ONLINE APPENDIX OF

GENERATIVE AI USAGE AND ACADEMIC PERFORMANCE

Completed Research Paper

Thirty-Third European Conference on Information Systems (ECIS 2025), Amman, Jordan

Variables	Definition	Rationale	Reference
Exam Score	Continuous measure of student performance in the final exam of the mandatory introductory accounting course indicating the percentage of points a student achieved in the final exam. The minimum is zero and the actual (achievable) maximum is 96.67 (100).	Exam scores reflect students' abilities to solve course-relevant problems, therefore approximating student performance. Consistent with related studies, we use it as the dependent variable.	Voshaar et al., (2024), Voshaar et al., (2023a), Cheng & Ding (2021), Eskew & Faley (1988), Hu et al., (2023), Lento (2018), Massoudi et al., (2017), Perera & Richardson (2010)
GenAI User	Indicator variable depicting student usage of GenAI applications for studying and crafting work that they intentionally should have written on their own. *GenAI User* is estimated by ZeroGPT, a GenAI detection system indicating the percentage of text identified as created by GenAI applications. *GenAI User* equals one if the estimate of ZeroGPT exceeds a threshold of 0.5 and is zero otherwise. In our robustness checks we use alternating thresholds of the ZeroGPT estimation (i.e., 0.4 and 0.6) to distinguish GenAI users from non-users.	Related studies document positive effects (e.g., simplification of complex topics and personalized learning) as well as negative effects (e.g., risk of superficial understanding and loss of independent problem-solving) of using GenAI applications in higher education. These factors can either enhance or impede students' performance. We use *GenAI User* as variable of interest in this study to examine the overall effect of GenAI usage on students' performance, which is unknown at present. We chose ZeroGPT over other GenAI detection due to three reasons. First, related studies rank ZeroGPT among the best GenAI detectors. Second, ZeroGPT is shown to minimize both false positive and false negative classifications texts created by GenAI applications and humans. Third, ZeroGPT is capable if correctly classify texts of German language.	Positive effects of GenAI usage: Fauzi et al., (2023), Gilson et al., (2023), AlAfnan et al., (2023), Pavlik (2023), Engelmann et al., (2023), Calderon et al., (2023), Sallam et al., (2023), Qadir (2023), Lund et al., (2023), Wu et al., (2023), Perkins (2023), Cotton et al., (2023), Ali et al., (2023), Sullivan et al., (2023) Negative effects of GenAI usage: Markauskaite et al., (2022), Eager & Brunton (2023), Jain & Kapoor (2013), Rasul et al., (2023), Crawford et al., (2023a), Sallam et al., (2023), Crawford et al., (2023b), Lund et al., (2023), Milano et al., (2023), Bangert-Drowns et al., (2004), Aremu (2023), Liang et al., (2023), Walters (2023), Weber-Wulff et al., (2023), Yeadon et al., (2024)
A-Level Grade	Continuous measure school performance prior to entering university, ranging from 4.0 (best grade) to 1.0 (worst grade).	*A-Level Grade* is used as proxy for general academic aptitude in this study. Related studies consistently document A-level grades to be predictive for exam performance.	Positive effect on exam performance: Azzali et al., (2023), Eskew & Faley (1988), Lento (2018), Massoudi et al., (2017), Papageorgiou & Halabi (2014), Tan & Laswad (2008)
Attempt	Continuous measure equals the number of times a student has registered for the final exam with a maximum value of five, as student must pass the examination within five semesters (i.e. five attempts) according to the regular examination regulations.	Students' academic behavior (e.g., the number of attempts) affect performance. We include *Attempt* into our analyses to control for prior academic performance and the level of general experience at university and particular course experience.	Positive effect on exam performance: Voshaar et al., (2023a), Dowling et al., (2003), Perera & Richardson (2010)

Variables	Definition	Rationale	Reference
Attendance	Continuous measure indicating the number of tutorials a given student attended scaled by the total number of tutorials offered to the students. Information on students' attendance was collected by conducting short in-class quizzes on the LMS comprising three questions regarding the specific tutorial contents in each tutorial. To participate in the quizzes, the students had to sign in to the LMS utilizing a QR-code presented to them in the corresponding tutorial.	Related studies document a statistical association between student behavior (e.g., attendance) and their exam performance. The students' attendance reflects their effort and motivation they spend on the given course Byrne & Flood (2008). We therefore include *Attendance* into our analyses to control for characteristics of student behavior.	Positive effect on exam performance: Voshaar et al., (2023a), Aldamen et al., (2015), Cheng & Ding (2021), Massoudi et al., (2017), Romer (1993)
Vocational Training	Indicator variable equal to one if a given student completed vocational training before entering university, and zero otherwise.	Students that have completed a vocational training prior to entering university are more experienced, mature, and have higher knowledge, which positively affects their exam performance. Therefore, we integrate *Vocational Training* into our analyses to control for experience, maturity, and knowledge previously accumulated through completing a vocational training.	Positive effect on exam performance: Voshaar et al., (2023b), Guney (2009), Hartnett et al., (2004)
Voluntary Service	Indicator variable equal to one if a student completed a voluntary service or spent a gap year prior to entering university, and zero otherwise.	Completing a voluntary service or a gap year before entering the university is another source of experience, maturity, and knowledge for students. Voluntary Service reflects higher self-organization skills, supporting students to structure university life and thus improve exam performance. We include Voluntary Service to control for this potential impact.	Positive effect on exam performance: Voshaar et al., (2023b), Guney (2009), Hartnett et al., (2004)

Variables	Definition	Rationale	Reference
Female	Indicator variable equal to one if a student is female, and zero otherwise.	Related studies provide evidence for students' gender affecting their exam performance. The results of these studies are versatile, indicating a positive, negative, or no effect of gender on performance. We control for a potential influence by integrating *Female* into our analyses.	Positive effect on exam performance: Wecks et al., (2023), Aldamen et al., (2015), Gammie et al., (2003), Gracia & Jenkins (2003), Mutchier et al., (1987), Premuroso et al., (2011), Tan & Laswad (2008), Tyson (1989) Negative effect on exam performance: Johansson et al., (2022), Koh & Koh (1999), Massoudi et al., (2017), Tan & Laswad (2008) No significant effect on exam performance: Azzali et al., (2023), Byrne & Flood (2008), Hu et al., (2023), Papageorgiou & Halabi (2014)
LinkedIn User	Indicator variable equal to one if a student has a LinkedIn profile, and zero otherwise.	Having a LinkedIn profile potentially impacts exam performance and is therefore included in our analyses. Related studies document LinkedIn usage being correlated with exam performance. Moreover, (new) social media usage (e.g., LinkedIn usage) can be utilized as operationalization for personal innovativeness, which in turn affects GenAI acceptance among students.	Study documenting an effect on exam performance: Paul et al., (2012) Study documenting an effect on personal innovativeness: Strzelecki (2023) Studies documenting an effect on social media usage: Aldahdouh et al., (2020), Wijesundara & Sun (2018)
Course of Study	Categorial variable equal to one if a given student is registered for business studies, two for economics, three for economics in complementary subject, four for engineering and management - production engineering, five for electrical engineering with management, six for information systems and management, and seven for others, respectively. We integrate Course of Study as fixed effects in our analyses.	Student characteristics vary considerably across different courses of study as each attract certain student sub-populations. We consider that this potentially influences students' exam performance and therefore include Course of Study fixed effects into our analyses.	Studies documenting an effect on exam performance: Cheng & Ding (2021), Duff (2004), Jackson & Cossitt (2015), Tan & Laswad (2008), Tan & Laswad (2015)

Variables	Definition	Rationale	Reference
German Detector	Indicator variable alternatively deployed as *GenAI User* estimated by the detection system developed at the University of Applied Sciences Wedel. *German Detector* equals one if the estimate of Originality.AI exceeds a threshold of 0.1 and is zero otherwise.	The students' essays we use to estimate *GenAI User* are characterized by the morphological and semantic peculiarities of the German language. We therefore repeat the robustness check using an AI detector particularly designed for these aspects.	Tlok et al., (2023)
Originality.AI	Indicator variable alternatively deployed as *GenAI User* estimated by the detection system Originality.AI. *Originality.AI* equals one if the estimate of Originality.AI exceeds a threshold of 0.5 and is zero otherwise.	We ensure the robustness of our findings by using alternative detection tools. We utilize Originality.AI as it is prominent in literature and claims to be multi-language.	Walters (2023)
Manual Computation	Continuous measure indicating GenAI application usage for text generation. *Manual Computation* is extracted from principal component analysis comprising *Adjectives*, *Fog Index*, and *Herdan's C*. We use this GenAI detection as an alternative estimation of *GenAI User* in our robustness checks.	We deploy *Manual Computation* as an alternative measure to estimate *GenAI User* to test the robustness of our main analysis. Systems detecting GenAI applications typically rely on numerous determinants to classify, if a given text is created be GenAI or by humans. Related studies document three relevant categories of determinants for GenAI classification. First, GenAI texts show a higher number of adjectives compared to human texts. Second, based on various readability scores, GenAI texts are less readable than human texts. And finally, GenAI created texts provide higher lexical richness. Creating a legitimate variable that depicts GenAI usage requires to include determinants stemming from these categories.	Gunning (1952), Herdan (1960), Markowitz et al., (2023), Martínez et al., (2024), Muñoz-Ortiz et al., (2023), Shah et al., (2023), Deveci et al., (2023), Pehlivanoğlu et al., (2023)
Adjectives	Continuous measure indicating the number of adjectives scaled by the total number of words of a text.	We regard *Adjectives* as a determinant for GenAI usage to create texts and include it into the principal component analysis to extract *Manual Computation*. Related studied characterize GenAI texts to encompass more adjectives. By scaling the absolute number of adjectives by the total number of words of a given text, we avoid a biased measure due to the text length.	Markowitz et al., (2023)

Variables	Definition	Rationale	Reference
Fog Index	Gunning (1952) readability index, calculated as: $\left(\frac{Total\ Words}{Total\ Sentences} + 100 * \frac{n_{wsy \geq 3}}{Total\ Words}\right) * 0.4$ where $n_{wsy} \geq 3$ is the number of words with three syllables or more. A higher (lower) *Fog Index* indicates an easier (a more difficult) read.	We include *Fog Index* into the principal component analysis to extract *Manual Computation*. Related studies have shown that GenAI texts are less readable. *Fog Index* is commonly used by numerous studies to approximate text readability.	Gunning (1952)
Herdan's C	Herdan (1960) lexical richness index, calculated as: $\frac{\log(Total\ Unique\ Terms)}{\log(Total\ Words)}$	*Herdan's C* is included into the principal component analysis to extract Detection Factor. Related studies provide evidence on lexical richness to be predictive for the use of GenAI applications to generate texts. We deploy *Herdan's C* to approximate lexical diversity and richness.	Herdan (1960)

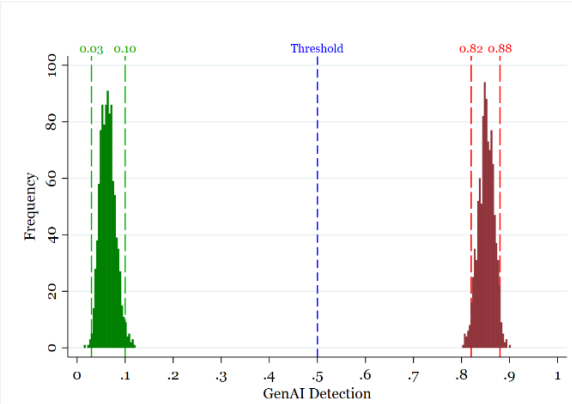
Appendix A. Variable Definitions.

Panel A: Descriptive Results of GenAI Detector Validation

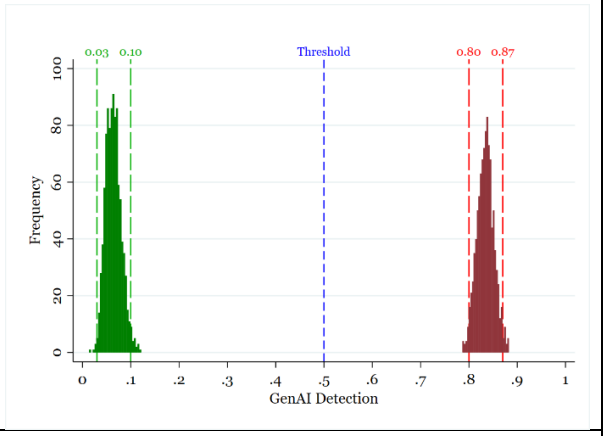
Variable	*N*	*Mean*	*Median*	*SD*	*P25*	*P75*	*Min*	*Max*
GenAI Detection Pre AI	30	0.0643	0.0364	0.0894	0.0000	0.0790	0.0000	0.3311
GenAI Detection AI Gen	30	0.8518	0.8629	0.0865	0.7739	0.9307	0.6689	0.9759
GenAI Detection AI Mod	30	0.8351	0.8548	0.0975	0.7482	0.8935	0.6420	0.9933

Panel B: Distribution of Student vs. GenAI Texts

Pre AI (student) vs. AI-generated Essays



Pre AI (student) vs. AI-modified Essays



Panel C: *t*-test Results of Pre AI (student) vs. GenAI Essays

	Mean	*|Diff.|*
GenAI Detection Pre AI vs. *GenAI Detection AI Gen*	0.0643 0.8518	0.7870*** (-1,074.97)
GenAI Detection Pre AI vs. *GenAI Detection AI Mod*	0.0643 0.8351	0.7715*** (-999.83)

Panel D: Prompts for GenAI Texts

Prompt for AI-generated Essays (English)	You are a student at a German university and are enrolled in the Master's degree program “Business Administration”. You have chosen the specialization Accounting and are taking part in the module “Advanced Seminar: Advanced Problems in Accounting”. The examination of the module is a term paper on the topic “[title of student essay]”. Write a coherent text of about 1000 words for the term paper.
Prompt for AI-generated Essays (German)	Du bist ein Student an einer deutschen Universität und immatrikuliert im Masterstudiengang "Betriebswirtschaftslehre". Du hast den Schwerpunkt Rechnungswesen gewählt und nimmst am Modul "Hauptseminar: Fortgeschrittenenprobleme im Rechnungswesen" teil. Die Prüfungsleistung des Moduls ist eine Hausarbeit zum Thema "[Titel der Hausarbeit]". Verfasse einen zusammenhängenden Text mit einem Umfang von etwa 1000 Worten für die Hausarbeit.
Prompt for AI-modified Essays (English)	I am a student at a German university and enrolled in the Master's degree program “Business Administration”. I have chosen the specialization Accounting and am taking part in the module “Advanced Seminar: Advanced Problems in Accounting”. The module examination is a term paper on the topic “[title of student essay]”. This is my previous draft: “[text of student essay]” Revise and improve my draft. Write a coherent text of about 1000 words.

(Table continued)

Prompt for AI-modified Essays (German)	Ich bin ein Student an einer deutschen Universität und immatrikuliert im Masterstudiengang "Betriebswirtschaftslehre". Ich habe den Schwerpunkt Rechnungswesen gewählt und nehme am Modul "Hauptseminar: Fortgeschrittenenprobleme im Rechnungswesen" teil. Die Prüfungsleistung des Moduls ist eine Hausarbeit zum Thema "[Titel der Hausarbeit]". Dies ist mein bisheriger Entwurf: "[Text der Hausarbeit]" Überarbeite und verbessere meinen Entwurf. Verfasse einen zusammenhängenden Text mit einem Umfang von etwa 1000 Worten.
Appendix B presents the results from the GenAI detector validation study. Variables included in this study were constructed based 30 original student essays written before the availability of GenAI, thus without AI assistance (pre-AI), 30 AI-generated essays using the titles from the pre-AI essays, and 30 AI-modified texts based on the original student essays. ZeroGPT's detector scores for each essay were retrieved, resulting in three variables *GenAI Detection Pre AI*, *GenAI Detection AI Gen, and GenAI Detection AI Mod* representing these distinct sample groups. Panel A provides the descriptive results for the GenAI detections across the three groups. Panel B visualizes the distributions of detections through histograms. The left histogram compares detector values of original pre-AI student essays (green) against AI-generated essays (red), while the right histogram compares the original pre-AI student essays (green) to AI-modified essays (red). Due to the limited sample size, histogram observations were bootstrapped (1,000 replications) for visualization purposes. Green and red dashed vertical lines indicate upper and lower bounds of 95% confidence intervals, calculated from non-bootstrapped detector values. The blue dashed vertical line represents the detector threshold used to classify essays as AI-generated in the main analyses. In Panel C we report two-tailed *t*-test results for differences in mean detector values between original pre-AI student-written essays and their AI-generated and AI-modified counterparts. The *t*-values are presented in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10% level, respectively. Panel D outlines the prompts used to generate both sets of AI-generated and AI-modified essays, presented in both German and English.	

Appendix B. Results of GenAI Detector Validation Study.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	VIF
(1) *GenAI User*	1.000								1.13
(2) *A-Level Grade*	-0.116 *	1.000							1.16
(3) *Attempt*	0.184 **	-0.152 **	1.000						1.19
(4) *Attendance*	-0.072	0.214 ***	-0.331 ***	1.000					1.29
(5) *Vocational Training*	-0.097	0.014	-0.100	0.049	1.000				1.20
(6) *Voluntary Service*	-0.056	0.057	-0.051	-0.066	-0.298 ***	1.000			1.16
(7) *Female*	-0.109	0.041	0.003	-0.001	-0.038	0.173 **	1.000		1.18
(8) *LinkedIn User*	0.048	0.106	0.057	-0.059	-0.038	0.071	0.067	1.000	1.05

Appendix C presents the pairwise Pearson correlations of the variable of interest *GenAI User* and the control variables. ***, **, and * indicate statistical significance at 1%, 5%, and 10% level, respectively. The last column indicates the Variance Inflation Factor (VIF) for each variable. The mean VIF across all Variables (1.16) is below any conventional textbook threshold. All variables are defined in Appendix A.

Appendix C. Correlation Matrix.

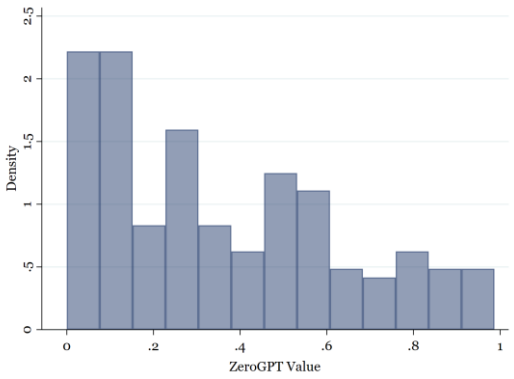
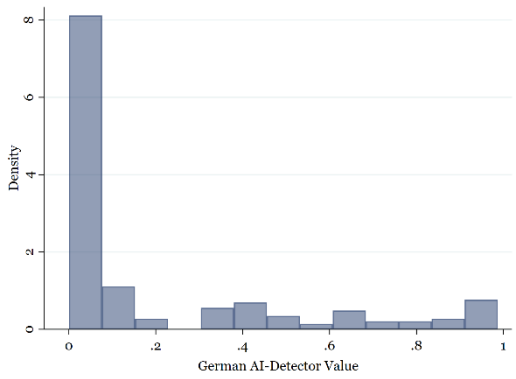
Panel A: Student Data by Essay Submission	*Essay submitter (within sample students)*		*Non-submitter (out of sample students)*		|*Diff.*|
Variables	*N*	*Mean*	*N*	*Mean*	
A-Level Grade	193	2.290	134	2.182	0.108
Attempt	193	1.441	134	2.388	-0.948 ***
Attendance (relative)	193	0.448	134	0.208	0.239 ***
Vocational Training	193	0.135	134	0.135	0.000
Voluntary Service	193	0.362	134	0.403	-0.041
Female	193	0.472	134	0.388	0.084
Panel B: Student Data by GenAI Usage	*GenAI User*		*Non-User*		|*Diff.*|
Variables	*N*	*Mean*	*N*	*Mean*	
A-Level Grade	59	2.184	134	2.337	0.152 *
Attempt	59	1.763	134	1.276	0.486 **
Attendance (relative)	59	0.413	134	0.463	0.051
Vocational Training	59	0.085	134	0.157	0.072
Voluntary Service	59	0.322	134	0.381	0.059
Female	59	0.390	134	0.508	0.118
LinkedIn User	59	0.221	134	0.179	0.041

Appendix D shows the results of two randomization tests two analyse potential sample selection biases. Panel A, shows students characteristics of essay submitting students (in our final sample) against those students, who do not submit an essay and are thus omitted in our further analyses. In Panel B, student characteristics disaggregated by GenAI usage. The last column presents the difference in mean values and the significance level of a two-tailed *t*-test (chi-squared test) for continuous (binary) variables. ***, **, * indicate statistical significance at the 1 %, 5 %, and 10 % level, respectively. Note that we cannot test for a difference in LinkedIn usage in Panel A as we only collected LinkedIn usage data for essay submitting students.

Appendix D. Randomization Test.

Students' GenAI usage	N	Yes	No
GenAI usage for general purposes within the course	30	8 (0.267)	22 (0.734)
GenAI usage for academic essay-writing within the course	30	9 (0.300)	21 (0.700)
Appendix E presents the student survey results on their GenAI usage.			

Appendix E. Survey Results.

ZeroGPT	German AI Detecot (FH Wedel)
	

Appendix F presents histogram plots for the AI Detector values of our main (ZeroGPT) as well as a tool specialized for German language (German AI Detector from FH Wedel).

Appendix F. Distribution of AI Detector

Variables	(1) Threshold 0.4	(2) Threshold 0.6	(3) Continuous Detection Score	(4) German Detector	(5) Manual Computation	(6) Balanced Sample
GenAI User	**-7.10 ****	**-7.17 ****	**-14.67** ***	**-8.53 *****	**-2.66 *****	**-6.51 ****
	(-2.43)	**(-2.02)**	**(-2.82)**	**(-2.90)**	**(-2.70)**	**(-2.07)**
A-Level Grade	11.64 ***	11.42 ***	11.46 ***	11.21 ***	11.62 ***	8.88 ***
	(4.92)	(4.78)	(4.81)	(4.75)	(4.87)	(3.03)
Attempt	0.89	0.84	0.96	0.94	0.79	2.05 **
	(0.74)	(0.69)	(0.79)	(0.79)	(0.66)	(1.98)
Attendance (relative)	17.70 ***	18.38 ***	14.85 ***	16.81 ***	15.55 ***	17.75 ***
	(3.75)	(3.87)	(3.78)	(3.58)	(3.25)	(2.95)
Vocational Training	10.37 **	11.25 ***	10.92 **	10.93 **	12.09 ***	8.75
	(2.41)	(2.63)	(2.57)	(2.58)	(2.86)	(1.65)
Voluntary Service	1.81	2.07	2.60	1.57	1.24	1.52
	(0.60)	(0.69)	(0.87)	(0.53)	(0.41)	(0.45)
Female	-10.14 ***	-9.29 ***	-9.40 ***	-10.20 ***	-10.18 ***	-10.49 ***
	(-3.50)	(-3.18)	(-3.24)	(-3.54)	(-3.50)	(-3.22)
LinkedIn User	10.75 ***	9.25 ***	9.34 ***	10.26 ***	7.89 **	10.53 ***
	(3.05)	(2.65)	(6.66)	(2.96)	(2.25)	(2.82)
Constant	Included	Included	Included	Included	Included	Included
Course of Study-FE	Included	Included	Included	Included	Included	Included
N	193	193	193	193	193	193
Adj. R^2	0.31	0.30	0.41	0.32	0.31	0.20

Appendix G presents the results of the robustness checks. In columns (1) and (2), we reduced (> 0.4) or increased (> 0.6) the threshold of the AI detector value to be classified in the *GenAI User* group. In column (3) we include the continuous detection score from ZeroGPT as our independent variable *GenAI User*. Column (4) uses alternative AI detectors. Column (5) includes a manual computed score that represents AI detection. In column (6), we again present our main results but with an entropy-balanced sample. Bold font indicates the variable of interest. ***, **, * indicate statistical significance at the 1 %, 5 %, and 10 % level (two-tailed), respectively. *t*-values are presented in parentheses. All variables are defined in Appendix A (https://tinyurl.com/zjehfa3n).

Appendix G. Results of Robustness Checks

Variables	(1) Higher A-Level Grade	(2) Lower A-Level Grade	(3) Higher Attendance	(4) Lower Attendance
ChatGPT User	-12.27 *******	2.09	-11.90 *******	-2.92
	(-2.73)	(0.46)	(-2.74)	(-0.64)
A-Level Grade	11.96 **	24.13 ***	11.58 ***	12.64 ***
	(2.37)	(3.25)	(3.60)	(3.31)
Attempt	-2.03	1.21	-1.78	2.27
	(-0.87)	(0.87)	(-0.71)	(1.44)
Attendance (relative)	17.18 **	12.07 *	18.62	34.93 *
	(2.24)	(1.89)	(1.65)	(1.83)
Vocational Training	4.41	24.40 ***	10.33 *	11.93 *
	(0.76)	(3.80)	(1.77)	(1.76)
Voluntary Service	-1.17	3.97	-2.80	8.47 *
	(-0.27)	(0.94)	(-0.67)	(1.83)
Female	-9.65 **	-9.44 **	-11.07 ***	-8.22 *
	(-2.25)	(-2.30)	(-2.76)	(-1.71)
LinkedIn User	9.87 **	6.00	16.44 ***	0.81
	(2.12)	(1.12)	(3.37)	(0.15)
Constant	Included	Included	Included	Included
Course of Study-FE	Included	Included	Included	Included
N	103	90	104	89
Adj. R^2	0.27	0.28	0.30	0.12

Appendix H presents the regression results using split samples. In columns (1) and (2), we repeat our main regression analysis on a restricted sample only containing students with above- (below-)median *A-Level Grade*. Columns (3) and (4) present the main regression separately for students with above- and below-median attendance. Bold font indicates the variable of interest. ***, **, * indicate statistical significance at the 1 %, 5 %, and 10 % level (two-tailed), respectively. *t*-values are presented in parentheses. All variables are defined in Appendix A (https://tinyurl.com/zjehfa3n).

Appendix H. Results of Split Sample Regressions

References

AlAfnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). ChatGPT as an Educational tool: Opportunities, Challenges, and Recommendations for Communication, Business Writing, and Composition Courses, *Journal of Artificial Intelligence and Technology* 3(2), 60-68. 10.1177/2158244019899441

Aldahdouh, T., Nokelainen, P., & Korhonen, V. A. (2020). Technology and Social Media Usage in Higher Education: The Influence of Individual Innovativeness, *SAGE Open* 10(1). 10.1177/2158244019899441

Aldamen, H., Al-Esmail, R., & Hollindale, J. (2015). Does Lecture Capturing Impact Student Performance and Attendance in an Introductory Accounting Course?, *Accounting Education* 24(4), 291-317. https://www.doi.org/10.1080/09639284.2015.1043563

Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. Q. (2023). Impact of ChatGPT on Learning Motivation, *Journal of English Studies in Arabia Felix*

Aremu, T. (2023). Unlocking Pandora's Box: Unveiling the Elusive Realm of AI Text Detection, *Available at SSRN 4470719*.

Azzali, S., Mazza, T., & Tibiletti, V. (2023). Student Engagement and Performance: Evidence from the First Wave of COVID-19 in Italy, *Accounting Education* 32(4), 479–500. https://doi.org/10.1080/09639284.2022.2081813

Bangert-Drowns, R., Hurley, M. M., & Wilkinson, B. (2004). The Effects of School-Based Writing-to-Learn Interventions on Academic Achievement: A Meta-Analysis, *Review of Educational Research* 74, 29 - 58

Byrne, M., & Flood, B. (2008). Examining the Relationships among Background Variables and Academic Performance of First-Year Accounting Students at an Irish University, *Journal of Accounting Education* 26(4), 202–212. https://doi.org/10.1016/j.jaccedu.2009.02.001

Calderon, T. G., Gao, L., & Cardoso, R. L. (2023). Generative Artificial Intelligence in the Classroom: A Financial Accounting Experience, in *Advances in Accounting Education: Teaching and Curriculum Innovations*, T.G. Calderon (ed.). Emerald Publishing Limited, 125-144. https://doi.org/10.1108/S1085-462220230000027006

Cheng, P., & Ding, R. (2021). The Effect of Online Review Exercises on Student Course Engagement and Learning Performance: A Case Study of an Introductory Financial Accounting Course at an International Joint Venture University, *Journal of Accounting Education* 54(1), 100699. 10.1016/j.jaccedu.2020.100699

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*

Crawford, J., Cowling, M., & Allen, K.-A. (2023a). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI), *Journal of University Teaching & Learning Practice* 20(3), 02

Crawford, J., Cowling, M., Ashton-Hay, S., Kelder, J.-A., Middleton, R., & Wilson, G. S. (2023b). Artificial intelligence and authorship editor policy: ChatGPT, bard bing AI, and beyond, *Journal of University Teaching & Learning Practice* 20(5), 1

Deveci, C. D., Baker, J. J., Sikander, B., & Rosenberg, J. (2023). A comparison of cover letters written by ChatGPT-4 or humans, *Danish Medical Bulletin (Online)* 70(11)

Dowling, C., Godfrey, J. M., & Gyles, N. (2003). Do Hybrid Flexible Delivery Teaching Methods Improve Accounting Students' Learning Outcomes?, *Accounting Education* 12(4), 373–391. https://doi.org/10.1080/0963928032000154512

Duff, A. (2004). Understanding Academic Performance and Progression of First-Year Accounting and Business Economics Undergraduates: The Role of Approaches to Learning and Prior Academic Achievement, *Accounting Education* 13(4), 409–430. https://doi.org/10.1080/0963928042000306800

Eager, B., & Brunton, R. (2023). Prompting higher education towards AI-augmented teaching and learning practice, *Journal of University Teaching & Learning Practice* 20(5), 02

Engelmann, B., Haak, F., Kreutz, C. K., Khasmakhi, N. N., & Schaer, P. (2023). Text Simplification of Scientific Texts for Non-Expert Readers, *available at arXiv: 2307.03569*.

Eskew, R. K., & Faley, R. H. (1988). Some Determinants of Student Performance in the First College-Level Financial Accounting Course, *The Accounting Review* 63(1), 137–147

Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A. M. A., & Hatta, H. R. (2023). Analysing the Role of ChatGPT in Improving Student Productivity in Higher Education, *Journal on Education*

Gammie, E., Paver, B., Gammie, B., & Duncan, F. (2003). Gender Differences in Accounting Education: An Undergraduate Exploration, *Accounting Education* 12(2), 177–196. https://doi.org/10.1080/0963928032000091765

Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment, *JMIR Medical Education* 9(1), e45312

Gracia, L., & Jenkins, E. (2003). A Quantitative Exploration of Student Performance on an Undergraduate Accounting Programme of Study, *Accounting Education* 12(1), 15–32. https://doi.org/10.1080/0963928032000049375

Guney, Y. (2009). Exogenous and Endogenous Factors Influencing Students' Performance in Undergraduate Accounting Modules, *Accounting Education* 18(1), 51–73. http://doi.org/10.1080/09639280701740142

Gunning, R. (1952). *The Technique of Clear Writing*. McGraw-Hill.

Hartnett, N., Römcke, J., & Yap, C. (2004). Student Performance in Tertiary-Level Accounting: An International Student Focus, *Accounting & Finance* 44(2), 163–185. https://doi.org/10.1111/j.1467-629X.2004.00104.x

Herdan, G. (1960). *Type-Token Mathematics*. The Hague: Mouton.

Hu, Y., Nath, N., Zhu, Y., & Laswad, F. (2023). Accounting Students' Online Engagement, Choice of Course Delivery Format and Their Effects on Academic Performance, *Accounting Education* 32, 1-36. https://doi.org/10.1080/09639284.2023.2254298

Jackson, M., & Cossitt, B. (2015). Is Intelligent Online Tutoring Software Useful in Refreshing Financial Accounting Knowledge?, *Advances in Accounting Education* 16, 1–19. 10.2139/ssrn.1960011

Jain, T., & Kapoor, M. (2013). The Impact of Study Groups and Roommates on Academic Performance, *Review of Economics and Statistics* 97, 44-54. www.jstor.org/stable/43554978

Johansson, E., Kanapathippillai, S., Khan, A., & Dellaportas, S. (2022). Formative Assessment in Accounting: Student Perceptions and Implications of Continuous Assessment, *Accounting Education* In Press, 1–29. 10.1080/09639284.2022.2091411

Koh, M. Y., & Koh, H. C. (1999). The Determinants of Performance in an Accountancy Degree Programme, *Accounting Education* 8(1), 13–29. 10.1080/096392899331017

Lento, C. (2018). Student Usage of Assessment-Based and Self-Study Online Learning Resources in Introductory Accounting, *Issues in Accounting Education* 33(4), 13–31. https://doi.org/10.2308/iace-52252

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT Detectors are Biased Against Non-Native English Writers, *Patterns* 4(7). 10.1016/j.patter.2023.100779

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S. R., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing, *Journal of the Association for Information Science and Technology* 74, 570 - 581. https://doi.org/10.1002/asi.24750

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., De Laat, M., Buckingham Shum, S., Gašević, D., & Siemens, G. (2022). Rethinking the Entwinement Between Artificial Intelligence and Human Learning: What Capabilities Do Learners Need for a World with AI?, *Computers and Education: Artificial Intelligence* 3, 100056. 10.1016/j.caeai.2022.100056

Markowitz, D. M., Hancock, J. T., & Bailenson, J. N. (2023). Linguistic Markers of Inherently False AI Communication and Intentionally False Human Communication: Evidence From Hotel Reviews, *Journal of Language and Social Psychology* 43(1), 63-82. https://doi.org/10.1177/0261927x231200201

Martínez, G., Hernández, J. A., Conde, J., Reviriego, P., & Merino, E. (2024). Beware of Words: Evaluating the Lexical Richness of Conversational Large Language Models, *available at arXiv: 2402.15518*.

Massoudi, D., Koh, S., Hancock, P. J., & Fung, L. (2017). The Effectiveness of Usage of Online Multiple Choice Questions on Student Performance in Introductory Accounting, *Issues in Accounting Education* 32(4), 1–17. https://doi.org/10.2308/iace-51722

Milano, S., McGrane, J. A., & Leonelli, S. (2023). Large Language Models Challenge the Future of Higher Education, *Nature Machine Intelligence* 5(4), 333-334. https://doi.org/10.1038/s42256-023-00644-2

Muñoz-Ortiz, A., Gómez-Rodríguez, C., & Vilares, D. (2023). Contrasting Linguistic Patterns in Human and LLM-Generated Text, *ArXiv*.

Mutchier, J. F., Turner, J. H., & Williams, D. D. (1987). The Performance of Female Versus Male Accounting Students, *Issues in Accounting Education* 2(1)

Papageorgiou, K., & Halabi, A. K. (2014). Factors Contributing toward Student Performance in a Distance Education Accounting Degree, *Meditari Accountancy Research* 22(2), 211–223. 10.1108/MEDAR-08-2013-0032

Paul, J. A., Baker, H. M., & Cochran, J. D. (2012). Effect of online social networking on student academic performance, *Comput. Hum. Behav.* 28, 2117-2127

Pavlik, J. V. (2023). Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education, *Journalism & Mass Communication Educator* 78(1), 84-93. https://doi.org/10.1177/10776958221149577

Pehlivanoğlu, M. K., Syakura, M. A., & Duru, N. (2023). Enhancing Paraphrasing in Chatbots Through Prompt Engineering: A Comparative Study on ChatGPT, Bing, and Bard, *2023 8th International Conference on Computer Science and Engineering (UBMK)*.

Perera, L., & Richardson, P. (2010). Students' Use of Online Academic Resources within a Course Web Site and Its Relationship with Their Course Performance: An Exploratory Study, *Accounting Education* 19(6), 587–600. 10.1080/09639284.2010.529639

Perkins, M. (2023). Academic integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond, *Journal of University Teaching and Learning Practice*

Premuroso, R. F., Tong, L., & Beed, T. K. (2011). Does Using Clickers in the Classroom Matter to Student Performance and Satisfaction When Taking the Introductory Financial Accounting Course?, *Issues in Accounting Education* 26(4), 701–723. 10.2308/iace-50066

Qadir, J. (2023). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education, *2023 IEEE Global Engineering Education Conference (EDUCON)*.

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heathcote, L. (2023). The Role of ChatGPT in Higher Education: Benefits, Challenges, and Future Research Directions, *Journal of Applied Learning and Teaching* 6(1). https://doi.org/10.37074/jalt.2023.6.1.29

Romer, D. (1993). Do Students Go to Class? Should They?, *The Journal of Economic Perspectives* 7(3), 167–174

Sallam, M., Salim, N. A., Barakat, M., & Ala'a, B. (2023). ChatGPT Applications in Medical, Dental, Pharmacy, and Public Health Education: A Descriptive Study Highlighting the Advantages and Limitations, *Narra J* 3(1). https://doi.org/10.52225/narra.v3i1.103

Shah, A., Ranka, P., Dedhia, U., Prasad, S., Muni, S., & Bhowmick, K. (2023). Detecting and Unmasking Ai-Generated Texts Through Explainable Artificial Intelligence Using Stylistic Features, *International Journal of Advanced Computer Science and Applications* 14(10). https://doi.org/10.14569/IJACSA.2023.01410110

Strzelecki, A. (2023). To Use or Not to Use ChatGPT in Higher Education? A Study of Students' Acceptance and Use of Technology, *Interactive Learning Environments* 31, 1-14. https://doi.org/10.1080/10494820.2023.2209881

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning,

Tan, L. M., & Laswad, F. (2008). Impact of Prior Content and Mega-Cognitive Knowledge on Students' Performance in an Introductory Accounting Course, *Pacific Accounting Review* 20(1), 63–74. 10.2139/ssrn.3331279

Tan, L. M., & Laswad, F. (2015). Academic Performance in Introductory Accounting: Do Learning Styles Matter?, *Accounting Education* 24(5), 383–402. 10.1080/09639284.2015.1075315

Tlok, T., Annuth, H., & Pawłowski, M. (2023). Robuste Erkennung von KI-generierten Texten in deutscher Sprache, in: *Master Thesis at Department of Data Science & Artificial Intelligence – FH Wedel*.

Tyson, T. (1989). Grade Performance in Introductory Accounting Courses: Why Female Students Outperform Males, *Issues in Accounting education* 4(1), 153–160

Voshaar, J., Knipp, M., Loy, T., Zimmermann, J., & Johannsen, F. (2023a). The impact of using a mobile app on learning success in accounting education, *Accounting Education* 222-247. https://doi.org/10.1080/09639284.2022.2041057

Voshaar, J., Wecks, J. O., Johannsen, F., Knipp, M., Loy, T., & Zimmermann, J. (2023b). Supporting Students in the Transition to Higher Education: Evidence from a Mobile App in Accounting Education, *ECIS 2023 Proceedings*.

Voshaar, J., Wecks, J. O., Plate, B. J., & Zimmermann, J. (2024). Tackling Professorial Expert Bias: The Role of ChatGPT in Simplifying Financial Accounting Exam Texts, *Issues in Accounting Education*, 1-31. https://doi.org/10.2308/issues-2023-091

Walters, W. H. (2023). The Effectiveness of Software Designed to Detect AI-Generated Writing: A Comparison of 16 AI Text Detectors, *Open Information Science* 7(1), 20220158

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of Detection Tools for AI-Generated Text, *International Journal for Educational Integrity* 19(1), 26. https://doi.org/10.1007/s40979-023-00146-z

Wecks, J. O., Voshaar, J., & Zimmermann, J. (2023). Using Machine Learning to Address Individual Learning Needs in Accounting Education, *SSRN*. https://dx.doi.org/10.2139/ssrn.4648223

Wijesundara, T. R., & Sun, X.-x. (2018). Social Networking Sites Acceptance: The Role of Personal Innovativeness in Information Technology, *International Journal of Business and Management* 13(8). https://doi.org/10.5539/ijbm.v13n8p75

Wu, H., Wang, W., Wan, Y., Jiao, W., & Lyu, M. R. (2023). ChatGPT or Grammarly? Evaluating ChatGPT on Grammatical Error Correction Benchmark, *ArXiv* abs/2303.13648

Yeadon, W., Agra, E., Inyang, O.-o., Mackay, P., & Mizouri, A. (2024). Evaluating AI and Human Authorship Quality in Academic Writing through Physics Essays, *arXiv preprint arXiv:2403.05458*