virginia were conducted to avoid references to the state of West Virginia, and those with respect to the state of Washington should avoid references to Washington, DC, or the District of Columbia.

In particular, we conducted three sets of Google internet searches: [4, 5]

Search A: (*ResourcesINTERNET*): keywords: "state name help write term paper OR project"

Search B: (*Policy1*): keywords: "state name college OR university plagiarism OR cheating policy"

Search C: (*Policy2*): keywords: "state name college OR university plagiarism OR cheating policy syllabus"

Search A captures the internet resources on the potential internet resources/benefits of plagiarism or copying, while searches B and C capture information about the potential costs

---

[2] The overall internet search strategy draws on the seminal related work of Goel et al. (2012) who used internet searches to determine cross country corruption awareness.

[3] Our searches were all conducted in one afternoon in November 2023. One should also bear in mind that the search algorithm used by Google differs from some other search platforms (see Goel et al. (2012)).

[4] Placing the search keywords in quotation marks cuts down noise in the results by ensuring that the key terms appear in every internet search result – for background, see https://www.theguardian.com/technology/2016/jan/15/how-to-use-search-like-a-pro-10-tips-and-tricks-for-google-and-beyond.

[5] The choice of the search strings is admittedly somewhat arbitrary. Without specific guidance from the literature, we relied on intuition regarding what a typical demander (student) or supplier (firms, websites) in this context would likely use to search for related information. We did try some alternative variations, and settled upon the strings used after examining the top search results (see the Appendix).

(with the difference between B and C lying in the use of the keyword "syllabus").[6] Some examples of the search results for these searches in Google are provided in the Appendix. The correlation between *Policy1* and *Policy2* is 0.73 (Table 2), suggesting that the searches are picking up similar but not identical aspects of anti-plagiarism/cheating policies. We turn next to a discussion of the formal model, the data employed, and the estimation strategy used.

3. The model, data, and estimation

3.1 The model

One should bear in mind that internet-based information (both on the resources to facilitate and to check cheating/plagiarism) is only one source of information on the phenomena of interest in the paper, although it is perhaps widely used in the current age of information flows. However, the internet-based nature of AI makes the use of digital technologies especially relevant. This coupled with the borderless and relatively anonymous nature of the internet drives the quantity and quality of information about the potential costs and benefits of resources related to plagiarism and cheating.

The relationship between the internet information facilitating cheating/plagiarism (*ResourcesINTERNET)* and information about the policies to check abuse (Policies) could well be negative (Deterrence effect) or positive (Facilitating effect).

The intuition for the deterrence effect is rather straightforward – greater and/or more stringent policies against cheating/plagiarism at colleges and universities would dissuade the generation and transmission of information facilitating such behavior. [7]

---

[6] It is possible that some commercial enterprises (e.g., advertising agencies, journalists etc.) might be using the internet writing aids to assist in their deliverables or increase productivity (Goel and Hsieh (2019)). In those instances, maintaining academic integrity would be less of a concern.

[7] For instance, the first internet search hit result for the state of Alabama, for *Policy1*, related to the plagiarism policy of the University of North Alabama - https://www.una.edu/english/plagiarism-policies.html; while the first result for the same state for *Policy2* related to the syllabus from the same university for the course COM 368 - https://www.una.edu/education/docs-syllabi/syllabus_COM368.pdf.

On the other hand, the Benefit internet search as defines in Table 1, for Alabama yield a number of commercial resources at the top, with the top 6 being (in order of appearance):
1. "Custom Essay Writing Services for US Community Colleges"; https://www.thepaperexperts.com/campus-custom-essay-writing/
2. "Exceptional Academic Writing Services | Academized - 15% OFF"; https://academized.com/
3. "Writing a Paper (for Teens) - Children's Health System"; https://kidshealth.org/ChildrensAlabama/en/teens/writing-papers.html
4. "Write My Term Paper for Me"; https://essays-panda.com/write-my-term-paper-for-me
5. "Essay Writer Expert - Essay Writing Loft"; https://www.linkedin.com/in/essaywritingloft/
6. "Writing Essentials Grade Booster Pack Study Aid"; https://www.universitysupplystore.com/shop_product_detail.asp?catalog_group_id=Mg&catalog_group_name=R2VuZXJhbCBCb29rcw&catalog_id=628&catalog_name=U3R1ZHkgQWlkcyAmIEJhcmNoYXQ0cw&pf_id=63381&produc

On the other hand, the Facilitating effect, involving a positive impact of policies on information about plagiarism resources, could exist due to several factors. One, there may be time lags between the recognition of information on policies and the generation of information facilitating cheating. This would, at the very least, undermine the negative effect and might result in no discernible relation between policies and outcomes. Two, the potential perpetrators might be focused on getting their term papers written in a hurry to meet deadlines, while not paying attention to related policies. Three, the global nature of the internet and the lack of effective teeth with anti-cheating policies would render them relatively less useful. Finally, many internet resources might contain fake information, promising more (illegal) stuff upfront but delivering legal resources in effect.

Based on the above discussion, we formulate our main hypothesis that we will test with the data, generated partly via the method discussed in Section 2.

H1: Greater awareness of the potential costs of plagiarism should dissuade plagiarism opportunities, ceteris paribus.

The broader theoretical underpinnings of our hypothesis can be seen as nested in the seminal work of Becker on crime and punishment (1968), with the economic aspects of plagiarism discussed by Collins et al. (2007); also see Sutherland-Smith (2011).

With individual observations at the state level, the general form of our estimated equation to test hypothesis H1 takes the following form:

Plagiarism resources or benefits (*ResourcesINTERNET*) = f(Plagiarism policy$_{1,2}$, Z, Unemployment (*UNEMP*), Diversity (*DIVERSITY*), *IVYleague*, *POLICE*, *CANADA*, *MEXICO*)

Z = *INCOME*, enrollment (*EDU*), Urbanization(*URBAN*), Population (*POP*), gender ratio (*GENDER*)

The dependent variable, *ResourcesINTERNET*, captures the resources available on the internet in a state about assistance with term papers or course projects. As discussed in the previous section, we recorded the number of Google search hits with "state name help write term paper OR project" as keywords (see normalized details below). This information identifies potential resources for a student looking for assistance (legitimate or otherwise) in writing term papers or projects. Furthermore, the information generated by the web searches need not necessarily

---

t_name=V3JpdGluZyBFc3NlbnRpYWxzIEdyYWRlIEJvb3N0ZXIgUGFjaBTdHVkeSBBBaWQ&type=3&target=shop_product_list.asp

Thus, we notice that the Benefits search results are more general, with results offering legal writing aids, and the academic institution appearing at number 6 (this order, of course, would vary across states).
This variation in the results policy for Policy and Benefits would explain why the relation between the two could very well be positive or negative.

pertain to online assistance – there could be physical locations offering assistance that are advertising on the internet.[8]

On the costs side, we ran two different internet searches about the potential costs of plagiarism (via be-based policies/information against copying or plagiarism), with the difference being the keyword "syllabus" to identify whether anti-plagiarism/copying policies were listed on course syllabi. Specifically, the respective internet search strings were: *Policy1*: "state name college OR university plagiarism OR cheating policy"; and *Policy2*: "state name college OR university plagiarism OR cheating policy syllabus". While many students might not pay attention to these policies (Brown and Howell (2001), Gullifer and Tyson (2014)), we would expect that overall they might act as a deterrent. Conversely, however, more (or more stringent) anti-plagiarism policies might invite greater internet activity to suggest/offer ways to bypass these policies. This is especially possible due to the borderless and relatively anonymous nature of the internet. We also consider the presence of law enforcement in a state as a potential deterrent and that is discussed further below.

The costs and benefits are normalized by internet users.[9] Besides providing a useful robustness check of our findings, the normalization by internet users enables us to account for the digital divide issues across states (Asmar et al. (2022)).

The average for *ResourcesINTERNET* is greater than the policy searches, with *Policy1* averaging more than double that for *Policy2* (see Table 1). This makes sense since not all course syllabi have explicit policies related to plagiarism/cheating (and some of the course syllabi might not be posted on the web).

A set of Z controls is included in all models to account for socio-economic aspects. Zillien and Hargittaii (2019) have noted that internet usage could vary across groups. INCOME captures economic prosperity, associated with the affordability and with the opportunity costs of breaking the rules, higher education enrollment (per capita) addresses market size and competition, as does population, while the urbanization rates account for greater access to resources in the urban areas. The urbanization rate in our sample stood at 72.4 percent, while the unemployment rate was 3.4 percent (Table 1). Finally, gender differences are controlled for by including the gender ratio across states. In our sample, the mean urbanization rate was about 72 percent, while the average gender ratio (men to women) was 97.8.

Relatedly, some models also control for state-level differences in diversity (*DIVERSITY*) and the unemployment rate (*UNEMP*). The unemployment rate would capture the costs and benefits related to breaking the law (in this case plagiarism).

---

[8] Note that there is the possibility of some small duplication whereby information posted by an entity is repeated on mirror sites.

[9] We also tried normalizing alternately by population. The main results were qualitatively similar and are not reported here. Details are available upon request.

Although police are generally not involved directly in monitoring/catching academic plagiarism, we include *POLICE* as a control for the overall indication of enforcement resources in a state. Furthermore, as a control for education quality in a state, *IVYleague* identifies the seven states that house the Ivy League institutions of higher education. The underlying logic is that these states would likely have more vigilant anti-plagiarism policies, ceteris paribus, as the Ivy League schools try to maintain their reputations and other institutions in these states are more vigilant to attract quality students.

Finally, *CANADA* and *MEXICO*, are dummy variables identifying states bordering Canada and Mexico, respectively, to account for issues related to language (Spanish in the case of Mexico and French for some states bordering Canada) and transient populations.

3.2 Data

The internet search data for the key variables in the analysis were described above. The top of Table 1 presents summary statistics of the results. Regarding benefits, the mean number of internet hits per estimated internet user across the 50 states (*ResourcesINTERNET*) stood at 1.21 and ranged from a low of 0.13 (Virginia) to a high of 5.35 (Alaska). On the cost side, the mean values in the data set were 0.08 (*Policy1*) and 0.03 (*Policy2*) with a low-high range of 0.002 (California) to 0.616 (Alaska) and 0.0005 (California) to 0.235 (Alaska), respectively.

Table 2 summarizes the correlation matrix for these key indicators. Results show that the correlation between the *ResourcesINTERNET* variable and the two cost measures was in the 0.70 to 0.76 range. The correlation between *Policy1* and *Policy2* stood at 0.73, indicating that these two measures offer a somewhat different perspective on the costs of plagiarism.

The other variables are from reputed sources that are routinely used in studies based on United States data. Details about these variables, including definitions, summary statistics, and data sources are provided in Table 1.

3.3 Estimation

We employ Ordinary Least Squares as our estimation technique to test hypothesis H1 from different variations of equation (1) outlined above. This estimation strategy makes sense given the cross-sectional nature of the underlying data sample. The statistical significance of individual estimated coefficients is based upon robust standard errors, while the overall strength of the estimated models is given by R^2's and F-values.

As mentioned above, the estimation results may be viewed as primarily establishing correlations – it is hard to determine the direction of causality between internet resources and policies about plagiarism/cheating due to the demand and supply influences and the lags involved. Yet, this study provides formal evidence on an issue of significant emerging importance.

4. Results

Goel-Nelson WP (AI and plagiarism)

4.1 Baseline models

The results from the baseline models, using variations of equation (1), are presented in Table 3.

The results from the baseline models, using variations of equation (1), are presented in Table 3. Models 1.1 and 1.2 summarize the results using the *Policy1* anti-plagiarism search variable while Models 1.3 and 1.4 present estimates for similar models using the *Policy2* search (includes syllabus) variable. In all four models, the estimated parameter estimate on the policy variable is positive and significant at the 5% level or better. Thus, our hypothesis H1 above is not supported. These findings are consistent with the notion that supply creates demand in the context of internet resources available for cheating/plagiarism.

Quantitatively, a one unit increase in *Policy1* is associated with 3.7 units increase in *ResourcesINTERNET* (Model 1.1), while a corresponding increase in *Policy2* leads to a 9.8 unit increase in 9.8 units increase in *ResourcesINTERNET* (Model 1.3). The corresponding elasticities are reported in the concluding section.

In other results, state population has a consistently negative and statistically negative association with internet searches for term papers/projects in all four models considered, suggesting that such search activity does not keep up with market size when measured by state population. There is some evidence that state per capita income is positively associated with internet search activity for term papers, but the variable fails to gain statistical significance at conventional levels in any of the models considered in Table 3. This was also the case for *UNEMP*.

Further, none of the other socio-economic control variables (*EDU*, *URBAN*, *DIVERSITY*, *GENDER*) consistently achieves statistical significance across the models considered.   In other words, these attributes do not significantly figure in the internet information/awareness related to resources for term papers.

4.2 Additional considerations

We enhance the set of controls considered in the baseline models by including indicators of state-level diversity (*DIVERSITY*), states with IVY League schools (*IVYleague*), and geographic controls (i.e., states bordering Canada or Mexico). The related results are reported in Table 4.

The results with *Policy1* and *Policy2* remain positive and significant, as shown in Table 3. There is some statistical support for the positive influence of non-English speakers (*nonENGLISH*) and mixed support for the border-state *Mexico* variable. On the other hand, the coefficients on *POLICE*, *CANADA*, and *IVYleague* were statistically insignificant. The insignificance of *POLICE* makes sense because most police are not tasked with fighting plagiarism.

4.3 Robustness check: Checking for the influence of outliers

It is possible that our results may be unduly influenced by outlying values, whereby some states might have either unusually low or high internet search activity. Accordingly, the baseline

models are rerun in Table 5 by excluding the one state with the highest and the lowest search results (normalized as above).

When the highest and the lowest values on benefits and costs are removed in Table 5, the findings with *Policy1* are qualitatively similar to what is reported above, both in terms of the estimated magnitude of the parameter estimate for that variable and its statistical significance in the models where it is used. In contrast, the results using *Policy2* in the model setup are still positive, but statistically weaker than what was reported above for that policy measure. This may have to do with many students not paying attention to the fine details in the syllabi.

The concluding section follows.

5. Conclusions

The advent of the internet (Goffe and Sosin (2005)), and more recently of artificial intelligence (AI), has challenged academic and other institutions to ensure ethical practices and reward/promote true merit. The borderless and relatively anonymous nature of the internet creates policing challenges, leading to the abuse of established rules and standards. The increased globalization is facilitated by the development of internet access in recent times. In the context of academia, this impacts the size and scope of resources to facilitate/check plagiarism and cheating, both from the demand and supply sides. In fact, a number of sites offering term paper help were indeed housed outside of the United States.

Adding some formal insights into the current topic of fundamental importance to maintaining academic integrity, this paper examines the association of anti-plagiarism/anti-cheating policies with resources that facilitate such behavior (legal or otherwise).

Using unique internet search indices of the policies and resources, we find that the two are positively associated – the associated resources ratchet up with the policies. This association is robust to different modeling formulations, including considerations of course syllabi. When sensitivity to outlying values of internet search results is tested (Table 5), it turns out that the policy search results without the term "syllabus" (*Policy1*) show a relatively more robust association with available internet resources.

Turning to the research question posed in the Introduction, we find that the internet resources facilitating academic cheating/plagiarism are positively associated with internet resources that are aimed at preventing such behavior. Thus, we are unable to find support for our main hypothesis that internet-based information on anti-plagiarism/anti-cheating policies is effective in undermining related information that promotes such behavior. This could be due to a lack of cognition or other communication channels (verbal, print, etc.) being relatively more effective.

In terms of relative magnitudes, the elasticities of *ResourcesINTERNET* with respect to *Policy1* (Model 1.1) and *Policy2* (Model 1.3) are quite similar (0.3 and 0.2, respectively – all elasticities

evaluated at the respective means (Table 1).   At a broader level, this speaks to the different channels of information flows and the diffusion of information.

The findings reinforce the view that policies to check plagiarism and cheating are likely to lack teeth and may be a step behind the resources that facilitate unethical behavior. An interesting result is that discussions of plagiarism/cheating on course syllabi have no different correlates than other instances. This may be attributed to a lack of cognition or the lag between when syllabi are first introduced at the (beginning of the course) and when possible cheating/plagiarism decisions might be made (later in the course when assignments and term papers are due).

Overall, this paper can be seen as making initial formal forays into the investigations of possible cheating and plagiarism in academic settings in the age of AI. As technologies evolve, impacting both the demand and supply of internet resources and facilitating research measures, additional insights may be gained over time. One interesting and relevant avenue in due course would be to distinguish the impacts of general AI from those of generative AI tools.

References

Asmar, A., Mariën, I., Van Audenhove, L., 2022. No one-size-fits-all! Eight profiles of digital inequalities for customized inclusion strategies, New Media & Society, 24(2), 279-310. DOI: 10.1177/14614448211063182.

Awasthi, S., 2019. Plagiarism and academic misconduct: A systematic review. DESIDOC Journal of Library & Information Technology, 39(2), 94-100, DOI: 10.14429/djlit.39.2.13622

Becker, G.S., 1968. Crime and punishment: An economic approach. Journal of Political Economy, 76(2), 169-217.

Brown, V.J., Howell, M.E., 2001. The efficacy of policy statements on plagiarism: Do they change students' views? Research in Higher Education, 42, 103-118.

Collins, A., Judge, G., Rickman, N., 2007. On the economics of plagiarism. European Journal of Law and Economics, 24, 93-107.

Cotton, D.R.E., Cotton, P.A., Shipway, J.R., 2023. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, in press, https://doi.org/10.1080/14703297.2023.2190148

de Maio, C., Dixon, K., Yeo, S., 2019. Academic staff responses to student plagiarism in universities: A literature review from 1990 to 2019. Issues in Educational Research, 29(4), 1131-1142.

Goel, R.K., Hsieh, E.W.-T., 2002. Internet growth and economic theory. Netnomics, 4(2), 221-225.

Goel, R.K., Hsieh, E.W.-T., 2019. Internet use and labor productivity growth: Recent evidence from the U.S. and other OECD countries. Netnomics, 20(2), 195-210.

Goel, R.K., Nelson, M.A., Naretta, M.A., 2012. The internet as an indicator of corruption awareness. European Journal of Political Economy, 28(1), 64-75.

Goffe, W.L., Sosin, K., 2005. Teaching with technology: May you live in interesting times. The Journal of Economic Education, 36(3), 278-291.

Gullifer, J.M., Tyson, G.A., 2014. Who has read the policy on plagiarism? Unpacking students' understanding of plagiarism. Studies in Higher Education, 39(7), 1202-1218.

Kaposi, D., Dell, P., 2012. Discourses of plagiarism: Moralist, proceduralist, developmental and inter-textual approaches. British Journal of Sociology of Education, 33(6), 813-830.

Lo, C.K., 2023. What Is the Impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), 410, https://doi.org/10.3390/educsci13040410

McGrail, E., McGrail, J.P., 2015. Exploring web-based university policy statements on plagiarism by research-intensive higher education institutions. Journal of Academic Ethics, 13, 167-196, https://doi.org/10.1007/s10805-015-9229-3

Merkel, W., 2021. Collage of confusion: An analysis of one university's multiple plagiarism policies. System, 96, 102399

Padillah, R., 2023. Ghostwriting: a reflection of academic dishonesty in the artificial intelligence era. Journal of Public Health, fdad169, https://doi.org/10.1093/pubmed/fdad169

Park, C., 2003. In Other (People's) Words: Plagiarism by university students--literature and lessons. Assessment & Evaluation in Higher Education, 28(5), 471-488.

Pecorari, D., 2015. Plagiarism in second language writing: Is it time to close the case? Journal of Second Language Writing, 30, 94-99, https://doi.org/10.1016/j.jslw.2015.08.003.

Sutherland-Smith, W., 2011. Crime and punishment: An analysis of university plagiarism policies. Semiotica, 187, 127-139, https://doi.org/10.1515/semi.2011.067

Tsertsvadze, V., Khurtsia, L., 2020. Education policy and effectiveness of the plagiarism detection system. Globalization and Business, 9, 151-155, https://doi.org/10.35945/gb.2020.09.018

Walker, M., Townley, C., 2012. Contract cheating: a new challenge for academic honesty? Journal of Academic Ethics, 10, 27-44.

Zillien, N., Hargittaii, E., 2019. Digital distinction: Status-specific types of internet usage. Social Science Quarterly, 90(2), 274-291.

Table 1
Variable definitions, summary statistics, and data sources

Variable	Mean (std. dev.)	Min./Max.	Source
Number of Google search hits by state for the query "*state name*" *help write "term paper OR project"* per estimated internet users in a state, November 21, 2023. [*ResourcesINTERNET*]	1.21 (1.4)	0.13 5.35	[1]
Number of Google search hits by state for the query "*state name*" "*college OR university*" "*plagiarism OR cheating policy*" per estimated internet users in a state, November 21, 2023. [*Policy1*]	0.08 (0.2)	0.00 0.62	[1]
Number of Google search hits by state for the query "*state name*" "*college OR university*" "*plagiarism OR cheating policy*" "*syllabus*" per estimated internet users in a state, November 21, 2023. [*Policy2*]	0.03 (0.1)	0.00 0.24	[1]
State per capita personal income (in thousands), 2022. [*INCOME*]	63.2 (8.6)	46.4 84.6	[2]
State population (in millions), 2022. [*POP*]	6.65 (7.5)	0.58 39.0	[2]
Total fall enrollment in degree-granting postsecondary institutions (percent of state population measured in thousands), 2020. [*EDU*]	58.23 (17.4)	30.1 135.1	[3]
Urban population (percent), 2020. [*URBAN*]	72.4 (14.8)	35.1 94.2	[4]
Gender ratio – ratio of men to women (100 = parity), 2021. [*GENDER*]	97.8 (3.2)	93.6 109.2	[5]
Unemployment rate, 2022 annual average (percent) [*UNEMP*]	3.4 (0.8)	2.1 5.4	[6]
State Diversity Index: The probability that two people chosen at random will be from different race and ethnic groups, 2020. [*DIVERSITY*]	49.0 (14.5)	18.5 76.0	[7]
Police and Sheriff's Patrol officers (percent of state population, measured in thousands, 2022. [*POLICE*]	1.89 (0.4)	1.2 2.9	[8]
Percentage of people 5 years and over who spoke a language other than English at home, 2019. [*nonENGLISH*]	15.2 (10.1)	2.6 44.5	[9]
State where one or more of the eight Ivy League schools are located (1 = yes, 0 = no). [*IVYleague*]	0.14 (0.35)	0.0 1.0	[10]
State has border with Mexico (1 = yes, 0 = no). [*MEXICO*]	0.08 (0.3)	0.0 1.0	
State has border with Canada (1 = yes, 0 = no). [*CANADA*]	0.26 (0.4)	0.0 1.0	

Notes: N = 50.

Goel-Nelson WP (AI and plagiarism)

Sources:

[1]. Google search hits: Authors' calculations – see text for details. Internet users: Internet penetration in a state as of November 2021, retrieved October 2, 2023, from https://statista-com.ezproxyz.uakron.edu:2443/statistics/184691/internet-usage-in-the us-by-state/.

[2]. U.S. Bureau of Economic Analysis. www.bea.gov. (accessed October 2023)

[3]. National Center for Education Statistics. https://nces.ed.gov/programs/digest/current_tables.asp - Table 304.10. (accessed November 2023)

[4]. https://www.visualcapitalist.com/sp/mapping-us-urbanization-by-state/#:~:text=Not%20surprisingly%2C%20the%20three%20most,%2C%20and%20Vermont%20(35.1%25). (accessed October 2023)

[5]. U.S. Census 2021 ACS 5-Year Survey (Table S01010), drawn from https://worldpopulationreview.com/state-rankings/sex-ratio-by-state. (accessed October 2023)

[6]. U.S. Bureau of Labor Statistics. https://www.bls.gov/lau/lastrk22.htm. (accessed October 2023)

[7]. U.S Census, 2020. https://www.census.gov/library/visualizations/interactive/racial-and-ethnic-diversity-in-the-united-states-2010-and-2020-census.html. (accessed November 2023)

[8]. U.S. Bureau of Labor Statistics. https://data.bls.gov/oes/#/occGeo/One%20occupation%20for%20multiple%20geographical%20areas (accessed November 2023)

[9]. U.S. Census. 2019 American Community Survey, 1-year estimates. https://www.census.gov/content/dam/Census/library/publications/2022/acs/acs-50.pdf. (accessed October 2023)

[10]. https://en.wikipedia.org/wiki/Ivy_League.

Table 2
Correlation of key variables

	ResourcesINTERNET	Policy1	Policy2
ResourcesINTERNET	1.00		
Policy1	0.76	1.00	
Policy2	0.70	0.73	1.00

Notes: See Table 1 for variable definitions.
N=50.

Table 3
Effectiveness of college anti-plagiarism/cheating policies: Baseline models
Dependent variable: _ResourcesINTERNET_

Model →	1.1	1.2	1.3	1.4
Policy1	**3.733**[**]	**3.622**[**]		
	(3.2)	**(3.2)**		
Policy2			**9.827**[**]	**9.763**[**]
			(2.5)	**(2.4)**
Per capita personal income [_INCOME_]	0.014	0.018	0.022	0.027
	(0.8)	(1.1)	(1.3)	(1.6)
Population [_POP_]	-0.035[**]	-0.040[**]	-0.047[**]	-0.053[**]
	(2.7)	(3.0)	(2.5)	(2.9)
Postsecondary enrollment [_EDU_]	-0.002	0.004	-0.011[**]	-0.003
	(0.5)	(0.9)	(3.2)	(0.7)
Urban population [_URBAN_]	-0.020	-0.032[*]	-0.018	-0.033
	(1.1)	(1.7)	(0.9)	(1.6)
Gender ratio [_GENDER_]	0.057	0.068[*]	0.051	0.062
	(1.5)	(1.8)	(1.1)	(1.3)
Unemployment rate [_UNEMP_]		0.165		0.194
		(1.4)		(1.3)
Population diversity [_DIVERSITY_]		0.12		0.016[*]
		(1.2)		(1.7)
Observations	50	50	50	50
F-statistic	12.28[**]	13.35[**]	19.00[**]	34.36[**]
R-squared	0.71	0.73	0.74	0.73

Notes: Variable definitions are provided in Table 1. All models are estimated via ordinary least squares and include a constant term (not reported). The numbers in parentheses are (absolute value) z-statistics based on robust standard errors.

<table>
<tr><td colspan="7" align="center">**Table 4**
Effectiveness of college anti-plagiarism/cheating policies: Extended models

Dependent variable: *ResourcesINTERNET*</td></tr>
</table>

Model →	2.1	2.2	2.3	2.4	2.5	2.6
Policy1	3.786** (3.4)	3.597** (3.2)	3.554** (2.8)			
Policy2				10.510** (2.9)	10.030** (3.1)	10.319** (2.6)
Per capita personal income [*INCOME*]	0.015 (0.8)	0.010 (0.6)	0.012 (0.5)	0.028* (1.8)	0.017 (1.2)	0.031 (1.4)
Population [*POP]*	-0.041** (2.6)	-0.058** (3.2)	-0.045** (2.1)	-0.066** (3.4)	-0.072** (4.0)	-0.071** (3.0)
Postsecondary enrollment [*EDU*]	-0.003 (0.6)	-0.001 (0.2)	-0.004 (0.7)	-0.013** (3.2)	-0.008** (2.5)	-0.012** (3.0)
Urban population [*URBAN*]	-0.019 (1.2)	-0.041* (1.9)	-0.019 (1.1)	-0.018 (1.1)	-0.042* (1.9)	-0.017 (1.0)
Gender ratio [*GENDER*]	0.033 (0.7)	0.047 (1.3)	0.052 (0.9)	0.011 (0.2)	0.035 (0.8)	0.009 (0.1)
Police officers [*POLICE*]	-0.363 (1.2)			-0.269 (0.8)		
Non-English speakers [*nonENGLISH*]		0.051* (1.8)			0.058** (2.7)	
Ivy league schools [*IVYleague*]			0.109 (0.3)			-0.172 (0.3)
Border with Mexico [*MEXICO*]	0.338 (0.9)		0.398 (0.8)	0.935** (2.6)		0.970** (2.4)
Border with Canda [*CANADA*]	0.157 (0.7)		0.194 (0.7)	0.161 (0.7)		0.234 (0.8)
Observations	50	50	50	50	50	50
F-statistic	9.58**	13.64**	7.78**	13.54**	25.64**	13.67**
R-squared	0.73	0.77	0.72	0.68	0.71	0.68

Notes: See Table 3.

Table 5
Effectiveness of college anti-plagiarism/cheating policies: Robustness analysis – removing outliers
Dependent variable: *ResourcesINTERNET*
Model →	**1.1A**	**1.3A**	**1.1B**	**1.3B**
Policy1	**3.310**^{**} **(2.9)**		**3.209**^{**} **(2.9)**	
Policy2		**7.255 (1.6)**		**6.921 (1.5)**
Per capita personal income [*INCOME*]	0.016 (1.0)	0.025 (1.5)	0.013 (0.8)	0.027 (1.6)
Population [*POP]*	-0.036^{**} (2.7)	-0.048^{**} (2.6)	-0.053^{**} (2.4)	-0.049^{**} (2.6)
Postsecondary enrollment [*EDU*]	0.000 (0.1)	-0.007[*] (2.0)	-0.001 (0.2)	-0.009^{**} (2.7)
Urban population [*URBAN*]	-0.020 (1.2)	-0.021 (1.1)	-0.019 (1.1)	-0.020 (1.1)
Gender ratio [*GENDER*]	0.025 (0.6)	0.037 (0.8)	0.020 (0.5)	0.048 (1.0)
Observations	48	48	48	48
Outlier states removed	*ResourcesINTERNET*– VA (low), AK (high)		*Policy1* – CA (low), AK (high)	*Policy2* – WA (low), AK (high)
F-statistic	11.16^{**}	11.86^{**}	11.78^{**}	12.60^{**}
R-squared	0.65	0.54	0.65	0.55

Notes: High and low states for ResourcesINTERNET(Models 1.1A and 1.3A), Policy1 (Model 1.1B), and Policy2 (Model 1.3B) are removed from data set as outliers. See Table 3 for other notes.

Goel-Nelson WP (AI and plagiarism)

APPENDIX

Sample internet search results

Given the relative novelty of our internet search procedure, it seems useful to give the reader a sense of the kind of information that the searches are generating. Accordingly, we present below the top search results with the three searches outlined above, using the state of Kansas as the representative state.

Search A (*ResourcesINTERNET*): keywords: "state name help write term paper OR project"

A1. Sponsored
     Need A Paper Written - Hire Your Personal Nerds 24/7
     Nerdifyit.com
A2. Sponsored
     Do My Paper For Me - 1h Ready
     essaywriter.org
A3. Sponsored
     Write My Term Paper - Experts Across All Fields
     edugenie.net
A4. Sponsored
     Get A Paper Written For Me - Accredited Professionals Only
     savemygrade.com
A5. Term Paper Help Kansas City, Missouri - 99papers.com
     99papers.com
A6. Term Paper Writing services Kansas City, Missouri
     99papers.com
A7. 27 Best Ghostwriters For Hire In Kansas
     Upwork
     https://www.upwork.com › hire › ghostwriters › kansas-us

We notice that several search results here are sponsored by commercial enterprises.

Search B (*Policy1*): keywords: "state name college OR university plagiarism OR cheating policy"

B1. College Readiness Skills and Resources: Plagiarism
     The University of Kansas
     https://guides.lib.ku.edu › c.php
B2. Plagiarism Policy: K-State CC 120 Textbook
     Kansas State University
     https://textbooks.cs.ksu.edu › ... › Course Information
B3. Plagiarism - Newman University
     newmanu.edu

B4. kansas state university faculty perspective, opinions, and
  Kansas State University
  https://www.k-state.edu › documents › hmdiss
B5. School of Law Honor Codes < University of Missouri-Kansas ...
  University of Missouri–Kansas City
  https://catalog.umkc.edu › academic-honesty › law-ho...
B6. Faculty & Student Handbook for KWU Online
  Kansas Wesleyan University
  https://www.kwu.edu › wp-content › uploads

Search C (*Policy2*): keywords: "state name college OR university plagiarism OR cheating policy syllabus"

The search added the keyword "syllabus" to Search B to pick up mentions of plagiarism on course syllabi.

C1. Plagiarism Policy :: K-State CC 120 Textbook
  Kansas State University
  https://textbooks.cs.ksu.edu › ... › Course Information
C2. Plagiarism - Newman University
  newmanu.edu
C3. kansas state university faculty perspective, opinions, and
  Kansas State University
  https://www.k-state.edu › documents › hmdiss
C4. Faculty & Student Handbook for KWU Online
  Kansas Wesleyan University
  https://www.kwu.edu › wp-content › uploads
C5. Central Christian College of Kansas Course Syllabus
  cloudfront.net
C6. Syllabus - FIU Honors College - Florida International University
  Florida International University

Thus, we see from these results that the internet search results are quite pertinent and would inform a potential student about the costs and benefits/resources of plagiarism.