Three Essays in Accounting and Taxation: Integrating Discipline-Specific Language through Digital Technologies

by

Lukas Schmidt

a Thesis submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy (Ph.D.)
in Business Administration

Approved Dissertation Committee

Andreas Seebeck, Professor of Accounting and Auditing, Constructor University

Adalbert Wilhelm, Professor of Statistics, Constructor University

Armin Varmaz, Professor of Finance, Hochschule Bremen

Date of Defense: 01.07.2026

School of Business, Social and Decision Sciences

Statement on the Use of Artificial Intelligence (AI)

During the preparation of this dissertation, the author used *ChatGPT*, *DeepL Translate*, *DeepL Write*, and *Grammarly* to improve the language and readability, exercising caution. After using these tools, the author reviewed and edited the content as needed, taking full responsibility for the dissertation's content.

Acknowledgements

My time as a Ph.D. student began with a mixture of excitement for what lay ahead and the uncertainty of how I would develop during this time. Looking back, I see this as a formative period, and I would like to express my sincere gratitude to all those who became part of this journey.

First and foremost, I would like to thank my Ph.D. supervisor and co-author, Andreas Seebeck. Thank you for giving me the freedom to grow and develop. Having the opportunity to make my own decisions, and inevitably my own mistakes, has been invaluable, and I am convinced that I learned a great deal through this process.

Furthermore, I would like to thank my colleague, Tomke Hartmann. It has been a pleasure to work with you, and I very much look forward to following your future projects. I would also like to thank all the student assistants who helped me with various tasks. My thanks also go to my co-authors, Marius Weiß, Johannes Voshaar, and Frank Hechtner. In different ways, each of you challenged my work and opened up new perspectives that helped me grow. I would also like to thank Robin Wolter for his feedback on my projects over the past few years and for his insightful comments on the introductory chapter of this dissertation.

I would also like to thank Adalbert Wilhelm and Armin Varmaz for their support throughout my Ph.D. process. I would additionally like to thank Cornelia Gerhards for her consistent support with departmental matters. I am also grateful to many others, too numerous to name individually, whose support is recognized in the acknowledgments to each essay included in this dissertation.

Moreover, I am deeply grateful to my family. You have supported me in different ways, and without that support, I would not have been able to pursue this path in the way I envisioned and ultimately achieved. Thank you for everything.

I would also like to thank my friends for the moments of balance and fun. Special thanks go to Natalie and Nicolas: You have been my safe space for so long. Thank you for that.

Finally, I would like to thank my partner, Tina. You challenge me, ground me, and continuously show me new and smart perspectives I would never have considered on my own. You were there for me during many of the difficult moments of my work, and for that I am deeply grateful. I am very much looking forward to all the chapters that lie ahead of us.

Lukas Schmidt

Danksagungen

Meine Zeit als Doktorand begann mit einer Mischung aus Vorfreude auf das, was vor mir lag, und Ungewissheit darüber, wie ich mich in dieser Zeit entwickeln würde. Rückblickend betrachte ich diese Zeit als prägende Phase, und ich möchte all jenen meinen aufrichtigen Dank aussprechen, die Teil dieser Reise waren.

Zuallererst möchte ich meinem Doktorvater und Co-Autor, Andreas Seebeck, danken. Danke, dass du mir die Freiheit gegeben hast, zu wachsen und mich weiterzuentwickeln. Die Möglichkeit, eigene Entscheidungen zu treffen und dabei zwangsläufig Fehler zu machen, war von großem Wert, und ich bin überzeugt, dass ich durch diesen Prozess sehr viel gelernt habe.

Außerdem möchte ich meiner Kollegin Tomke Hartmann danken. Es war mir eine Freude, mit dir zusammenzuarbeiten, und ich freue mich sehr darauf, deine zukünftigen Projekte zu verfolgen. Ebenso danke ich allen studentischen Hilfskräften, die mich bei verschiedenen Aufgaben unterstützt haben. Mein Dank gilt auch meinen Co-Autoren Marius Weiß, Johannes Voshaar und Frank Hechtner. Jeder von euch hat meine Arbeit auf unterschiedliche Weise hinterfragt und mir neue Perspektiven eröffnet, die mir dabei geholfen haben, mich weiterzuentwickeln. Ich möchte mich auch bei Robin Wolter für sein Feedback zu meinen Projekten in den letzten Jahren und für seine aufschlussreichen Anmerkungen zum Einführungskapitel dieser Dissertation bedanken.

Ich möchte mich auch bei Adalbert Wilhelm und Armin Varmaz für ihre Unterstützung im Promotionsprozess bedanken. Mein Dank gilt zudem Cornelia Gerhards für ihre beständige Unterstützung in allen Angelegenheiten des Lehrstuhls. Ich bin auch vielen anderen dankbar, die hier nicht namentlich genannt werden können, deren Unterstützung jedoch in den Danksagungen zu den einzelnen Aufsätzen dieser Dissertation gewürdigt wird.

Darüber hinaus bin ich meiner Familie zutiefst dankbar. Ihr habt mich auf vielfältige Weise unterstützt, und ohne diese Unterstützung wäre es mir nicht möglich gewesen, diesen Weg so zu beschreiten, wie ich ihn mir vorgestellt und letztlich erreicht habe. Danke für alles.

Ich möchte auch meinen Freunden für die Momente der Ausgeglichenheit und des Spaßes danken. Ein besonderer Dank gilt Natalie und Nicolas: Die enge Verbundenheit mit euch bedeutet mir sehr viel. Danke dafür.

Schließlich möchte ich meiner Partnerin Tina danken. Du forderst mich heraus, gibst mir Halt und zeigst mir immer wieder neue und kluge Perspektiven auf, die ich allein niemals in Betracht gezogen hätte. Du warst in vielen schwierigen Momenten meiner Arbeit für mich da, und dafür bin ich dir zutiefst dankbar. Ich freue mich sehr auf all die Kapitel, die vor uns liegen.

Lukas Schmidt

Table of Contents

Abstract

In recent years, the growing availability and capability of digital technologies have fundamentally transformed how discipline-specific language is acquired, applied, and analyzed. Given the central role of specialized terminology in accounting and taxation, both education and research increasingly require tools that can effectively integrate such discipline-specific language. This dissertation is motivated by the research question of how digital technologies can be used to integrate discipline-specific language into the fields of accounting and taxation.

Against this background, the three essays of this dissertation examine complementary perspectives on this integration. *Essay I* focuses on the acquisition of accounting-specific language in an educational context by analyzing a wiki-based collaborative learning environment in introductory accounting education. *Essay II* shifts to the practical application of tax-specific language, employing dictionary-based text analysis to construct a novel cross-country dataset on corporate environmental tax legislation. *Essay III* advances this perspective by developing and applying a domain-specific large language model (LLM) to analyze qualitative corporate tax disclosures.

The findings highlight the effectiveness of digital technologies across these disciplines. *Essay I* demonstrates that collaborative, wiki-based glossary construction significantly enhances students' academic performance and supports deeper engagement with accounting terminology, while also revealing challenges related to coordination and content reliability. Building on a dataset created using digital technologies and tax-specific language, *Essay II* reveals that the implementation of corporate environmental taxes across countries is not driven by climate risk, but rather by institutional and political factors. These factors include tax competition, governance quality, and female political representation. *Essay III* provides evidence that domain-specific LLMs substantially outperform general-purpose language models and traditional text analysis techniques in analyzing qualitative corporate tax disclosures, enabling more accurate and economically meaningful insights.

Overall, the dissertation illustrates that digital technologies can effectively support both the acquisition and analysis of discipline-specific language in the fields of accounting and taxation. While educational technologies facilitate structured and collaborative learning of terminology, advanced natural language processing methods enable scalable and context-sensitive analysis of specialized language in practice.

The three essays are intended for a broad audience, including academics, educators, and policymakers. *Essay I* offers a replicable framework for integrating collaborative digital tools into accounting education. *Essay II* provides a novel empirical foundation for understanding the determinants of corporate environmental taxation and informs policy debates. *Essay III* contributes methodological guidance for developing specialized LLMs and highlights their potential for future research. Collectively, the dissertation underscores the importance of aligning digital technologies with the unique linguistic characteristics of accounting and taxation.

Zusammenfassung

In den letzten Jahren haben die zunehmende Verfügbarkeit und Leistungsfähigkeit digitaler Technologien die Art und Weise, wie fachspezifische Sprache erworben, angewendet und analysiert wird, grundlegend verändert. Angesichts der zentralen Rolle der Fachterminologie in den Bereichen Rechnungs- und Steuerwesen benötigen Lehrende und Forschende zunehmend Werkzeuge, die diese fachspezifische Sprache effektiv integrieren können. Diese Dissertation untersucht die Forschungsfrage, wie digitale Technologien genutzt werden können, um fachspezifische Sprache in die Bereiche Rechnungs- und Steuerwesen zu integrieren.

Vor diesem Hintergrund untersuchen die drei Aufsätze dieser Dissertation komplementäre Perspektiven dieser Integration. In *Aufsatz I* wird der Erwerb fachspezifischer Sprache im Rechnungswesen im Bildungskontext untersucht. Dazu wird eine Wiki-basierte kollaborative Lernumgebung in der einführenden Lehre des Rechnungswesens analysiert. *Aufsatz II* wendet sich der praktischen Anwendung von fachspezifischer Sprache im Steuerwesen zu und nutzt dabei eine lexikonbasierte Textanalyse, um einen neuartigen länderübergreifenden Datensatz zur Umweltsteuergesetzgebung für Unternehmen zu erstellen. *Aufsatz III* führt diese Perspektive weiter, indem ein domänenspezifisches Large Language Model (LLM) entwickelt und angewendet wird, um qualitative Unternehmenssteuerberichterstattung zu analysieren.

Die Ergebnisse unterstreichen die Wirksamkeit digitaler Technologien in diesen Bereichen. *Aufsatz I* zeigt, dass die gemeinschaftliche Erstellung eines Wiki-basierten Glossars die akademischen Leistungen der Studierenden verbessert und eine intensivere Auseinandersetzung mit der Fachterminologie des Rechnungswesens fördert. Gleichzeitig bestehen Herausforderungen hinsichtlich der Koordination und der inhaltlichen Zuverlässigkeit der Wiki-Beiträge. Aufbauend auf einem Datensatz, der mithilfe digitaler Technologien und steuerfachspezifischer Sprache erstellt wurde, zeigt *Aufsatz II*, dass die Einführung von Umweltsteuern für Unternehmen in verschiedenen Ländern nicht durch Umweltrisiken, sondern vielmehr durch institutionelle und politische Faktoren bestimmt wird. Zu diesen Faktoren zählen der internationale Steuerwettbewerb, die Qualität der Regierung und der Anteil von Frauen in politischen Entscheidungsgremien. *Aufsatz III* liefert Belege dafür, dass domänenspezifische LLMs bei der Analyse qualitativer Angaben zur Unternehmensbesteuerung deutlich besser abschneiden als Allzwecksprachmodelle und traditionelle Textanalysetechniken, was genauere und wirtschaftlich aussagekräftigere Erkenntnisse ermöglicht.

Insgesamt zeigt die Dissertation, dass digitale Technologien sowohl die Erfassung als auch die Analyse fachspezifischer Sprache in den Bereichen Rechnungs- und Steuerwesen effektiv unterstützen können. Während Bildungstechnologien das strukturierte und kollaborative Erlernen von Terminologie erleichtern, ermöglichen fortschrittliche Methoden der natürlichen Sprachverarbeitung eine skalierbare und kontextsensitive Analyse von Fachsprache in der Praxis.

Die drei Aufsätze richten sich an ein breites Publikum, darunter Wissenschaftler, Lehrkräfte und politische Entscheidungsträger:innen. *Aufsatz I* bietet einen reproduzierbaren Rahmen für die Integration kollaborativer digitaler Werkzeuge in die Ausbildung im Bereich des Rechnungswesens. *Aufsatz II* liefert eine neuartige empirische Grundlage für das Verständnis der Determinanten der Implementierung von Umweltsteuern für Unternehmen und generiert zugleich relevante Erkenntnisse für wirtschaftspolitische Debatten. *Aufsatz III* enthält methodische Leitlinien für die Entwicklung spezialisierter LLMs und hebt deren Potenzial für die zukünftige Forschung hervor. Insgesamt unterstreicht die Dissertation, wie wichtig es ist, digitale Technologien auf die besonderen sprachlichen Merkmale des Rechnungs- und Steuerwesens abzustimmen.

List of Tables

List of Figures

I. Introduction

Discipline-specific language[1] refers to the specialized vocabulary, phrases, and communication styles unique to a particular field of study or profession. In accounting and taxation, discipline-specific language arises from a combination of discipline-exclusive terminology, terms that acquire specialized meanings distinct from general usage, and legal jargon reflecting the field's regulatory foundations (Huang et al., 2023; Loughran & McDonald, 2011). For example, the term "allowance" carries different meanings depending on the context. In general usage, it often refers to a regular sum of money given for a specific purpose, such as a student allowance. In accounting, it denotes a financial provision set aside to cover anticipated credit losses from unpaid customer invoices, commonly known as an "allowance for doubtful accounts." In taxation, the term "allowance" refers to a portion of income that is exempt from tax. Consequently, taking this discipline-specific nature into account is important for the various users of this language in accounting and taxation.

For aspiring accounting and tax professionals, *acquiring* the discipline's terminology is crucial before applying it in professional contexts (e.g., Bardien & Lubbe, 2025; Peters et al., 2013; Seebeck & Wolter, 2022). Likewise, accounting and tax researchers must consider these specificities when *analyzing* practitioners' use of the language, for example, in legislative texts or corporate reporting (e.g., Hasan, 2020; Huang et al., 2023; Loughran & McDonald, 2011; Lynch et al., 2025; Swenson et al., 2024).

In this context, digital technologies[2] can help improve the integration of discipline-specific terminology in the process of *acquiring* and *analyzing* it.[3] On the one hand, digital technologies can facilitate the acquisition of discipline-specific language. For example, in introductory accounting education, digital technologies such as digital audience response systems (Premuroso et al., 2011), online learning management systems (Kelly et al., 2023), and mobile learning applications (Voshaar et al., 2023) can support students with the

[1] In this dissertation, the term "discipline-specific" is used interchangeably with "domain-specific" or "specialized," and "language" is used interchangeably with "terminology" or "vocabulary." However, it should be noted that there are valid distinctions between these terms, particularly from a technical perspective. In the context of this dissertation, though, these distinctions are negligible.

[2] Digital technologies are tools and systems that convert real-world information (e.g., text) into digital form, allowing it to be processed, stored, and transmitted by computers (Li et al., 2024).

[3] Please note that this list of processes is not exhaustive. The processes mentioned represent the foci of this dissertation.

acquisition of discipline-specific language by enabling repeated, interactive engagement with fundamental accounting definitions, concepts, and principles.

On the other hand, digital technologies enable the analysis of discipline-specific language as applied in practical contexts. For example, natural language processing (NLP)[4] techniques are increasingly used in accounting and tax research to analyze firm-generated textual outputs by transforming unstructured text into structured data (e.g., Bae et al., 2023; Bochkay et al., 2023; Li, 2010; Loughran & McDonald, 2011). Approaches such as dictionary-based text analysis[5] and, more recently, specialized large language models (LLMs)[6] allow researchers to capture both the frequency and contextual meaning of discipline-specific language, thereby facilitating the measurement of economically relevant constructs in, for example, legislative texts or corporate reporting (e.g., Hasan, 2020; Huang et al., 2023; Lei & Shu, 2024; Lynch et al., 2025; Ranta et al., 2023; Swenson et al., 2024).

Building on the preceding, this dissertation is motivated by the following research question: *How can digital technologies be used to integrate discipline-specific language into the fields of accounting and taxation?*

The dissertation consists of three essays that address different research gaps in response to the dissertation's research question. *Essay I* focuses on the acquisition of accounting-specific language. Specifically, it examines the use of wiki technology in introductory accounting courses, where students collaboratively develop a glossary of fundamental accounting definitions, concepts, and principles. *Essays II* and *III* focus on the analysis of tax-specific language as applied in practice. Specifically, *Essay II* employs dictionary-based text analysis technology to create a novel dataset from corporate tax legislation, facilitating empirical investigation. *Essay III* develops and employs a specialized LLM to analyze qualitative corporate tax disclosures.

Overall, the three essays in this dissertation demonstrate that digital technologies can effectively support both the *acquisition* and the *analysis* of discipline-specific language in

[4] NLP is a subfield of computer science that uses algorithms to process human (natural) language in the form of text or speech (Bochkay et al., 2023).

[5] Dictionary-based text analysis is a method of textual analysis in which a predefined list of words, the dictionary, is used to classify and quantify a text's content (Loughran & McDonald, 2011).

[6] LLMs are deep-learning NLP models that learn semantic and syntactic relationships from large text corpora, enabling them to capture context and word order rather than treating words independently (Huang et al., 2023).

accounting and taxation. *Essay I* shows that wiki-based approaches enable the collaborative learning and structuring of accounting terminology, while *Essays II* and *III* illustrate how NLP-based methods, including dictionary-based techniques and specialized LLMs, can facilitate the systematic, transparent, and scalable analysis of tax-related language in practice.

The introduction proceeds as follows. In Section 1, I outline the dissertation's structure and how the three essays interrelate. Part of this section provides an overview of the three essays (Table I-1) and a more detailed discussion of each individual essay (Sections 1.1 to 1.3). Finally, Section 2 discusses the dissertation's contribution to the academic literature and practice.

1. Three Essays in Accounting and Taxation

The three essays in this dissertation examine the role of digital technologies in integrating discipline-specific language into accounting and tax research. However, perspectives, foci, and methodological approaches differ. Table I-1 provides an overview of the three essays in accounting and taxation that are part of this dissertation.

Table I-1. Overview of the Dissertation

Essays in Accounting and Taxation: Integrating Discipline-Specific Language through Digital Technologies			
	Essay I	*Essay II*	*Essay III*
Title	Wiki-based collaborative learning in accounting education: impact on academic performance and student perceptions	Determinants of Corporate Environmental Taxes: Country-Level Empirical Evidence	How to design and employ Specialized Large Language Models for Accounting and Tax Research: The Example of TaxBERT
Research question(s)	How does wiki-based collaborative learning affect the academic performance of undergraduate students in an introductory accounting course, and how do students perceive the wiki approach?	What country-level factors explain cross-country variation in the implementation of corporate environmental taxes?	How can researchers design, evaluate, and apply domain-specific large language models (LLMs) for accounting and tax research?
Methodology	Mixed-methods approach: Empirical-archival analyses combined with qualitative thematic analysis	Empirical-archival analysis; Hypothesis development	Design Science Research; Empirical-archival analysis

Essays in Accounting and Taxation: Integrating Discipline-Specific Language through Digital Technologies			
	Essay I	*Essay II*	*Essay III*
Main results	Greater engagement in wiki-based collaborative learning significantly improves academic performance. Students report enhanced collaboration and learning, though challenges such as free-riding and content reliability remain.	Corporate environmental tax implementation is not driven by climate risk. It is negatively associated with corporate tax competition and positively associated with governance quality and female political representation.	TaxBERT significantly outperforms generic language models and traditional NLP methods in analyzing qualitative corporate tax disclosures.
Discipline-specific language	Accounting terminology in accounting education	Environmental tax terminology in tax legislation	Corporate tax terminology in corporate reporting
Digital technology	Wiki platform (Moodle)[7]	Dictionary-based text analysis (Python)	Large language models (Python)

Note: This table provides an overview of the three essays in accounting and taxation that are part of this dissertation.

Essay I examines the dissertation's research question from the perspective of acquiring discipline-specific language through digital technologies. Specifically, it focuses on the process of acquiring accounting-specific terminology (e.g., accounting definitions, concepts, and principles) in introductory accounting education. As a digital technology, a Moodle-based wiki platform functions as a collaborative online learning environment for undergraduate students to create an accounting glossary.[8] A mixed-methods approach is used to evaluate this educational intervention. Empirical-archival analyses are conducted to examine the wiki's effect on students' academic performance, complemented by a qualitative thematic analysis exploring students' perceptions of using the wiki for glossary creation.

In contrast, *Essays II* and *III* examine the dissertation's research question from the perspective of integrating discipline-specific language into digital text analysis technologies. Their purpose is to analyze this specialized language as it is used in practice (e.g., legislative texts or corporate reporting). *Essay II* focuses on integrating environmental tax terminology into dictionary-based text analysis to construct a novel dataset for empirical investigation. To be precise, the approach allows us to construct a global index scoring the implementation of

[7] Moodle is the university's digital learning platform through which the wiki platform was implemented. Briefly, Moodle is a digital learning platform that provides educators, administrators, and learners with a robust, secure, and integrated system for creating personalized learning environments.

[8] Briefly, wikis are web-based platforms that support collaborative content creation, allowing users to contribute, edit, and refine shared knowledge (Ioannou et al., 2015).

corporate environmental taxes. Methodologically, we utilize this self-constructed dataset to conduct hypothesis-based empirical-archival analyses.

Essay III examines the integration of corporate tax terminology into LLMs to better understand the specialized language used in corporate reporting. Specifically, a general-purpose LLM base model is augmented through tax-specific vocabulary expansion to better capture the nuanced language of qualitative corporate tax disclosures. Methodologically, this study combines a design science research approach with empirical-archival analyses. Given the importance of textual context, the digital technology employed in *Essay III* advances over that used in *Essay II* through the use of modern LLMs, which incorporate contextual and semantic information to facilitate a more comprehensive analysis of qualitative corporate tax disclosures (Bochkay et al., 2023; Huang et al., 2023).

Overall, the three essays jointly address the dissertation's research question by examining the integration of discipline-specific language through digital technologies from complementary perspectives.

1.1. Essay I: Wiki-based collaborative learning in accounting education: impact on academic performance and student perceptions

The first essay, "Wiki-based collaborative learning in accounting education: impact on academic performance and student perceptions," is co-authored with Andreas Seebeck and Johannes Voshaar. In response to the dissertation's research question, this essay examines how digital technologies can support collaborative learning of accounting definitions, concepts, and principles in introductory accounting education. Specifically, it centers around the question of how wiki-based collaborative learning affects the academic performance of undergraduate students in an introductory accounting course, and how students perceive the wiki approach. Table I-2 provides an overview of *Essay I*.

Table I-2. Overview of *Essay I*

Wiki-based collaborative learning in accounting education: impact on academic performance and student perceptions	
Authors	Lukas Schmidt, Andreas Seebeck, Johannes Voshaar
Abstract	This study investigates how wiki-based collaborative learning impacts students' academic performance and how students perceive the wiki in an undergraduate introductory accounting course. We design the wiki based on key dimensions of the social constructivism theory to scaffold the process of building an accounting glossary. Using a mixed-methods approach involving wiki engagement data, student data, and semi-structured interviews, we find that greater wiki engagement significantly enhances academic performance. In addition, thematic analysis of the semi-structured interviews sheds light on how students perceive the wiki. This study makes three contributions to

Wiki-based collaborative learning in accounting education: impact on academic performance and student perceptions	
	accounting education. First, it addresses a gap in the collaborative learning literature by examining wikis as a central tool for glossary-building. Second, it empirically evaluates the effectiveness of this wiki-based approach, responding to calls for research on IT-enabled collaboration. Third, it documents the wiki design and scaffolding materials, providing educators with a replicable framework.
Keywords	academic performance, collaborative learning, social constructivism, student perceptions, wiki-based learning
Publication status	Working Paper, submitted to "Accounting Education" on April 24, 2025. Status: Under 3rd-Round Review
Conference Contributions	Presented at the European Accounting Association's (EAA) Annual Congress 2026; British Accounting & Finance Association's (BAFA) Annual Conference (2023)
Contribution	Each section was joint work

Note: This table provides an overview of the first essay in accounting and taxation that is part of this dissertation.

The key motivation of this study is the central role of discipline-specific language in accounting. Prior research shows that accounting relies on specialized terminology and that early mastery of this language is essential for developing conceptual understanding (Bardien & Lubbe, 2025; Huang et al., 2023; Peters et al., 2013; Seebeck & Wolter, 2022). Accordingly, the study focuses on a wiki-based glossary approach to support students in acquiring and interrelating fundamental accounting knowledge.

The wiki design centers on collaborative glossary building to support mastery of discipline-specific accounting terminology. Students create structured wiki entries (i.e., accounting definitions, concepts, and principles) that promote student-centered learning, elaboration,[9] and retrieval practice[10] through repeated writing, linking, reviewing, revising, and testing (Reigeluth, 1979; Roediger & Karpicke, 2006). Guided by social constructivism,[11] the process combines individual and collaborative learning in three phases: an initial individual writing stage, followed by two collaborative writing rounds with rotating partners, each accompanied by peer review and revision. Complementing this, instructor scaffolding, such as clear wiki guidelines, exemplary entries, and review templates, supports students' learning while reducing cognitive load and ensuring format consistency and academic quality.

[9] Elaboration is the process of creating meaningful connections or associations with a specific idea or concept (Reigeluth, 1979).

[10] Retrieval practice is the process of actively recalling information from memory (e.g., through testing), which strengthens long-term retention more effectively than simply restudying the material (Roediger & Karpicke, 2006).

[11] Social constructivism theory posits that learning is fundamentally a social process, shaped through collaborative activities and interaction with peers and instructors (Adams, 2006; Helliar, 2013; Swan, 2005; Vygotsky, 1978).

We employ a mixed-methods approach, combining ordinary least squares (OLS) regression with multiple controls to assess the impact of the wiki-based collaborative learning environment on academic performance. Results show that greater wiki engagement is significantly associated with higher exam scores. Specifically, each additional wiki contribution increases academic performance by about 1.1–1.3 percentage points. Students with multiple wiki contributions outperform others by 4.6–6.4 percentage points, equivalent to roughly one numerical grade level. These findings are robust across various model specifications and additional tests. We complement the quantitative results with a qualitative analysis of 16 interviews, showing that the wiki enhances collaboration and learning beyond positive effects on academic performance. Students value peer interaction but report issues such as free-riding and uneven workload. Instructor scaffolding is generally appreciated, though students request more flexibility and clearer guidance on peer review. While the wiki technology supports exam preparation, concerns about content credibility suggest the need for stronger expert input and closer integration with course materials.

1.2. Essay II: Determinants of Corporate Environmental Taxes: Country-Level Empirical Evidence

The second essay, "Determinants of Corporate Environmental Taxes: Country-Level Empirical Evidence," is co-authored with Andreas Seebeck and Marius Weiß. This essay responds to the dissertation's research question by utilizing digital technologies and discipline-specific language to create a novel cross-country dataset for empirical investigation. Specifically, it investigates why countries differ in the implementation of corporate environmental taxes, a key instrument for internalizing environmental externalities (Brown et al., 2022; Jacob & Zerwer, 2024; Martinsson et al., 2024; Pigou, 2017). Table I-3 provides an overview of *Essay II*.

Table I-3. Overview of *Essay II*

Determinants of Corporate Environmental Taxes: Country-Level Empirical Evidence	
Authors	Lukas Schmidt, Andreas Seebeck, Marius Weiß
Abstract	This study investigates why countries differ in the implementation of corporate environmental taxes, a key instrument for internalizing environmental externalities. Using text mining, web scraping, and multi-stage human coding, we develop a novel index that captures the presence of corporate energy, pollution, resource, and transport taxes across 149 countries. We link cross-country variation in corporate environmental tax implementation to climate risk, corporate tax competition, governance quality, and female political representation using logistic regression. The results show that climate risk does not explain the implementation of corporate environmental taxes. In contrast, greater corporate tax competition significantly decreases the likelihood of implementation, while

Determinants of Corporate Environmental Taxes: Country-Level Empirical Evidence	
	higher governance quality and stronger female political representation increase it. Additional analyses indicate that the gender effect is likely to operate through parliamentary mechanisms and emerges only once a critical mass of women is present. These findings provide the first global, instrument-specific evidence on the determinants of corporate environmental taxes, highlighting key political and institutional drivers.
Keywords	environmental taxes; climate risk; corporate tax competition; governance quality; female political representation; critical mass theory
Publication status	Working Paper, submitted to "Accounting Forum" on December 20, 2024. Status: 2nd Revise and Resubmit
Conference Contributions	Presented at the EAA Annual Congress (2025; 2024); Virtual Doctoral Tax Seminar, Mannheim (2023)
Contribution	Each section was joint work

Note: This table provides an overview of the second essay in accounting and taxation that is part of this dissertation.

Corporate activities create negative environmental externalities, such as pollution and greenhouse gas emissions, not reflected in market prices (Chava, 2014; Crowley, 2000; Levitus et al., 2001). Corporate environmental taxes aim to internalize these costs by making firms pay for environmental damage (Intergovernmental Panel on Climate Change (IPCC), 2022b; Jacob & Zerwer, 2024; Pigou, 2017). Evidence shows these taxes can reduce emissions and energy use (Martin et al., 2014; Martinsson et al., 2024), and they are widely supported by international organizations (European Environment Agency (EEA), 2022; International Monetary Fund (IMF), 2019). However, their adoption varies across countries, and their specific determinants remain largely unexplored (Fredriksson & Neumayer, 2013; Fredriksson & Svensson, 2003).

Our self-constructed Corporate Environmental Tax Index (CET-Index) measures whether countries have implemented corporate environmental taxes across four dimensions: (i) energy, (ii) pollution, (iii) resource, and (iv) transport, as of the end of 2022 (IMF, 2024; Organization for Economic Co-operation and Development (OECD), 2024). It is constructed in four steps: (1) developing and rigorously validating a discipline-specific word list (177 terms) that operationalizes environmental tax terminology for Python-based dictionary-based text analysis, (2) collecting and text-mining corporate tax laws and government materials, (3) manually coding the presence of each tax dimension through multiple independent scoring rounds, and (4) validating results using external tax databases and expert consultations. The final index includes four indicators specific to the environmental tax dimensions and two higher-level measures that capture whether any environmental tax dimension is present and how many dimensions are implemented. Overall, the CET-Index provides a coherent but multidimensional measure of corporate environmental tax implementation across 149 countries.[12]

The findings indicate that climate risk does not explain cross-country differences in corporate environmental tax implementation. This result contradicts expectations from public interest theory (Breyer, 1982; Stern, 2008) and the theory of focusing events (Birkland, 2006), suggesting that exposure to climate hazards alone is insufficient to trigger preventive fiscal policies. Instead, structural constraints, administrative limitations, and policy inertia likely hinder such responses (Baumgartner & Jones, 1991; IPCC, 2022a, 2022b). In contrast, corporate tax competition is consistently associated with a lower likelihood of implementing these taxes, in line with the theory of tax competition (Glaeser et al., 2023; Wilson & Wildasin, 2004), indicating that governments competing for mobile capital tend to avoid additional cost burdens on firms. Further, higher governance quality increases the likelihood of implementing corporate environmental taxes, supporting public interest theory, which posits that effective institutions are better able to address environmental externalities (Breyer, 1982; Stern, 2008). Finally, greater female political representation is positively associated with implementation, with evidence indicating that a critical mass of around 20–40% is required to achieve a measurable impact, consistent with gender socialization theory[13] (Eagly, 1987; Eagly & Karau, 1991) and critical mass theory[14] (Kanter, 1977a, 1977b).

1.3. Essay III: How to design and employ Specialized Large Language Models for Accounting and Tax Research: The Example of TaxBERT

The third essay, "How to design and employ Specialized Large Language Models for Accounting and Tax Research: The Example of TaxBERT," is co-authored with Frank Hechtner, Andreas Seebeck, and Marius Weiß. This essay responds to the dissertation's research question by designing and employing a domain-specific large language model (LLM)

[12] This sample accounts for approximately 93% of the total gross domestic product (GDP) and greenhouse gas (GHG) emissions of United Nations (UN) member states between 2017 and 2021. However, certain countries remain excluded from our sample, presenting an inherent limitation.

[13] Briefly, gender socialization theory assumes that societal expectations shape human behavior differently on the basis of gender (Eagly, 1987; Eagly & Karau, 1991).

[14] Briefly, critical mass theory suggests that a certain threshold or critical mass of underrepresented group members is necessary for these individuals to influence group dynamics and outcomes meaningfully (Kanter, 1977a, 1977b).

specialized for analyzing qualitative corporate tax disclosures. Table I-4 provides an overview of *Essay III*.

Table I-4. Overview of *Essay III*

How to design and employ Specialized Large Language Models for Accounting and Tax Research: The Example of TaxBERT	
Authors	Frank Hechtner, Lukas Schmidt, Andreas Seebeck, Marius Weiß
Abstract	Large Language Models (LLMs) are emerging as the new gold standard for accounting researchers utilizing Natural Language Processing (NLP) techniques to analyze the characteristics of financial and non-financial disclosures. This study provides an "A-to-Z" description of how to design and employ specialized Bidirectional Encoder Representations of Transformers (BERT) models that are suitable for practical implementation by accounting and tax researchers and are environmentally sustainable. We begin by highlighting some NLP-related key challenges, pitfalls, and shortcomings. Next, we provide a user's guide to LLMs and BERT models. Using the case-based approach of TaxBERT, a domain-specific LLM pretrained for analyzing qualitative corporate tax disclosures, we provide a detailed discussion on the efficient design, evaluation, and application of specialized BERT models tailored to specific domains. We show that TaxBERT leads to significant performance improvements over generic LLMs, FinBERT, and traditional bag-of-words approaches. By showcasing the development as well as the potential of domain-specific models, we aim to inspire researchers to develop and apply specialized LLMs in their research.
Keywords	large language models (LLMs); machine learning; tax disclosure; text mining; natural language processing (NLP)
Publication status	Discussion Paper, available at SSRN under http://dx.doi.org/10.2139/ssrn.5146523, submitted to "European Accounting Review" on June 16, 2025. Status: 1st Revise and Resubmit
Conference Contributions	Presented at the 2026 International Conference on Computer Auditing (ICAEA); By Co-Authors: Presented at the EAA Annual Congress (2025; 2024); Diginomics Summer Conference (2025); Annual Mannheim Taxation Conference (2025); PSP-Doctoral Seminar (2025); Arqus-Annual Meeting (2025); AuBaNüDu-Doctoral Seminar (2025)
Contribution	Each section was joint work

Note: This table provides an overview of the third essay in accounting and taxation that is part of this dissertation.

This study is motivated by the evolution of textual analysis in accounting and tax research and the growing recognition that traditional NLP approaches, such as dictionary-based methods and bag-of-words models, are insufficient for capturing the contextual complexity of financial and non-financial disclosures (Huang et al., 2018; Li, 2010; Loughran & McDonald, 2011). While modern LLMs, such as GPT and BERT models, substantially improve the analysis of textual data by incorporating contextual and semantic relationships (Devlin et al., 2019; Siano & Wysocki, 2021), their reliance on general-domain pretraining limits their ability to accurately interpret highly specialized accounting and tax language

(Beltagy et al., 2019; Huang et al., 2023).[15] Existing domain-adapted models, such as FinBERT, represent an important step forward but remain insufficient to capture the nuanced, context-dependent terminology of subdisciplines such as taxation, auditing, and corporate governance (Huang et al., 2023). Consequently, there is a clear need for more specialized, domain-specific LLMs to ensure accurate and reliable analysis as disclosure requirements become increasingly complex and domain-specific (Fiechter et al., 2022).

In response to this need, we develop TaxBERT, a domain-specific language model tailored to qualitative corporate tax disclosures, by following a structured, multi-stage design process. First, we assess the necessity of specialization by analyzing the unique linguistic and institutional characteristics of tax disclosures. This includes, for example, legal complexity, jurisdictional heterogeneity, and non-standardized reporting formats, and assemble a large, diverse corpus of public tax-related texts spanning mandatory and voluntary disclosures (e.g., 10-K filings, Global Reporting Initiative (GRI) 207, Australian Tax Transparency Code (TTC), United Kingdom (UK) Tax Strategy) (Hoopes et al., 2024; Müller et al., 2020). Second, we construct a heterogeneous training dataset capturing broad contextual variation across jurisdictions and disclosure types. Third, we select an efficient yet high-performing base model (i.e., DistilRoBERTa) to balance predictive performance with computational feasibility (Sanh et al., 2020). Fourth, we perform domain-adaptive pretraining, including targeted vocabulary augmentation and careful hyperparameter tuning, to embed tax-specific language and semantics into the model (Gururangan et al., 2020; Howard & Ruder, 2018). Finally, we fine-tune TaxBERT on annotated downstream tasks, such as text classification (e.g., tax vs non-tax text) and sentiment detection (e.g., tax risk), ensuring both construct validity and superior empirical performance relative to existing models, including FinBERT and general-purpose LLMs (Devlin et al., 2019; Huang et al., 2023).

As a proof of concept, we apply TaxBERT to 67,784 U.S. 10-K filings to measure tax risk sentiment and examine its relationship with corporate tax avoidance. The model demonstrates clear empirical advantages. It more accurately distinguishes tax-related from non-tax-related content and avoids systematic misclassification observed in benchmark models

[15] General-purpose LLMs are pretrained on broad, general-domain corpora (e.g., Wikipedia and BookCorpus), which may limit their ability to generalize effectively across diverse or specialized contexts (Bochkay et al., 2023; Devlin et al., 2019).

such as FinBERT, leading to more reliable identification of both tax risk and risk-mitigation disclosures. Consistent with its superior classification performance, TaxBERT generates sentiment measures that differ meaningfully from those of general and finance-focused BERT models, reflecting its tax-specific understanding. Using these refined measures, we document a significant negative association between tax avoidance and disclosed tax risk sentiment, suggesting that firms engaging more aggressively in tax avoidance tend to downplay tax-related risks in their narratives. This relationship is not captured by traditional approaches or less specialized models, underscoring the importance of domain-specific LLMs for uncovering economically meaningful patterns in qualitative corporate tax disclosures.

2. Contribution

The three essays in this dissertation contribute to addressing the research question of how digital technologies can be used to integrate discipline-specific language into the fields of accounting and taxation in different yet complementary ways. In addition, the essays contribute insights that go beyond answering this research question.

Essay I makes three contributions. First, we show how wiki-based glossary building can support the collaborative acquisition and structuring of accounting terminology, addressing a gap in prior research where wikis have not been used as a central tool for this purpose. In addition, we respond to calls for further research on information technology (IT)-enabled collaborative learning in accounting education (Sangster et al., 2020). Finally, we provide a detailed, replicable wiki design, including scaffolding materials, that offers practical guidance for educators implementing similar approaches in introductory accounting education and beyond.

Essay II contributes in four ways. First, using a dictionary-based text analysis approach, it introduces a novel index, the CET-Index, enabling cross-country comparisons and future research. Second, it extends sustainability-related tax research by identifying key country-level determinants of corporate environmental taxes, shifting the focus beyond firm behavior (e.g., Adrian et al., 2023; Bird & Davis-Nozemack, 2018; Hardeck et al., 2024). Third, it contributes to the environmental policy literature by examining the determinants of a specific instrument rather than aggregate policies (e.g., Bättig & Bernauer, 2009; Gazmararian & Milner, 2024). Finally, it provides policy implications, emphasizing the roles of tax competition and gender diversity in advancing corporate environmental taxation (IPCC, 2022a, 2022b; United Nations Entity for Gender Equality and the Empowerment of Women (UN WOMEN), 2022).

Essay III makes two contributions. First, we address concerns regarding reproducibility and transparency in NLP-based accounting and tax research by providing a fully documented, open, and methodologically transparent framework for developing domain-specific BERT models, thereby mitigating the black-box limitations of proprietary LLMs (Hail et al., 2020). Second, we introduce TaxBERT, a specialized, resource-efficient LLM for processing qualitative corporate tax disclosures. It enables accurate, scalable, and cost-effective analysis, thereby facilitating broader and more accessible application of advanced NLP techniques in tax research.

Overall, this dissertation contributes to the accounting and tax literature by examining how digital technologies facilitate the integration of discipline-specific language into accounting and taxation. Beyond its academic contributions, the dissertation offers practical implications for a broader audience, including educators and policymakers.

3. References

Adams, P. (2006). Demystifying Constructivism. In L. Tan Wee Hin & R. Subramaniam (Eds.), *Handbook of Research on Literacy in Technology at the K-12 Level* (pp. 493–514). IGI Global.

Adrian, C., Garg, M., Pham, A. V., Phang, S.-Y., & Truong, C. (2023). Do Natural Disasters Affect Corporate Tax Avoidance? The Case of Drought. *Journal of Business Ethics*, *186*(1), 105–135. https://doi.org/10.1007/s10551-022-05250-7

Bae, J., Yu Hung, C., & van Lent, L. (2023). Mobilizing text as data. *European Accounting Review*, *32*(5), 1085–1106. https://doi.org/10.1080/09638180.2023.2218423

Bardien, M., & Lubbe, I. (2025). Knowledge in accounting: using semantic density to explore complexity in threshold concepts and terminology. *Accounting Education*. https://doi.org/10.1080/09639284.2025.2482686

Bättig, M. B., & Bernauer, T. (2009). National Institutions and Global Public Goods: Are Democracies More Cooperative in Climate Change Policy? *International Organization*, *63*(2), 281–308. https://doi.org/10.1017/S0020818309090092

Baumgartner, F. R., & Jones, B. D. (1991). Agenda Dynamics and Policy Subsystems. *Journal of Politics*, *53*(4), 1044–1074. https://doi.org/10.2307/2131866

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A Pretrained Language Model for Scientific Text. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1903.10676

Bird, R., & Davis-Nozemack, K. (2018). Tax Avoidance as a Sustainability Problem. *Journal of Business Ethics*, *151*(4), 1009–1025. https://doi.org/10.1007/s10551-016-3162-2

Birkland, T. A. (2006). *Lessons of Disaster: Policy Change after Catastrophic Events*. Georgetown University Press.

Bochkay, K., Brown, S. v., Leone, A. J., & Tucker, J. W. (2023). Textual Analysis in Accounting: What's Next?. *Contemporary Accounting Research*, *40*(2), 765–805. https://doi.org/10.1111/1911-3846.12825

Breyer, S. (1982). *Regulation and Its Reform*. Harvard University Press.

Brown, J. R., Martinsson, G., & Thomann, C. (2022). Can Environmental Policy Encourage Technical Change? Emissions Taxes and R&D Investment in Polluting Firms. *Review of Financial Studies*, *35*(10), 4518–4560. https://doi.org/10.1093/rfs/hhac003

Chava, S. (2014). Environmental Externalities and Cost of Capital. *Management Science*, *60*(9), 2223–2247. https://doi.org/10.1287/mnsc.2013.1863

Crowley, T. J. (2000). Causes of Climate Change Over the Past 1000 Years. *Science*, *289*(5477), 270–277. https://doi.org/10.1126/science.289.5477.270

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1810.04805

Eagly, A. H. (1987). *Sex Differences in Social Behavior*. Psychology Press.

Eagly, A. H., & Karau, S. J. (1991). Gender and the emergence of leaders: A meta-analysis. *Journal of Personality and Social Psychology*, *60*(5), 685–710. https://doi.org/10.1037/0022-3514.60.5.685

European Environment Agency (EEA) (2022). *The role of (environmental) taxation in supporting sustainability transitions,* Publications Office of the European Union. https://doi.org/10.2800/24487

Fiechter, P., Hitz, J.-M., & Lehmann, N. (2022). Real Effects of a Widespread CSR Reporting Mandate: Evidence from the European Union's CSR Directive. *Journal of Accounting Research*, 60(4), 1499–1549. https://doi.org/10.1111/1475-679X.12424

Fredriksson, P. G., & Neumayer, E. (2013). Democracy and climate change policies: Is history important? *Ecological Economics*, *95*, 11–19. https://doi.org/10.1016/j.ecolecon.2013.08.002

Fredriksson, P. G., & Svensson, J. (2003). Political instability, corruption and policy formation: the case of environmental policy. *Journal of Public Economics*, *87*(7–8), 1383–1405. https://doi.org/10.1016/S0047-2727(02)00036-1

Gazmararian, A. F., & Milner, H. v. (2024). Political Cleavages and Changing Exposure to Global Warming. *Comparative Political Studies*. *58*(9), 1965–1999. https://doi.org/10.1177/00104140241283014

Glaeser, S. A., Olbert, M., & Werner, A.-C. (2023). Tax Competition and Employment. *Accounting Review*, *98*(5), 267–296. https://doi.org/10.2308/TAR-2020-0575

Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. ArXiv Working Paper. https://doi.org/10.48550/arXiv.2004.10964

Hardeck, I., Hechtner, F., Seebeck, A., & Weiß, M. (2024). Does Public Tax Sustainability Disclosure Deter Corporate Tax Avoidance? Evidence from GRI 207 Reporting. SSRN Working Paper. https://ssrn.com/abstract=4910591

Hasan, M. M. (2020). Readability of Narrative Disclosures in 10-K Reports: Does Managerial Ability Matter? *European Accounting Review*, *29*(1), 147–168. https://doi.org/10.1080/09638180.2018.1528169

Helliar, C. (2013). The Global Challenge for Accounting Education. *Accounting Education*, *22*(6), 510–521. https://doi.org/10.1080/09639284.2013.847319

Hoopes, J. L., Robinson, L., & Slemrod, J. (2024). Corporate Tax Disclosure. *Journal of the American Taxation Association*, 46(2), 31–61. https://doi.org/10.2308/JATA-2022-037

Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1801.06146

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach. *Management Science*, *64*(6), 2833–2855. https://doi.org/10.1287/mnsc.2017.2751

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A Large Language Model for Extracting Information from Financial Text*. *Contemporary Accounting Research*, *40*(2), 806–841. https://doi.org/10.1111/1911-3846.12832

Intergovernmental Panel on Climate Change (IPCC) (2022a). *Climate Change 2022: Impacts, Adaptation and Vulnerability Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. https://www.ipcc.ch/report/ar6/wg2/downloads/report/IPCC_AR6_WGII_FullReport.pdf

Intergovernmental Panel on Climate Change (IPCC) (2022b). *Climate Change 2022 Mitigation of Climate Change Working Group III contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. https://www.ipcc.ch/report/ar6/wg3/downloads/report/IPCC_AR6_WGIII_FullReport.pdf

International Monetary Fund (IMF) (2019). *Fiscal Monitor: How to Mitigate Climate Change. Washington*, DC, October. https://www.imf.org/-/media/Files/Publications/fiscal-monitor/2019/October/English/text.ashx

International Monetary Fund (IMF) (2024, November 28). *Environmental Taxes*. https://climatedata.imf.org/datasets/3fb1ed30d3394574b3145246846023b1_0/about

Ioannou, A., Brown, S. W., & Artino, A. R. (2015). Wikis and forums for collaborative problem-based activity: A systematic comparison of learners' interactions. *Internet and Higher Education*, *24*, 35–45. https://doi.org/10.1016/j.iheduc.2014.09.001

Jacob, M., & Zerwer, K. L. (2024). Emission Taxes and Capital Investments: The Role of Tax Incidence. *Accounting Review*, *99*(5), 247–278. https://doi.org/10.2308/TAR-2023-0053

Kanter, R. M. (1977a). *Men and women of the corporation*. Basic Books.

Kanter, R. M. (1977b). Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women. *American Journal of Sociology*, *82*(5), 965–990. https://doi.org/10.1086/226425

Kelly, O., Hall, T., & Connolly, C. (2023). PACE-IT: designing blended learning for accounting education in the challenging context of a global pandemic. *Accounting Education*, *32*(6), 626–645. https://doi.org/10.1080/09639284.2022.2090851

Lei, L. G., & Shu, S. Q. (2024). Does audit regulation improve corporate decision making? Evidence from the disclosure of tax-related key audit matters. *Journal of Business Finance & Accounting*, *51*, 2055–2092. https://doi.org/10.1111/jbfa.12774

Levitus, S., Antonov, J. I., Wang, J., Delworth, T. L., Dixon, K. W., & Broccoli, A. J. (2001). Anthropogenic Warming of Earth's Climate System. *Science*, *292*(5515), 267–270. https://doi.org/10.1126/science.1058154

Li, F. (2010). Textual analysis of corporate disclosures: A survey of the literature. *Journal of Accounting Literature*, *29*, 143–165.

Li, H., Li, Q., Xu, Z., & Ye, X. (2024). Digital technologies. *Journal of Digital Economy*, *3*, 240–248. https://doi.org/10.1016/j.jdec.2025.02.001

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Lynch, D. P., Mandell, A., & Rousseau, L. M. (2025). The consequences of expanded audit reporting: implications of tax key audit matters for tax attribute valuation and auditor-provided tax services. *Review of Accounting Studies*, *30*(4), 3894–3953. https://doi.org/10.1007/s11142-025-09884-6

Martin, R., de Preux, L. B., & Wagner, U. J. (2014). The impact of a carbon tax on manufacturing: Evidence from microdata. *Journal of Public Economics*, *117*, 1–14. https://doi.org/10.1016/j.jpubeco.2014.04.016

Martinsson, G., Sajtos, L., Strömberg, P., & Thomann, C. (2024). The Effect of Carbon Pricing on Firm Emissions: Evidence from the Swedish CO2 Tax. *Review of Financial Studies*, *37*(6), 1848–1886. https://doi.org/10.1093/rfs/hhad097

Müller, R., Spengel, C., & Vay, H. (2020). On the Determinants and Effects of Corporate Tax Transparency: Review of an Emerging Literature. ZEW - Centre for European Economic Research Discussion Paper No. 20-063. https://doi.org/http://dx.doi.org/10.2139/ssrn.3736747

Organization for Economic Co-operation and Development (OECD) (2024). *Environmentally related tax revenue - Dataset documentation*. https://stats.oecd.org/wbos/fileview2.aspx?IDFile=4f959e36-628a-4196-9012-b422b2e0ad40

Peters, P., Smith, A., Middledorp, J., Karpin, A., Sin, S., & Kilgore, A. (2013). Learning essential terms and concepts in Statistics and Accounting. *Higher Education Research & Development*, *33*(4), 742–756. https://doi.org/10.1080/07294360.2013.863838

Pigou, A. (2017). *The economics of welfare*. Routledge.

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. https://doi.org/10.2308/iace-50066

Ranta, M., Ylinen, M., & Järvenpää, M. (2023). Machine Learning in Management Accounting Research: Literature Review and Pathways for the Future. *European Accounting Review*, *32*(3), 607–636. https://doi.org/10.1080/09638180.2022.2137221

Reigeluth, C. M. (1979). In Search of a Better Way to Organize Instruction: The Elaboration Theory. *Journal of Instructional Development*, *2*(3), 8–15. https://www.jstor.org/stable/30220561

Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Sangster, A., Stoner, G., & Flood, B. (2020). Insights into accounting education in a COVID-19 world. *Accounting Education*, *29*(5), 431–562. https://doi.org/10.1080/09639284.2020.1808487

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1910.01108

Seebeck, A., & Wolter, R. M. (2022). Financial education of founders, is it important? A case study of Jacobs Startup Competition. *International Journal of Entrepreneurial Venturing*, *14*(1), 122–143. https://doi.org/10.1504/IJEV.2022.122018

Siano, F., & Wysocki, P. (2021). Transfer Learning and Textual Analysis of Accounting Disclosures: Applying Big Data Methods to Small(er) Datasets. *Accounting Horizons*, *35*(3), 217–244. https://doi.org/10.2308/HORIZONS-19-161

Stern, N. (2008). The Economics of Climate Change. *American Economic Review*, *98*(2), 1–37. https://doi.org/10.1257/aer.98.2.1

Swan, K. (2005). A constructivist model of thinking about learning online. In J. Bourne & J. Moore (Eds.), *Elements of quality online education: Emerging communities* (pp. 13–30). Sloan-C.

Swenson, C., Qu, H., & Song, X. (2024). Complexity and the Tax Code: A Textual Analysis. *Public Finance Review*, *52*(4), 466–538. https://doi.org/10.1177/10911421241233110

United Nations Entity for Gender Equality and the Empowerment of Women (UN WOMEN). (2022, February 27). *Explainer: How gender inequality and climate change are interconnected*. https://wrd.unwomen.org/explore/insights/explainer-how-gender-inequality-and-climate-change-are-interconnected

Voshaar, J., Knipp, M., Loy, T., Zimmermann, J., & Johannsen, F. (2023). The impact of using a mobile app on learning success in accounting education. *Accounting Education*, *32*(2), 222–247. https://doi.org/10.1080/09639284.2022.2041057

Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press.

Wilson, J. D., & Wildasin, D. E. (2004). Capital tax competition: bane or boon. *Journal of Public Economics*, *88*(6), 1065–1091. https://doi.org/10.1016/S0047-2727(03)00057-4

II. Essay I

Wiki-based collaborative learning in accounting education: impact on academic performance and student perceptions

Lukas Schmidt, Andreas Seebeck, Johannes Voshaar

Abstract

This study investigates how wiki-based collaborative learning impacts students' academic performance and how students perceive the wiki in an undergraduate introductory accounting course. We design the wiki based on key dimensions of the social constructivism theory to scaffold the process of building an accounting glossary. Using a mixed-methods approach involving wiki engagement data, student data, and semi-structured interviews, we find that greater wiki engagement significantly enhances academic performance. In addition, thematic analysis of the semi-structured interviews sheds light on how students perceive the wiki. This study makes three contributions to accounting education. First, it addresses a gap in the collaborative learning literature by examining wikis as a central tool for glossary-building. Second, it empirically evaluates the effectiveness of this wiki-based approach, responding to calls for research on IT-enabled collaboration. Third, it documents the wiki design and scaffolding materials, providing educators with a replicable framework.

Acknowledgements

We thank Patricia Everaert, Lize Vanderstraeten, Amelie Vanhove, Hans van der Heijden, Greg Stoner, and Barbara Flood for their valuable guidance. We thank the participants of the BAFA Accounting Education SIG annual conference (2023, Durham) for their constructive comments on an earlier version of this paper. We want to thank the University of Erlangen-Nuremberg and Constructor University Bremen for providing us with the opportunity to present and discuss our work, as well as for their input on the interview guide. We sincerely appreciate the valuable comments and suggestions from the Editors and anonymous reviewers of "Accounting Education," which have greatly contributed to improving this study.

1. Introduction

As accounting educators face growing demands for inclusive and technology-supported teaching, investigating the effectiveness and exploring student perceptions of student-centered, IT-enabled collaboration approaches has become an important area of inquiry (Musundwa, 2025; Sangster et al., 2020). In response, this study investigates a wiki-based collaborative learning approach by addressing the following research question: *How does wiki-based collaborative learning affect the academic performance of undergraduate students in an introductory accounting course, and how do students perceive the wiki approach?*

Our motivation for this study is threefold. First, accounting has a unique disciplinary terminology (Hechtner et al., 2025; Huang et al., 2023), and early mastery of this terminology is essential for conceptual understanding (Peters et al., 2013; Seebeck & Wolter, 2022). Our wiki-based collaborative learning approach engages students in creating an accounting glossary that promotes both the acquisition of key terms and the exploration of their interrelationships. Second, increasing diversity within student cohorts further amplifies the need for inclusive, student-centered learning approaches (Matheson, 2018; Musundwa, 2025). Third, evolving digital technologies have increasingly enabled hybrid and online teaching in accounting education. The COVID-19 pandemic accelerated this development by prompting a rapid transition to remote teaching, requiring accounting educators to adopt digital learning approaches (Lux et al., 2023; Martins et al., 2024). This shift has also led to calls for further research on IT-enabled collaborative learning in accounting education (Sangster et al., 2020).

We empirically examine a wiki-based collaborative learning approach as it aligns well with student-centered teaching by promoting active learning, knowledge sharing, and teamwork, which are essential for academic development and professional practice (Camacho et al., 2016; Cilliers, 2017; Zheng et al., 2015). Wikis can facilitate shared knowledge construction and peer-to-peer interaction, encouraging students to take ownership of their learning (Lin & Reigeluth, 2019). Beyond their pedagogical affordances, wikis offer a cost-effective and collaborative alternative to traditional, teacher-centered approaches, which is particularly valuable in resource-constrained learning environments (Ainsworth, 2021). Although IT-enabled collaboration is increasingly adopted in accounting education, the effectiveness of wiki-based approaches remains underexplored. Much of the existing research considers wikis as supplementary rather than primary tools for collaborative learning. We address this research gap by examining how wiki-based collaborative learning affects the academic performance of undergraduate students in an introductory accounting course and how students perceive the wiki approach. Our wiki is designed according to the principles of social constructivism, which views learning as a socially situated activity shaped by contextual factors and active engagement among learners and digital platforms (Barak, 2017; Bozkurt, 2017). It enables the collaborative creation of an accounting glossary through a mix of individual and collaborative writing activities.

We use a mixed-methods approach to answer our research question. First, we run an ordinary least squares (OLS) regression with several controls for student success to examine the impact of the wiki intervention on academic performance. We find that with increasing wiki engagement, as measured by the number of glossary articles contributed, students achieve significantly higher scores in the final exam. Depending on the model specification, each additional wiki contribution is associated with a 1.1 to 1.3 percentage point increase in academic performance. In additional tests, we find that students contributing more than once, labeled as comprehensive contributors, score significantly higher on the final exam than noncomprehensive contributors. Depending on the model specification, comprehensive contributors score between 4.6 and 6.4 percentage points higher, which is pedagogically meaningful, given that it corresponds roughly to a two-grade-level difference. Our regression results remain robust across various robustness checks, such as the use of alternative independent variables, the inclusion of additional controls, entropy balancing, and propensity score matching.

We complement the quantitative analysis with a qualitative exploration of student perceptions, as regression alone cannot capture the complexity of wiki-based collaborative learning. Thematic analysis of 16 semi-structured interviews indicates that our wiki fosters collaboration and enhances learning beyond measurable exam performance gains. Guided by social constructivism, we examine three themes: student collaboration, instructor scaffolding, and learning experiences. First, students value peer feedback and knowledge sharing, but issues such as free-riding and workload imbalance indicate a need for contribution tracking and group consent. Second, instructor scaffolds, including guidelines and templates, are appreciated, though students seek greater flexibility and clearer peer review support. Third, while the wiki aids exam preparation, concerns about content credibility point to integrating expert feedback and stronger links to course materials. Our findings highlight the potential of scaffolded, wiki-based collaborative learning for enhancing student academic performance and enriching course experience, while emphasizing the need for effective design.

1.1. Contribution

This study makes three key contributions. First, we contribute to the collaborative learning literature in accounting education by addressing a research gap: the use of wikis as a central vehicle for supporting the collaborative acquisition of discipline-specific accounting terminology. While prior research has examined face-to-face collaboration (e.g., Nsor-Ambala, 2022) and, more recently, IT-enabled approaches (e.g., Coetzee et al., 2023), wiki-based approaches have only been used in supplementary ways and not explicitly for glossary-building (e.g., Wyness & Dalton, 2018). Second, this study responds to the call for research extending the evaluation of IT-enabled collaborative learning approaches in accounting education (Sangster et al., 2020). Third, we detail our approach to wiki-based collaborative learning, documenting the design of the wiki, along with scaffolding materials provided as an exhibit. This documentation offers educators, especially in introductory courses requiring glossary knowledge, a replicable framework applicable across diverse learning environments, enhancing the approach's practical relevance for accounting education and beyond.

The paper proceeds as follows. Section 2 provides the theoretical background and literature review. The research method is presented in Section 3, which outlines the wiki design, study setting, sample, and data. We present our results in Section 4. Finally, Section 5 concludes the paper by discussing the findings, acknowledging limitations, and outlining directions for future research.

2. Theoretical Background and Literature Review

2.1. Social Constructivism Theory

Social constructivism theory posits that learning is fundamentally a social process, shaped through collaborative activities and interaction with peers and instructors (Adams, 2006; Helliar, 2013; Swan, 2005; Vygotsky, 1978). It acknowledges the need for peer engagement, as social engagement enables learners to gradually take greater control of their own learning (Torrance & Pryor, 1998). Consequently, knowledge is socially constructed through discussion and collaboration with others (Bozkurt, 2017). Within this paradigm, the instructor facilitates a supportive environment and scaffolds learning through guidance, peer interaction, and digital platforms (Adams, 2006; Barak, 2017). In sum, social constructivism recognizes the instructor as a guide rather than a director or dictator (Powell & Kalina, 2009). According to cognitive load theory, scaffolds must be carefully designed to minimize extraneous load (e.g., unclear instructions) and manage the complexity of collaborative tasks so that learners can focus their limited cognitive resources on meaningful knowledge construction (Sweller, 2010).

Recent accounting education research has applied social constructivism to promote student collaboration and enrich their learning experiences. For example, Christensen et al. (2019) implement team-based learning to create structured opportunities for student collaboration within accounting courses. In this setting, instructors scaffold the process, allowing students to create meaningful learning experiences through peer interaction and team-based problem solving. Oosthuizen et al. (2021) adopt a social constructivist perspective to explore how instructor-scaffolded teamwork process controls influence accounting students' engagement in simulated business meetings. The study emphasizes the instructor's role in scaffolding teamwork dynamics, supporting collaboration, and enriching student learning. Mali and Lim (2022) apply a social constructivist framework through a research-informed teaching intervention that prompts students to engage in reflective group discussions. The intervention encourages instructor scaffolding in guiding critical thinking and supports the construction of richer learning experiences by challenging traditional textbook-based views of accounting. Hu et al. (2024) employ a social constructivist approach to examine how various online platforms facilitate student collaboration and instructor interaction. They investigate the level of engagement and the quality of learning experiences that students develop through dialogue with their peers and instructors.

Informed by recent applications of social constructivist perspectives in accounting education research, our study operationalizes three core dimensions of the theory: student collaboration, instructor scaffolding, and learning experiences. These dimensions align with key constructivist principles: knowledge is co-constructed through peer interaction (e.g., Christensen et al., 2019; Hu et al., 2024), learning requires guided support from instructors (e.g., Mali & Lim, 2022; Oosthuizen et al., 2021), and well-structured, scaffolded collaboration can enrich student learning experiences (e.g., Christensen et al., 2019; Mali & Lim, 2022).

These three dimensions inform our design of the wiki, guide the development of the semi-structured interview guide, and shape the interpretation of our qualitative findings.

2.2. Collaborative Learning in Accounting Education

In accounting education research, collaborative learning is typically studied in two delivery contexts: face-to-face and IT-enabled (Gomez et al., 2010). For example, research shows that face-to-face collaborative learning can positively influence students' perceptions of the course (Ainsworth, 2021; Opdecam & Everaert, 2012), foster the development of critical thinking and employability skills essential to the accounting profession (Ainsworth, 2021; Christensen et al., 2019; De Klerk et al., 2025; Goosen & Steenkamp, 2025; Shawver, 2020; Tan, 2019), and enhance student academic performance (Nsor-Ambala, 2022; Opdecam & Everaert, 2019; Opdecam et al., 2014).

More recently, IT-enabled collaborative learning is increasingly recognized as an alternative to face-to-face collaborative learning (e.g., Coetzee et al., 2023; Kelly et al., 2023; Tello et al., 2026). For instance, Coetzee et al. (2023) explore the use of messenger apps by students to facilitate peer feedback during online assessments and investigate the types of feedback provided. The researchers chose this delivery format of social constructivist-based collaborative learning due to its benefits as a learning tool, such as flexible use, personalization, and active participation, enabling productive conversations and collaborations between students.

Kelly et al. (2023) examine the design and implementation of a blended learning intervention developed in response to the disruptions caused by the COVID-19 pandemic. Targeted at undergraduate accounting students, the initiative integrates a structured, scaffolded model of blended learning with a strong focus on digital literacy and peer engagement. The researchers combined asynchronous and synchronous digital communication platforms to support continuous learning and peer dialogue. Findings indicate that students benefited from the flexibility and accessibility of the digital platforms, which supported collaborative learning.

Tello et al. (2026) investigate the implementation of online discussion board activities in postgraduate accounting courses as a means to foster students' professional skills, such as critical thinking and teamwork. The researchers select the discussion board format for its ease of use, scalability, and ability to facilitate reflective, peer-driven engagement. The discussion board format supports the co-construction of knowledge. Students were required to critically comment on current news articles relevant to their coursework and interact meaningfully with peers' posts. The researchers find that the activity significantly enhances critical thinking and collaborative learning.

2.3. Wiki-based Collaborative Learning

IT-enabled, social constructivist-based collaborative learning approaches are becoming increasingly popular in accounting education. However, not every collaboration format is suitable for every setting. For example, in introductory accounting courses, acquiring the unique accounting terminology and interrelationships between terms and concepts is essential (Peters et al., 2013; Seebeck & Wolter, 2022). In this context, non-linear collaboration[16] formats, such as wikis, may be particularly effective, allowing students to co-construct an accounting glossary and link related terms within the wiki network. Thereby, this approach contrasts with linear collaboration through, for example, messaging apps, online forums, and discussion boards, which facilitate sequential discussions (Camarero et al., 2012; Coetzee et al., 2023; Kelly et al., 2023; Tello et al., 2026).

Wikis are web-based platforms that support collaborative content creation, allowing users to contribute, edit, and refine shared knowledge (Ioannou et al., 2015). Grounded in social constructivism, wikis offer real-time editing, version tracking, and integrated discussion forums to facilitate collaboration (Zheng et al., 2015). Furthermore, wikis promote scaffolding through peer interaction and instructor support (Luo & Chea, 2020), enabling the co-construction of knowledge (Biasutti, 2017). However, effective design is essential (Lin & Reigeluth, 2019; Rott & Weber, 2013; Salaber, 2014). For example, Rott and Weber (2013) demonstrate that successful wiki-based collaboration necessitates clear guidance and other forms of scaffolding.

Wikis can promote teamwork and knowledge sharing (Garcia & Privado, 2023). Active contributors tend to show stronger collaboration skills and engagement than in traditional group work (Chu et al., 2017). However, uneven participation remains a concern, highlighting the need for structured peer assessment (Gielen & De Wever, 2015). Furthermore, wikis allow instructors and peers to guide learning through feedback and revision tracking (Huang, 2019;

[16] Non-linear collaboration allows content to be co-created and connected in a network-like structure, whereas linear collaboration formats organize peer interaction in a sequential order, such as threaded discussions.

De Arriba, 2017). Moreover, wikis can enhance comprehension, retention, and writing skills by encouraging active learning (Kassem, 2017; Page & Reynolds, 2015).

In the context of accounting education, prior studies utilize wikis for various instructional purposes, including problem-based learning, cooperative teamwork, and knowledge management (Osgerby, 2013; Turner & Baskerville, 2013; Wyness & Dalton, 2018). Osgerby (2013) reports that while students acknowledge the value of wikis for structuring their understanding of accounting concepts, they express concerns about the reliability of peer-generated content. Turner and Baskerville (2013) implement wikis to facilitate teamwork in accounting courses, demonstrating that students benefit from student collaboration, but they require scaffolded interventions to navigate group dynamics effectively. Wyness and Dalton (2018) employ wikis as collaborative tools within a problem-based learning module on sustainability accounting, grounded in social constructivist pedagogy, to help students develop skills such as critical thinking and teamwork.

Overall, prior research in accounting education has examined face-to-face collaborative learning (Ainsworth, 2021; Opdecam & Everaert, 2012, 2019; Nsor-Ambala, 2022), as well as IT-enabled collaborative learning approaches more recently (Coetzee et al., 2023; Kelly et al., 2023; Tello et al., 2026). However, wiki-based approaches have only been explored in limited, supplementary ways in accounting education (Osgerby, 2013; Turner & Baskerville, 2013; Wyness & Dalton, 2018). To the best of our knowledge, no study has investigated a wiki-based collaborative learning approach to support the acquisition of discipline-specific accounting terminology, which is a fundamental requirement for undergraduate students in introductory accounting courses (Peters et al., 2013; Seebeck & Wolter, 2022). Our study addresses this gap by positioning the wiki as the central vehicle of a social constructivist, IT-enabled collaborative learning approach explicitly designed to facilitate the collaborative acquisition of accounting terminology.

3. Research Method

3.1. Wiki Design

We chose glossary-building as the core wiki activity for several reasons. First, the accounting discipline relies on unique terminology (Hechtner et al., 2025; Huang et al., 2023). Second, early mastery of accounting fundamentals is crucial for developing a conceptual understanding (Peters et al., 2013; Seebeck & Wolter, 2022). Third, glossary building reflects elaboration theory, which holds that learning deepens when students actively connect new material to prior knowledge (Reigeluth, 1979). Our structured entries, which include definitions, examples, and the creation of testing questions for peers, foster elaborative processing. Fourth, it also draws on retrieval-practice theory, which shows that repeated recall strengthens long-term retention (Roediger & Karpicke, 2006). Cross-linked entries, repeated authoring, reviewing, revising, and navigating wikis require students to revisit and apply terms multiple times, thereby embedding retrieval practice.

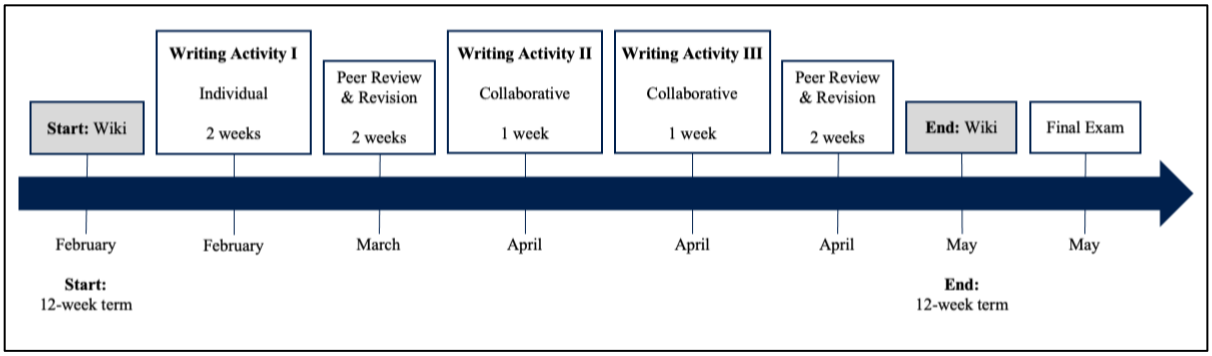
Overall, the wiki design is guided by three key dimensions of social constructivism identified in recent accounting education research applying the theory: student collaboration, instructor scaffolding, and learning experiences. The glossary-building process involves three phases of writing activities.[17] At the beginning of each activity, instructors assign a glossary term to each participant.[18] First, students individually contributed to the wiki over a two-week writing period. Second, they engaged in a one-week collaborative writing activity. Third, a second one-week collaborative writing activity followed. This phased approach is designed to promote individual accountability through solo contributions, followed by peer collaboration and feedback during the group writing activities. This approach reflects active learning, co-construction of knowledge, and learning through social interaction as key tenets of social constructivism (Adams, 2006; Helliar, 2013; Swan, 2005; Torrance & Pryor, 1998; Vygotsky, 1978). For the collaborative writing activities, students worked in randomly selected pairs, with different partners for each collaborative writing activity. We choose this approach to discourage students from dividing the work and writing a single article individually, rather than collaborating with their peers on two articles. Thereby, we offer multiple opportunities for collaboration, both through the co-construction of a shared accounting glossary and through collaborative writing activities. Following the individual writing activity and both collaborative writing activities, there was a one-week peer review period, followed by a one-

[17] For the writing activities, students were permitted to use generative artificial intelligence (AI) tools, such as ChatGPT. Students were provided with online tutorials offering guidance on their appropriate use. However, as the intervention semester (Spring 2023) was shortly after the launch of ChatGPT as the first broadly accessible AI tool, the dissemination and use of such tools were likely more limited than today. For instance, Google's Bard, the predecessor of Gemini, was only made available in the EU in July 2023.

[18] Two researchers compiled lists of introductory accounting terms from course materials (e.g., lecture and tutorial slides). By consensus, we finalized a glossary, excluding overly specialized terms (e.g., cash flow offsetting) to ensure comparable difficulty. The terms were then ordered by their relevance to the course, with the most relevant (e.g., double-entry bookkeeping) first. Students were randomly assigned to terms in this order.

week revision period after each review. Figure II-1 shows the main elements of the wiki-based collaborative learning approach and its implementation timeline.

Figure II-1. Timeline of the Wiki-based Collaborative Learning Approach



To scaffold the collaborative co-construction of an accounting glossary, we implemented the wiki in the university's online learning management system (Moodle). We provided scaffolding materials and ex-ante instructions on the wiki platform, as well as on-demand support via a wiki-integrated Frequently Asked Questions (FAQ) section. All of these were introduced and explained in one lecture and online tutorial dedicated exclusively to the wiki project. For example, we provide contribution guidelines as an instructional scaffold (e.g., word limit, citation and referencing style, use of academic sources, contribution structure) to promote academic validity and ensure consistency across the wiki contributions regarding formatting and structure.[19] We do so to alleviate previous concerns expressed by accounting students that wikis are of low academic quality (Osgerby, 2013). In addition, we provide instructor-written exemplary wiki contributions, thought of as an expert scaffold.[20] Moreover, we provide a peer review template, which we deliver as a resource scaffold.[21] These scaffolding strategies reflect the instructor's role as a facilitator in a social constructivist learning environment, supporting students within their zone of proximal development through structured guidance and expert modeling (Adams, 2006; Barak, 2017; Powell & Kalina, 2009). In line with cognitive load theory, these scaffolds are designed to reduce extraneous load by providing clear instructions, consistent formatting requirements, and structured feedback templates, thereby ensuring that students can devote their cognitive resources to engaging with accounting concepts rather than navigating task ambiguity (Sweller, 2010).

[19] A detailed summary of the wiki guidelines is presented in Appendix II-1.
[20] One example of the six exemplary wiki contributions is presented in Appendix II-2.
[21] The peer review template is presented in Appendix II-3.

The wiki was introduced to all enrolled students in the first lecture through a hands-on overview. Voluntary participation and registration details were shared via lectures, online tutorials, Moodle, email, and Microsoft Teams. The wiki-based glossary building process spanned the entire 12-week term, from the first week of the semester to the close of the lecture period, preceding the exam preparation period. Our teaching intervention and research project were supported by third-party funding and approved by the Constructor University Bremen (Project No. 50835). Moreover, this study followed ethical procedures given by the Helsinki Declaration (World Medical Association, 2013). Students provided informed consent via Moodle, confirming voluntary participation and their understanding of how their data would be used for research purposes.

3.2. Study Setting

We implemented the wiki in an undergraduate introductory accounting course delivered in English during the Spring 2023 semester. The introductory accounting course is for first-year students in three business programs and an elective for students in other programs. In line with accounting education research, we award bonus points for wiki engagement to incentivize participation and ensure effort (e.g., Hong, 2024; Voshaar et al., 2025).[22] We calculate the bonus points based on the quality of their wiki contributions. Assessment is via a 120-minute final paper-based exam combining multiple-choice and open-ended questions, testing both conceptual understanding and the ability to apply fundamental accounting concepts in computational tasks.[23]

3.3. Sample and Data

3.3.1. Sample Selection

Our initial sample consists of 128 students enrolled in the introductory accounting course. Of these, 110 took the final exam. The dropout rate of 18 students is consistent with prior cohorts. We exclude data from 3 students due to missing values in some of the control variables, leaving us with a final sample size for our quantitative analyses of 107.[24] Of these,

[22] We award a maximum improvement of 20% on the final exam score.

[23] Examples of final exam questions include a multiple-choice item asking: “A balance sheet is designed to show: (a) what the entity could be sold for, (b) what it would cost to set up a similar entity, (c) the financial position of an entity, or (d) information on the financial results of an entity’s business activities over a period of time.” Another example is an open-ended question requiring students to state the basic accounting equation.

[24] We assess the adequacy of our sample size with a post hoc power analysis in G*Power (version 3.1.9.6) (Faul et al., 2009). Using the observed effect sizes for Model 1 (f^2=0.65) and Model 2 (f^2=0.72) from Table II-4,

89 actually contributed to the wiki-based accounting glossary. We invited all wiki contributors enrolled in the course to participate in the qualitative research of this study after the course concluded and final exam results were released. A total of 16 students, hereafter referred to as interview participants, volunteered.[25]

3.3.2. Data Collection

To investigate the quantitative part of our research question, how wiki-based collaborative learning affects academic performance, we draw on a variety of data sources. First, we collect students' wiki engagement data from the wiki platform. Second, we calculate the student's academic performance as the final exam score, excluding the bonus points. Third, we collect students' socio-demographic and structural data (i.e., age, gender, region of origin) as well as additional factors that have been found to impact student academic performance (i.e., university performance, pre-university performance, study program, accounting-related work experience, and English language proficiency) through the university student database and hand-collection directly from the students. A member of the research team not involved in the analyses merged and pseudonymized the data. The final pseudonymized dataset does not allow for the identification of individual students.

To investigate the qualitative part of our research question, how students perceive the wiki-based collaborative learning approach, we collect in-depth data through semi-structured interviews (Bayne et al., 2022; Oosthuizen et al., 2021) and analyze it thematically. The development of the interview guide is informed by three key social constructivist dimensions identified in accounting education research: student collaboration, instructor scaffolding, and learning experiences. To evaluate the appropriateness of the interview questions, we supplemented the author team's expertise in accounting education research by developing the interview questions with several students enrolled in the course in previous semesters, as well as pilot-testing them with two accounting educators from external institutions. We made minor

five predictors, α=0.05, and N=107, we find statistical power well above the recommended 0.80 threshold (Cohen, 1992), indicating the sample was sufficient to detect the observed effects.

[25] Of the 16 interviewees, 9 (56%) are female and 7 (44%) male; 15 (94%) are enrolled in business-related programs. Participants represent 12 countries across five world regions. Twelve participants (75%) contributed three articles, three (19%) contributed two, and one (6%) contributed one, averaging 2.69 contributions each. This suggests strong potential for analyzing student collaboration, though uneven contribution levels may limit insights from less active participants. In the overall wiki, contributors with three articles constitute 55%, those with two articles 33%, and those with one article 12%, indicating the interview sample does not fully reflect the actual distribution and may limit the generalizability of findings.

adjustments to enhance the clarity and wording of certain questions based on feedback. Finally, we use a semi-structured interview guide consisting of nine questions, presented in Appendix II-4. All interviews began with a request for consent to use their responses in this study. All interviewees agreed to be recorded, and none subsequently requested to withdraw their participation. In line with Bayne et al. (2022), we conduct semi-structured one-on-one online interviews in English in July 2023. To minimize potential bias, a faculty member uninvolved in the teaching intervention and module coordination conducted the interviews. The average length of the interviews was 39 minutes, with a minimum of 30 minutes and a maximum of 65 minutes. Data saturation was evident during the latter interviews as no new information or ideas were noted, and participants kept providing similar answers for specific questions (Saunders et al., 2018).

3.3.3. Data Analysis

We investigate the quantitative part of our research question by estimating students' academic performance in the final exam using an OLS regression. In our empirical model, we control for several factors that are associated with student academic performance in accounting education literature (e.g., Coetzee et al., 2018; Tan & Laswad, 2015; Voshaar et al., 2023). Our empirical model is specified as follows (Equation II-1):

$$
\begin{aligned}
ACADEMIC\_PERFORMANCE_i & \\
&= \beta_0 + \beta_1 CONTRIBUTIONS_i + \beta_2 AGE_i + \beta_3 FEMALE_i + \beta_4 GPA_i \\
&+ \beta_5 PRE\_UNI\_PERFORMANCE_i + STUDY\_PROGRAM\ FE + REGION\ FE \\
&+ \varepsilon_i
\end{aligned}
\tag{II-1}
$$

where i is a subscript indicating the ith student and ε_i is a random error term. Consistent with previous studies, the dependent variable *ACADEMIC_PERFORMANCE* represents a student's academic success, which is measured as the percentage of points earned in the final exam (e.g., Eskew & Faley, 1988; Lento, 2018; Perera & Richardson, 2010; Voshaar et al., 2023, 2024).[26] The key independent variable of interest, *CONTRIBUTIONS*, ranges from 0 to

[26] A student's academic performance of at least 45% (54 points) in the final exam indicates a successful completion of the course.

3 and quantifies the number of wiki articles a student has contributed. In addition, we construct an alternative binary independent variable, *COMPR_CONTRIBUTOR*, which indicates whether a student has contributed to the wiki comprehensively. Specifically, *COMPR_CONTRIBUTOR* equals "1" if a student contributed more than one contribution to the wiki, and "0" otherwise. The alternative binary variable *COMPR_CONTRIBUTOR* helps us to distinguish between students who meaningfully engaged with the wiki-based accounting glossary and those who disengaged after an initial contribution or did not contribute at all.

We include control variables commonly used in prior accounting education literature. Moreover, we hand-collected additional control variables directly from the students regarding prior work experience in accounting jobs and their English language proficiency.

First, we control for the student age (*AGE*). The literature provides mixed results regarding the effect of age on students' academic performance, with Guney (2009) reporting a positive relationship and Liu et al. (2013) reporting the opposite. Therefore, we do not have a clear expectation about the direction of the effect. Second, the relationship between gender and academic performance in accounting remains inconclusive, with studies reporting outperformance of females (e.g., Aldamen et al., 2015; Premuroso et al., 2011), outperformance of males (e.g., Ewelt-Knauer et al., 2025; Massoudi et al., 2017; Tan & Laswad, 2015; Voshaar et al., 2023, 2025; Wecks et al., 2025), and finding insignificant effects (e.g., Papageorgiou & Halabi, 2014). Consequently, the ex-ante effect of *FEMALE* – coded as "1" for female students and "0" otherwise – is theoretically ambiguous. Third, given the positive link between prior and future academic performance shown in prior accounting education studies (e.g., Engel, 2018; Koh & Koh, 1999), we control for the student's university grade point average (*GPA*).[27] Fourth, besides *GPA*, we include *PRE_UNI_PERFORMANCE*, the share of tuition covered by merit-based entry scholarships. These reflect pre-university academic achievement and extracurricular engagement. This measure is relevant because first-year GPAs may not yet capture actual ability due to diverse backgrounds and adjustment to the university. *PRE_UNI_PERFORMANCE* also partly accounts for prior motivation, which may correlate with wiki engagement. Fifth, we include the program of study (*STUDY_PROGRAM*) as a fixed effect to control for variations in academic background, as

[27] All university course grades up to and including Spring 2023 are included, excluding the grade for the course under study.

supported by previous research (e.g., Duff, 2004; Jackson & Cossitt, 2015; Tan & Laswad, 2008, 2015; Voshaar et al., 2024, 2025; Wecks et al., 2025). Finally, we include the student region of origin (*REGION*) as a fixed effect to control for the socio-demographic background (Voshaar et al., 2023). Detailed variable definitions and measurements are presented in Appendix II-5.

For the qualitative part of our research question, we analyze semi-structured interviews thematically. Interviews were transcribed in real time. Thematic analysis was performed using MAXQDA, following Tello et al. (2026) in a two-phase sequence: deductive, then inductive. In the deductive phase, data were coded according to the predefined social constructivist dimension that informed our semi-structured interview guide. In the inductive phase, a line-by-line review assigned second-level codes to recurring patterns in student perceptions (Marks & Yardley, 2004). Two researchers independently coded the data in each phase, comparing results and resolving discrepancies through discussion informed by recent qualitative research in accounting education. A third, non-coding researcher provided an external perspective and participated in majority decisions when consensus was not reached, ensuring analytical rigor and agreement on the final coding scheme.

4. Results

4.1. Wiki-based Collaborative Learning and Academic Performance

4.1.1. Descriptive Statistics, Univariate Analysis, and Pairwise Correlations

Table II-1 presents the descriptive statistics for the dependent, independent, and control variables used in the regression analyses. The final sample comprises 107 students, with an average final exam score (*ACADEMIC_PERFORMANCE*) of about 66.5%.[28] On average, as measured by *CONTRIBUTIONS*, students participated in two of the three writing activities, with a standard deviation of one contribution, indicating contribution variability across the sample. 78 students (73%) contributed more than once, making them comprehensive contributors (*COMPR_CONTRIBUTOR*).

[28] Students passed the exam if they achieved at least 54 out of 120 points (45%). Non-contributors (N=18) averaged 63.3%, while contributors (N=89) averaged 67.2%.

Table II-1. Descriptive Statistics: Full Sample

Variables	N	Mean	Std. Dev.	Min	Max
ACADEMIC_PERFORMANCE	107	66.523	8.695	46	87
CONTRIBUTIONS	107	2.019	1.116	0	3
COMPR_CONTRIBUTOR	107	0.729	0.447		
AGE	107	19.991	1.186	18	22
FEMALE	107	0.467	0.501		
GPA	107	2.474	0.617	1.400	3.540
PRE_UNI_PERFORMANCE	107	0.262	0.122	0.000	0.400

Notes: This table presents the descriptive statistics of the variables used in our main OLS regression model. For binary variables, only means and standard deviations are presented. All variables are defined in Appendix II-5.

We conduct an analysis of variance (ANOVA) test to determine whether the level of wiki engagement impacts academic performance. We find a statistically significant difference in *ACADEMIC_PERFORMANCE* across levels of *CONTRIBUTIONS* ($F=3.25$; $p=0.025$), indicating a positive association between student wiki engagement and academic performance.[29]

Next, we conduct a two-tailed t-test to compare *ACADEMIC_PERFORMANCE* between comprehensive contributors, students who contributed to the wiki more than once (*COMPR_CONTRIBUTOR*), and noncomprehensive contributors. As shown in Table II-2, comprehensive contributors score about 5.5 percentage points higher on average, a statistically significant difference. Table II-2 also reports t-tests for the control variables. Aside from a marginally significant *GPA* difference, no other variables differ significantly. Thus, selection into either group does not appear systematically driven by underlying student characteristics.

Table II-2. Descriptive Statistics by Comprehensive and Noncomprehensive Wiki Contributors

	Comprehensive Contributor		Noncomprehensive Contributor		
Variables	N	Mean	N	Mean	| *Diff.* |
ACADEMIC_PERFORMANCE	78	68.013	29	62.517	5.4956***
AGE	78	19.885	29	20.276	0.3912
FEMALE	78	0.487	29	0.414	0.0734
GPA	78	2.536	29	2.310	0.2252*

[29] Specifically, a Dunnett's test reveals a statistically significant difference between non-contributors and full contributors ($p<0.05$).

Variables	Comprehensive Contributor N	Comprehensive Contributor Mean	Noncomprehensive Contributor N	Noncomprehensive Contributor Mean	| *Diff.* |
PRE_UNI_PERFORMANCE	78	0.265	29	0.252	0.0137

Notes: This table presents the descriptive statistics and the results from a two-tailed t-test for differences in the means of the demographics and background variables for the full sample distinguished by wiki engagement. The differences between comprehensive and noncomprehensive wiki contributors are reported as absolute values. ***, **, * indicate statistical significance at the 0.01, 0.05, and 0.1 levels (two-tailed), respectively. All variables are defined in Appendix II-5.

Table II-3 presents the Pearson pairwise correlations. The results show statistically significant correlations between the dependent variable (*ACADEMIC_PERFORMANCE*) and the two independent variables of interest (*CONTRIBUTIONS* and *COMPR_CONTRIBUTOR*). Correlations among the control variables are below 0.6, thus do not indicate a risk of multicollinearity.

Table II-3. Pearson Correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) *ACADEMIC_ PERFORMANCE*	1.0000						
(2) *CONTRI- BUTIONS*	0.2625***	1.0000					
(3) *COMPR_ CONTRIBUTOR*	0.2822***	0.9001***	1.0000				
(4) *AGE*	-0.1313	-0.1354	-0.1474	1.0000			
(5) *FEMALE*	-0.0415	0.0854	0.0654	-0.2148**	1.0000		
(6) *GPA*	0.4846***	0.1753*	0.1629*	-0.1399	0.2190**	1.0000	
(7) *PRE_UNI_ PERFORMANCE*	0.2990***	0.0470	0.0501	-0.3781***	0.2844***	0.3087***	1.0000

Notes: This table reports Pearson correlations. ***, **, * indicate statistical significance at the 0.01, 0.05, and 0.1 levels (two-tailed), respectively. All variables are defined in Appendix II-5.

4.1.2. Regression Results

To evaluate the impact of wiki engagement on academic performance, we run an OLS regression with fixed effects (Equation II-1). Results in Table II-4, Model 1, show that each additional contribution is associated with a 1.3-percentage-point increase in the final exam score. To test robustness and provide further insight, we replace *CONTRIBUTIONS* with the binary variable *COMPR_CONTRIBUTOR*, indicating whether a student contributed more than once. The positive, statistically significant coefficient ($p<0.05$) suggests that comprehensive wiki engagement corresponds to a 4.6-percentage-point increase in the final exam score, consistent with the initial findings. This effect is also pedagogically meaningful, roughly equivalent to a two-grade-level difference.[30]

For the control variables, we note that *AGE* does not affect *ACADEMIC_PERFORMANCE*. Second, our results indicate that *FEMALE* students score significantly lower ($p<0.05$) in the final exam. These findings align with prior and most recent studies reporting negative effects (Ewelt-Knauer et al., 2025; Massoudi et al., 2017; Tan & Laswad, 2015; Voshaar et al., 2023, 2025; Wecks et al., 2025). Nonetheless, given the mixed evidence in accounting education research (Aldamen et al., 2015; Papageorgiou & Halabi, 2014; Premuroso et al., 2011), our results are unlikely to be broadly generalizable. Third, we observe a positive and statistically significant ($p<0.01$) coefficient estimate of *GPA*, supporting the notion of previous academic performance being one of the most substantial determinants of subsequent academic performance (Fortin et al., 2019; Hu et al., 2024). Finally, in line with previous literature, we find that the coefficient estimate of *PRE_UNI_PERFORMANCE* is positive and statistically significant ($p<0.05$), indicating that students' achievement in secondary education and pre-university extracurricular engagement is a suitable predictor of future academic success (e.g., Voshaar et al., 2023, 2024).

Table II-4. Wiki Engagement and Academic Performance (OLS)

		ACADEMIC_PERFORMANCE	
Variables	Expected Sign	Model 1	Model 2
CONTRIBUTIONS	+	**1.339****	
		(2.049)	
COMPR_CONTRIBUTOR	+		**4.608****
			(2.615)
AGE	?	0.397	0.501
		(0.599)	(0.766)
FEMALE	?	-3.108**	-3.054**
		(-2.141)	(-2.167)
GPA	+	6.488***	6.361***
		(5.646)	(5.519)
PRE_UNI_PERFORMANCE	+	15.026***	14.769**
		(2.689)	(2.602)
Constant		37.358***	34.975**
		(2.666)	(2.482)

[30] The results remain statistically significant when using alternative measures of wiki engagement as independent variables. In addition to a binary variable equal to 1 if a student contributed at least one wiki article and 0 otherwise (participation measure), we also tested a quality measure based on the bonus points awarded for the wiki contributions, with up to 20 points available depending on quality.

Variables	Expected Sign	ACADEMIC_PERFORMANCE Model 1	Model 2
N		107	107
R^2		0.469	0.489
Adj. R^2		0.395	0.418
STUDY_PROGRAM FE		Yes	Yes
REGION FE		Yes	Yes

Notes: This table presents an overview of the OLS regression results estimating the relationship between students' wiki engagement, their academic performance, and several control variables. Bold font indicates the variables of interest. ***, **, * indicate statistical significance at the 0.01, 0.05, and 0.1 levels (two-tailed), respectively. The robust t-statistics are presented in parentheses. All variables are defined in Appendix II-5.

We use the Ramsey-RESET test (Ramsey, 1969) to check for model misspecification. The null hypothesis of correct specification is not rejected, suggesting no evidence of omitted variables. To further address potential omitted variable bias, we add additional controls hand-collected from the students. Following recent accounting education research (e.g., Choo et al., 2026), we include *WORK_EXPERIENCE*, indicating whether students worked in an accounting-related role before or during the intervention semester, and *ENGLISH_PROFICIENCY*, based on native language status and whether secondary education was primarily in English. Due to data limitations, these variables substantially reduce the sample size and are therefore excluded from the main regression. In Table II-5, we first add each control separately, then both together. In all cases, *CONTRIBUTIONS* remains significantly positive, confirming the robustness of its relationship with academic performance and reducing concerns about omitted variable bias.[31]

Table II-5. Wiki Engagement and Academic Performance (OLS) with Additional Controls

Variables	Expected Sign	ACADEMIC_PERFORMANCE Model 1	Model 2	Model 3
CONTRIBUTIONS	+	**1.255***	**1.268***	**1.267***
		(1.830)	**(1.834)**	**(1.827)**
AGE	?	0.823	0.871	0.868
		(1.109)	(1.093)	(1.080)
FEMALE	?	-3.549**	-4.166**	-4.213**
		(-2.265)	(-2.559)	(-2.543)
GPA	+	6.941***	7.058***	7.037***
		(5.939)	(5.690)	(5.666)

[31] These results hold for all models when using the alternative independent variable *COMPR_CONTRIBUTOR*.

Variables	Expected Sign	ACADEMIC_PERFORMANCE Model 1	Model 2	Model 3
PRE_UNI_PERFORMANCE	+	13.814**	12.942**	13.011**
		(2.371)	(2.174)	(2.204)
WORK_EXPERIENCE	+	0.577		-0.634
		(0.213)		(-0.207)
ENGLISH_PROFICIENCY	+		0.426	0.443
			(0.419)	(0.432)
Constant		28.566*	27.623*	27.793*
		(1.799)	(1.687)	(1.676)
N		93	89	89
R^2		0.493	0.521	0.522
Adj. R^2		0.402	0.431	0.423
STUDY_PROGRAM FE		Yes	Yes	Yes
REGION FE		Yes	Yes	Yes

Notes: This table presents an overview of the OLS regression results estimating the relationship between students' wiki engagement, their academic performance, the control variables from Equation II-1, and additional control variables. Bold font indicates the variable of interest. ***, **, * indicate statistical significance at the 0.01, 0.05, and 0.1 levels (two-tailed), respectively. The robust t-statistics are presented in parentheses. All variables are defined in Appendix II-5.

4.1.3. Robustness Checks

In additional robustness checks, we address potential endogeneity by applying entropy balancing (EB) and propensity score matching (PSM) to reduce estimation bias from self-selection. The t-test results in Table II-2 show a marginally significant difference ($p<0.1$) in prior university success between comprehensive wiki contributors and noncomprehensive contributors, suggesting that academic ability may influence participation in the teaching interventions. This potential self-selection, where higher-achieving students are both more likely to participate and perform better, could bias our initial results. To mitigate this, we balance covariates using EB and PSM to compare only similar students.

Following prior accounting education research (e.g., Wecks et al., 2025), we first use EB (Hainmueller, 2012) based on the covariates in Equation II-1. Specifically, we conduct entropy balancing for continuous treatments (EBCT) with our main variable of interest (*CONTRIBUTIONS*), using weights that remove Pearson correlations between covariates and the treatment (Tübbicke, 2022).[32] We also apply EB to *COMPR_CONTRIBUTOR*, assigning

[32] EBCT diagnostics confirmed that all covariates were balanced. After weighting, the regression of the treatment on covariates yielded $R^2=0.000$, $F=0.000$, and $p=1.000$, indicating that observable covariate differences were eliminated.

weights so that treatment and control groups have identical means, variances, and skewness for all covariates.[33] The regression results in Table II-6, which employed entropy balancing, are consistent with our primary regressions, confirming the initial findings.

Next, consistent with prior studies (e.g., Hu et al., 2024; Miihkinen & Virtanen, 2018), we implement PSM. Using a generalized propensity score (GPS) approach, we estimate an ordered logit model with *CONTRIBUTIONS* as the dependent variable and the Equation II-1 covariates as predictors. The predicted probabilities yield inverse probability weights for a weighted outcome regression estimating the effect of *CONTRIBUTIONS* on *ACADEMIC_PERFORMANCE*, controlling for covariates and absorbing *STUDY_PROGRAM* and *REGION* fixed effects. As shown in Table II-6, Model 3, the coefficient on *CONTRIBUTIONS* remains positive and significant ($p<0.05$), supporting our initial findings.

We then perform nearest-neighbor one-to-one PSM for *COMPR_CONTRIBUTOR* using a probit model to calculate propensity scores (Rosenbaum & Rubin, 1983), defined as the probability of contributing more than once to the wiki. Covariate balance tests show that, aside from *AGE* ($p<0.1$), differences are statistically insignificant, indicating largely effective matching. Re-estimating Equation II-1 with the matched sample (Table II-6, Model 4) yields a positive, significant coefficient for *COMPR_CONTRIBUTOR* ($p<0.05$), further confirming our results.

Table II-6. Robustness Check Results Using Entropy Balancing and Propensity Score Matching

		ACADEMIC_PERFORMANCE			
Variables	Expected Sign	Model 1 (EBCT)	Model 2 (EB)	Model 3 (GPS)	Model 4 (PSM)
CONTRIBUTIONS	+	**1.118***		**1.649****	
		(1.899)		**(2.611)**	
COMPR_CONTRIBUTOR	+		**3.186****		**6.398*****
			(2.232)		**(3.242)**
AGE	?	0.214	0.276	-0.222	0.247
		(0.330)	(0.531)	(-0.341)	(0.273)
FEMALE	?	-2.767**	-3.649***	-3.249*	-4.731**
		(-1.996)	(-2.633)	(-1.882)	(-2.266)
GPA	+	6.496***	4.668***	5.755***	5.796***
		(5.414)	(4.132)	(4.915)	(3.196)

[33] For the binary covariate *FEMALE*, only the mean is balanced. Untabulated results indicate that the covariate balance was achieved.

Variables	Expected Sign	ACADEMIC_PERFORMANCE Model 1 (EBCT)	Model 2 (EB)	Model 3 (GPS)	Model 4 (PSM)
PRE_UNI_PERFORMANCE	+	13.717**	19.090***	13.516**	15.907**
		(2.338)	(3.103)	(2.426)	(2.071)
Constant		41.867***	44.266***	51.136***	40.734**
		(3.061)	(3.886)	(3.911)	(2.136)
N		107	107	107	58
R^2		0.457	0.533	0.438	0.509
Adj. R^2		0.381	0.426	0.360	0.364
STUDY_PROGRAM FE		Yes	Yes	Yes	Yes
REGION FE		Yes	Yes	Yes	Yes

Notes: This table presents an overview of the OLS regression results with an entropy-balanced (EBCT; EB) sample, a matched sample using generalized propensity scores (GPS), and a matched sample using propensity scores (PSM), estimating the relationship between students' wiki engagement, their academic performance, and several control variables. Bold font indicates the variables of interest. ***, **, * indicate statistical significance at the 0.01, 0.05, and 0.1 levels (two-tailed), respectively. The robust t-statistics are presented in parentheses. All variables are defined in Appendix II-5.

4.2. Student Perceptions of Wiki-based Collaborative Learning

Table II-7 presents an introductory overview of the sub-themes that we identified regarding students' perceptions of wiki-based collaborative learning.

Table II-7. Overview of Qualitative Results: Perceptions on Student Collaboration, Instructor Scaffolding, and Learning Experiences

Central themes	Sub-themes	Key findings
Student collaboration	Learning ownership	Students embrace the individual writing activity, taking ownership of their work and valuing the autonomy it provides over structure and content.
	Peer review	Students value peer review for exposing them to diverse perspectives and enhancing clarity and critical thinking, though some question the depth and reliability of feedback.
	Free-Riding	Students frequently encounter free-riding, which undermines collaboration and prompts many to prefer individual work to maintain control and fairness.
Instructor scaffolding	Instructional scaffolds	Students rely on the contribution guidelines to structure their work and focus on key concepts, though some feel constrained by their prescriptiveness.
	Resource scaffolds	Students use the peer review template to guide evaluations, standardize expectations, and deliver focused, constructive feedback.
	Expert scaffolds	Students consult exemplary contributions to model their writing, clarify expectations, and align their work with academic standards.
Learning experiences	Affective experiences	Students describe the wiki as a motivating and rewarding tool, though unclear entries sometimes trigger frustration and reduce trust in peer-created content.

Central themes	Sub-themes	Key findings
	Behavioral experiences	Students engage with the wiki strategically, especially during exam preparation, using it to reinforce their understanding of accounting fundamentals.
	Environmental experiences	Students appreciate the wiki platform's structured interface but report technical and usability issues that occasionally limit their wiki engagement.

Notes: This table presents an overview of the sub-themes identified through the thematic analysis of the semi-structured interviews. The semi-structured interview guide is presented in Appendix II-4.

4.2.1. Student Collaboration

Learning ownership

A recurring sub-theme in the interviews is participants' sense of ownership over their learning, especially when comparing individual and collaborative writing tasks (Interviewees 1, 3). They describe strong ownership of their individual contributions to the collaboratively developed accounting glossary, feeling greater control over both the writing process and outcome (Interviewee 16). This autonomy enabled independent decision-making without needing approval or accommodating differing views (Interviewees 2, 9). Overall, participants find this self-reliance empowering, as reflected in the following quote:

> (…) the biggest thing I liked about it [the individual writing activity] was that it was totally my responsibility to do it. I could do things the way I wanted them to be, and I didn't have to check if my partner liked it or not. You just do it, and you can double-check yourself. This way, you take responsibility. (Interviewee 1)

Peer review

Comparing individual and collaborative writing enables a nuanced understanding of students' perceptions of peer collaboration. While individual writing promotes learning ownership, many interviewees note that the lack of collaboration reduces exposure to alternative viewpoints (Interviewees 2, 4, 6, 9, 10, 13, 16) that could have enriched their wiki contributions. In addition, the benefits of "proofreading" (Interviewee 6) are frequently cited as essential to collaborative writing, with several participants reporting that partner feedback can improve contribution quality (Interviewees 1, 5, 6, 10, 15). For example:

> (…) that was really good for me because I think I like to learn from other people and understanding their perspective and what people have to say. (…) There's always room for improvement (…) (Interviewee 6)

Peer review is not only viewed positively by all students. Some question the credibility of feedback due to peers' limited accounting expertise (Interviewee 2). Feedback delays occasionally reduced revision time (Interviewee 13). Conversely, writing reviews can deepen students' understanding of topics and enrich their contributions (Interviewees 5, 14, 15). Despite variations in quality, peer review is generally seen as valuable for fostering collaborative learning, critical thinking, and content revision. Beyond improving wiki contributions, several students note that collaboration promoted meaningful interaction (Interviewees 11, 16) and broadened understanding of accounting concepts through exposure to diverse cultural and academic perspectives (Interviewee 14).

Free-Riding

Despite the benefits of collaboration, free-riding is a major concern. Some students value the division of labor (Interviewees 1–3, 7) for reducing individual workload (Interviewees 1, 6, 15). In contrast, others faced unequal contributions, with partners contributing little and leaving them to complete most of the work alone (Interviewees 2, 5, 15). For example:

> (…) I would have preferred to do it [the collaborative wiki contribution] individually because I ended up doing most of the work. (Interviewee 2)

Students often report coordination difficulties, typically due to unresponsive or disengaged partners (Interviewees 2, 7). As a result, some prefer individual writing, which avoids unreliable partners and ensures control over their work (Interviewee 11).

Overall, the findings align with social constructivism, which holds that knowledge is co-constructed through dialogue and engagement with peers (Vygotsky, 1978; Bozkurt, 2017). Students' positive and critical views of peer collaboration illustrate how interaction exposes them to alternative perspectives, promotes reflection, and deepens learning. While some preferred individual control, most recognized interaction as essential for refining and extending understanding. Collaborative writing and peer review thus demonstrate how learning is shaped by social context and co-construction, as emphasized in social constructivism theory.

4.2.2. Instructor Scaffolding

Instructional scaffolds

Interview participants consistently view the contribution guidelines as an essential scaffold, providing a clear, practical structure for the writing process (Interviewees 1–3, 5–7, 10, 16). Beyond that, the guidelines enhance conceptual clarity. Although some students initially questioned their specificity, they later valued the framework for focusing attention on key aspects of the contribution (Interviewees 2, 5). From a cognitive load theory perspective, the step-by-step design helped reduce extraneous load by filtering extensive information and guiding students toward more focused glossary contributions (Sweller, 2010). For example:

> The first time I read it [the contribution guidelines], I wondered why they were being so specific. But then (…) I thought, yeah, this makes sense. (…) it allowed us to focus on simple things rather than start going crazy on all the aspects of the concept [the accounting concept]. (Interviewee 2)

Resource scaffolds

The review template is widely regarded as a critical scaffold, guiding students through the review process. A recurring sub-theme is its role in reducing uncertainty by providing a clear framework for evaluating wiki contributions. Several participants note initial concerns about the subjectivity of peer review and the difficulty of assessment without explicit guidelines (Interviewees 1, 9, 13), concerns that were alleviated by the template, as illustrated below:

> I didn't know how we would provide a review because I didn't know about the existence of that table [the review template], and I was just (…) thinking it would be way too subjective, right? (…) And then the professor [the instructor] sent us a video of how we would have to provide reviews. That table is actually really helpful, and it covered a lot of points. (Interviewee 1)

The review template enhances clarity and serves as a scaffold for reinforcing academic writing expectations. Several participants note that the structured checklist helped them identify key elements of a well-developed wiki article (Interviewees 7, 12, 16). By standardizing evaluation criteria, it fostered a shared understanding of academic validity, content credibility, and contribution structure. Yet, the interviewees suggest areas for improvement, such as clarifying the review process, encouraging more detailed feedback, adding summary or rating sections, and making the format more user-friendly (Interviewees 4, 9, 10, 12–15).

Expert scaffolds

Exemplary wiki contributions are often cited as valuable expert scaffolds for clarifying writing task expectations, especially at the outset when participants were unsure how to begin (Interviewees 2, 5–7, 10, 11, 15). By modeling best practices, they guided students on content structure, length, and citation formatting, offering clearer expectations than guidelines alone (Interviewees 10, 15, 16). While guidelines specify formal requirements, the examples demonstrate their practical application. For example:

> What I really liked were the exemplary contributions because just reading the guidelines [the contribution guidelines] still didn't give me a clear idea of what was wanted, but looking at the exemplary contributions really helped because I had some idea of what to do. (Interviewee 15)

Students used the examples as benchmarks for content quality, including word count, citation placement, and structure (Interviewees 2, 4, 7, 10). They can guide students in identifying key points and determining what to omit within word limits. Some students sought a broader range of exemplary articles to reflect diverse topics and styles (Interviewee 16), particularly more calculation-based examples alongside textual explanations of fundamental accounting concepts. Students regard these examples as essential support, providing clear, structured references to understand expectations, organize work, refine writing, and meet academic standards.

Overall, students value structured contribution guidelines, peer review templates, and exemplary contributions. These features reduce ambiguity, focus attention, and uphold academic standards. This aligns with social constructivism theory, which views the instructor as a facilitator who supports learning by providing guidance (Adams, 2006; Powell & Kalina, 2009). This kind of support enables meaningful learning and reinforces the theory that learners gain autonomy through guided practice.

4.2.3. Learning Experiences

Affective experiences

Students' emotional responses to the wiki-based collaborative learning approach reflect appreciation, motivation, and frustration. Many find the experience personally rewarding, citing satisfaction from engaging with unfamiliar terms, broadening their understanding, and reinforcing concepts from lectures or exams (Interviewees 1, 6, 8, 11, 16). For some, the wiki fosters learning progress by presenting new terminology in an accessible, student-centered format. However, unclear, complex, or incoherent entries caused frustration, raising doubts about reliability and reducing trust in peer-created content. Consequently, some relied on lecture slides or external materials instead (Interviewees 2, 5, 10, 13). Despite these issues, many still value the wiki-based accounting glossary as a selective learning aid. For example:

> I think (…), it's a learning process and really helpful for your learning. (…) especially when you challenge yourself because some words [the accounting concepts] are really difficult to understand. (…) sometimes it [the wiki] helps you remember the words. So if it [the accounting concept] comes up on the exam, you would actually remember it because you studied it by heart. (Interviewee 6)

Behavioral experiences

Students engaged with the wiki in varied, strategic ways shaped by learning styles, time constraints, and exam preparation. Some used it regularly to check definitions or review difficult terms (Interviewees 1, 4, 6, 7, 11), while others focused on articles they authored or those linked to reviewed reference words. Many valued the wiki for reinforcing learning before finals, though some noted information overload (Interviewees 2, 5, 9, 12, 15). For most, the wiki-based accounting glossary serves best as a supplementary study resource. For example:

> Let's say I was studying and I come across a term [the accounting concept] that I want more insight about, (…) I remember checking that specifically on the wiki contribution and I was able to understand the term better after looking at the wiki contributions. (…) That was pretty good for me. (Interviewee 11)

Environmental experiences

Students' experiences with the wiki's technical and organizational environment, particularly its Moodle integration, are mixed. Some value its structured format and found the editing and review processes intuitive (Interviewees 2, 6, 11), while others struggled with navigation and accessibility (Interviewees 4, 13, 14). Common issues include repeated logins, poor search functionality, and difficulty locating features. Such barriers sometimes discouraged wiki engagement (Interviewees 2, 5, 15). Although these problems did not stop usage entirely, students consistently note that a more user-friendly interface could further enrich students' learning experiences.

Overall, reflections on wiki engagement align with principles of social constructivism theory, where learning is personally constructed through meaningful, socially mediated activities (Swan, 2005; Barak, 2017). Peer-created content, the balance of individual and collaborative work, and even emotional reactions to usability illustrate how tools and peers can shape students' learning experiences. Differences in strategies and participation further underscore that knowledge construction is an active, personal, and social process.

5. Discussion, Conclusion, and Limitations

5.1. Discussion

Our quantitative results show that greater wiki engagement is associated with enhanced academic performance. The qualitative results help explain this effect. For example, consistent with social constructivism, students report that peer feedback and exposure to diverse perspectives (Interviewees 1, 2, 4–6, 9–11, 13–16), as well as instructor scaffolds, deepened their understanding of accounting terminology (Interviewees 1–3, 5–7, 10, 15, 16).

However, the qualitative findings also reveal some challenges that require attention. First, while student collaboration enables feedback and diverse views, many prefer individual writing for control and coherence. Second, a major issue is free-riding, which undermines fairness and teamwork. Third, although students value structured tools like guidelines and templates, some find them too rigid and suggest more flexible formats and better support for qualitative feedback. Finally, the wiki is considered a helpful exam preparation resource. However, concerns about reliability, inconsistency, information overload, and technical problems hinder its wider use.

We outline lessons learned that may inspire accounting educators to refine wiki-based collaborative learning. First, a confirmation step before each writing activity could ensure only committed students proceed, reducing free-riding and promoting fairness. In line with Lambert et al. (2014), instructors could also use the platform's logging features to monitor contributions and support transparent, equitable assessment.[34] Second, an initial in-class collaboration session can foster early communication, align goals, and build group cohesion, as emphasized by Interviewees 2, 4–11, 13–16. Third, to support varied collaboration strategies and reduce uncertainty, instructors could provide short guides outlining alternative collaboration models. Fourth, incorporating expert reviews would validate content and increase trust in wiki contributions (Interviewees 1–4, 6, 7, 10–14, 16). Fifth, linking wiki content to lecture and tutorial materials could enhance accessibility and promote sustained use, as highlighted by Interviewees 1, 2, 4–13, 15, 16. Finally, during our intervention semester, AI tools were not yet widely available, frequently exhibited hallucination issues, and lacked features such as internet connectivity or advanced reasoning modes. Over the past two years, however, their capabilities and application areas have advanced substantially. At the same time, the number of available models has increased, with many universities now providing premium access to proprietary tools through collaborations with providers such as OpenAI or Google, or developing in-house solutions to address concerns around ethics, data security, and copyright. These developments need to be considered when integrating wiki-based approaches into accounting education. Potential strategies may include requiring students to submit an AI usage

[34] However, in our setting, students often drafted content in Word and later transferred it to the wiki, undermining the accuracy of the wiki logs. Targeted training in the use of the wiki interface would be necessary to ensure meaningful log data and thus a fair assessment.

reflection and documentation report, which could help reinforce responsible AI use, uphold academic integrity, and foster AI literacy in higher education.

5.2. Conclusion

This study set out to address the research question: How does wiki-based collaborative learning affect the academic performance of undergraduate students in an introductory accounting course, and how do students perceive the wiki approach? Drawing on a mixed-methods research design, we investigate the impact on academic performance and explore student perceptions associated with the implementation of a wiki-based collaborative learning approach. First, quantitative analysis reveals that increased wiki engagement significantly enhances academic performance, suggesting that the collaborative building of a wiki-based accounting glossary can be an effective mechanism for reinforcing accounting knowledge. Second, we complement these findings by providing qualitative evidence on how students perceive the wiki-based accounting glossary. We find that students value the opportunities for peer feedback, reflection, and knowledge exchange, but also note persistent challenges, such as free-riding, rigid instructional material, and concerns about the reliability of peer-generated content. We outline several lessons learned to inspire ways to strengthen participation, increase the credibility of student-generated content, and support collaboration. Our study makes three contributions to accounting education. First, it addresses a gap in the collaborative learning literature by examining wikis as a central tool for glossary-building. Second, it empirically evaluates the effectiveness of this wiki-based approach, responding to calls for research on IT-enabled collaboration. Third, it documents the wiki design and scaffolding materials, providing educators with a replicable framework.

5.3. Limitations

This study is subject to limitations. First, the wiki is implemented in a setting with mostly international students for whom English is a second language. While this may limit generalizability to native English-speaking cohorts, it reflects the growing internationalization of accounting education and enhances the relevance of our findings for diverse learning environments. Second, the wiki focuses on glossary creation, which suits undergraduate students acquiring foundational accounting terminology. Although less applicable to postgraduate courses, the framework is adaptable to more advanced, problem-oriented tasks. Third, the data is limited to one semester. Thus, the lessons learned regarding the wiki design have not yet been tested. Nevertheless, the sample size and analytical rigor support robust conclusions and meaningful contributions.

Future research could examine the academic performance effect and explore the perceptions of using wikis in postgraduate and other instructional contexts, and test the lessons learned regarding the wiki design across multiple semesters to evaluate broader applicability.

6. References

Adams, P. (2006). Demystifying Constructivism. In L. Tan Wee Hin & R. Subramaniam (Eds.), *Handbook of Research on Literacy in Technology at the K-12 Level* (pp. 493–514). IGI Global.

Ainsworth, J. (2021). Team-Based Learning in professional writing courses for accounting graduates: positive impacts on student engagement, accountability and satisfaction. *Accounting Education*, *30*(3), 234–257. https://doi.org/10.1080/09639284.2021.1906720

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://doi.org/10.1080/09639284.2015.1043563

Barak, M. (2017). Science Teacher Education in the Twenty-First Century: a Pedagogical Framework for Technology-Integrated Social Constructivism. *Research in Science Education*, *47*(2), 283–303. https://doi.org/10.1007/s11165-015-9501-y

Bayne, L., Birt, J., Hancock, P., Schonfeldt, N., & Agrawal, P. (2022). Best practices for group assessment tasks. *Journal of Accounting Education*, *59*, 100770. https://doi.org/10.1016/j.jaccedu.2022.100770

Biasutti, M. (2017). A comparative analysis of forums and wikis as tools for online collaborative learning. *Computers & Education*, *111*, 158–171. https://doi.org/10.1016/j.compedu.2017.04.006

Bozkurt, G. (2017). Social constructivism: Does it succeed in reconciling individual cognition with social teaching and learning practices in mathematics? *Journal of Education and Practice*, *8*(3), 210–218.

Camacho, M. E., Carrión, M. D., Chayah, M., & Campos, J. M. (2016). The use of wiki to promote students' learning in higher education (Degree in Pharmacy). *International Journal of Educational Technology in Higher Education*, *13*(1), 23. https://doi.org/10.1186/s41239-016-0025-y

Camarero, C., Rodríguez, J., & San José, R. (2012). An exploratory study of online forums as a collaborative learning tool. *Online Information Review*, *36*(4), 568–586. https://doi.org/10.1108/14684521211254077

Choo, N., Fieger, P., Wells, P., & Tseng, K. (2026). The influence of prior knowledge on the self-efficacy of students enrolled in a tertiary introductory accounting course. *Accounting Education*, *35*(1), 34–58. https://doi.org/10.1080/09639284.2024.2405603

Christensen, J., Harrison, J. L., Hollindale, J., & Wood, K. (2019). Implementing team-based learning (TBL) in accounting courses. *Accounting Education*, *28*(2), 195–219. https://doi.org/10.1080/09639284.2018.1535986

Chu, S. K. W., Zhang, Y., Chen, K., Chan, C. K., Lee, C. W. Y., Zou, E., & Lau, W. (2017). The effectiveness of wikis for project-based learning in different disciplines in higher education. *Internet and Higher Education*, *33*, 49–60. https://doi.org/10.1016/j.iheduc.2017.01.005

Cilliers, L. (2017). Wiki acceptance by university students to improve collaboration in higher education. *Innovations in Education and Teaching International*, *54*(5), 485–493. https://doi.org/10.1080/14703297.2016.1180255

Coetzee, S. A., Schmulian, A., & Coetzee, R. (2018). Web conferencing-based tutorials: student perceptions thereof and the effect on academic performance in accounting education. *Accounting Education*, *27*(5), 531–546. https://doi.org/10.1080/09639284.2017.1417876

Coetzee, S. A., Schmulian, A., & Janse van Rensburg, C. (2023). The use of mobile instant messaging for peer feedback in online assessments for learning. *Accounting Education*, *32*(5), 563–577. https://doi.org/10.1080/09639284.2023.2223580

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

De Arriba, R. (2017). Participation and collaborative learning in large class sizes: wiki, can you help me? *Innovations in Education and Teaching International*, *54*(4), 364–373. https://doi.org/10.1080/14703297.2016.1180257

De Klerk, C., Ker-Fox, J., & Steenekamp, L. (2025). Enhancing critical thinking through collaborative learning: the impact of a partial pre-release assessment format. *Accounting Education*, *34*(4), 499–532. https://doi.org/10.1080/09639284.2024.2361647

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

Engel, A. M. (2018). Literature Review of Student Characteristics and Performance in an Accounting Course. *Community College Journal of Research and Practice*, *42*(10), 748–751. https://doi.org/10.1080/10668926.2017.1328320

Eskew, R. K., & Faley, R. H. (1988). Some Determinants of Student Performance in the First College-Level Financial Accounting Course. *Accounting Review*, *63*(1), 137–147. https://www.jstor.org/stable/247684

Ewelt-Knauer, C., Herrmann, F., Mai, S., Pleger, A., & Schuetz, H. (2026). Keeping the balance: the impact of an instructional edutainment-based video series on students' motivation and performance. *Accounting Education*, *35*(4), 489–525. https://doi.org/10.1080/09639284.2025.2482690

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fortin, A., Viger, C., Deslandes, M., Callimaci, A., & Desforges, P. (2019). Accounting students' choice of blended learning format and its impact on performance and satisfaction. *Accounting Education*, *28*(4), 353–383. https://doi.org/10.1080/09639284.2019.1586553

Garcia, C., & Privado, J. (2023). Predicting cooperative work satisfaction of autonomous groups using a wiki tool in higher education. *Interactive Learning Environments*, *31*(1), 117–128. https://doi.org/10.1080/10494820.2020.1764590

Gielen, M., & De Wever, B. (2015). Scripting the role of assessor and assessee in peer assessment in a wiki environment: Impact on peer feedback quality and product improvement. *Computers & Education*, *88*, 370–386. https://doi.org/10.1016/j.compedu.2015.07.012

Gomez, E. A., Wu, D., & Passerini, K. (2010). Computer-supported team-based learning: The impact of motivation, enjoyment and team contributions on learning outcomes. *Computers & Education*, *55*(1), 378–390. https://doi.org/10.1016/j.compedu.2010.02.003

Goosen, R., & Steenkamp, G. (2025). How can course design facilitate the development of teamwork skills for diligent students? *Accounting Education*, *34*(3), 387–409. https://doi.org/10.1080/09639284.2024.2335224

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

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*, *20*(1), 25–46. https://doi.org/10.1093/pan/mpr025

Hechtner, F., Schmidt, L., Seebeck, A., & Weiß, M. (2025). How to Design and Employ Specialized Large Language Models for Accounting and Tax Research: The Example of TaxBERT. SSRN Working Paper. https://doi.org/10.2139/ssrn.5146523

Helliar, C. (2013). The Global Challenge for Accounting Education. *Accounting Education*, *22*(6), 510–521. https://doi.org/10.1080/09639284.2013.847319

Hong, D. (2024). Using in-class group assignments to improve low-achieving students' performance in an introductory accounting course. *Accounting Education*, *33*(1), 66–83. https://doi.org/10.1080/09639284.2022.2119090

Hu, Y., Nath, N., Zhu, Y., & Laswad, F. (2024). Accounting students' online engagement, choice of course delivery format and their effects on academic performance. *Accounting Education*, *33*(5), 649–684. https://doi.org/10.1080/09639284.2023.2254298

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A Large Language Model for Extracting Information from Financial Text. *Contemporary Accounting Research*, *40*(2), 806–841. https://doi.org/10.1111/1911-3846.12832

Huang, K. (2019). Design and investigation of cooperative, scaffolded wiki learning activities in an online graduate-level course. *International Journal of Educational Technology in Higher Education*, *16*(1), 11. https://doi.org/10.1186/s41239-019-0141-6

Ioannou, A., Brown, S. W., & Artino, A. R. (2015). Wikis and forums for collaborative problem-based activity: A systematic comparison of learners' interactions. *Internet and Higher Education*, *24*, 35–45. https://doi.org/10.1016/j.iheduc.2014.09.001

Jackson, M., & Cossitt, B. (2015). Is Intelligent Online Tutoring Software Useful in Refreshing Financial Accounting Knowledge? *Advances in Accounting Education*, 16, 1–19. https://doi.org/10.1108/S1085-462220150000016001

Kassem, M. A. M. (2017). Developing Business Writing Skills and Reducing Writing Anxiety of EFL Learners through Wikis. *English Language Teaching*, *10*(3), 151. https://doi.org/10.5539/elt.v10n3p151

Kelly, O., Hall, T., & Connolly, C. (2023). PACE-IT: designing blended learning for accounting education in the challenging context of a global pandemic. *Accounting Education*, *32*(6), 626–645. https://doi.org/10.1080/09639284.2022.2090851

Koh, M. Y., & Koh, H. C. (1999). The determinants of performance in an accountancy degree programme. *Accounting Education*, *8*(1), 13–29. https://doi.org/10.1080/096392899331017

Lambert, S. C., Carter, A. J., & Lightbody, M. (2014). Taking the Guesswork Out of Assessing Individual Contributions to Group Work Assignments. *Issues in Accounting Education*, *29*(1), 169–180. https://doi.org/10.2308/iace-50637

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

Lin, C., & Reigeluth, C. M. (2019). Scaffolding learner autonomy in a wiki-supported knowledge building community and its implications for mindset change. *British Journal of Educational Technology*, *50*(5), 2667–2684. https://doi.org/10.1111/bjet.12713

Liu, C. Z., Rowe, B. J., Serrett, R., & Shelton, M. L. (2013). Leveling the playing field: Can students succeed in highly technical business courses delivered online? *Global Perspectives on Accounting Education*, *10*, 135–149.

Luo, M. M., & Chea, S. (2020). Wiki use for knowledge integration and learning: A three tier conceptualization. *Computers & Education*, *154*, 103920. https://doi.org/10.1016/j.compedu.2020.103920

Lux, G., Callimaci, A., Caron, M.-A., Fortin, A., & Smaili, N. (2023). COVID-19 and emergency online and distance accounting courses: a student perspective of engagement and satisfaction. *Accounting Education*, *32*(2), 115–149. https://doi.org/10.1080/09639284.2022.2039729

Mali, D., & Lim, H. (2022). Can the introduction of a research-informed teaching intervention enhance student performance and influence perceptions? *Accounting Education*, *31*(3), 322–346. https://doi.org/10.1080/09639284.2021.2014914

Marks, D. F., & Yardley, L. (2004). Content and thematic analysis. In D. F. Marks & L. Yardley (Eds.), *Research methods for clinical and health psychology* (pp. 56–69). SAGE Publications.

Martins, A., Gomes, S., Pacheco, L., & Martins, H. (2024). Higher education students' perceptions of accounting online learning: the emergency context of the COVID-19 pandemic. *Accounting Education*, *33*(5), 731–765. https://doi.org/10.1080/09639284.2023.2244947

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

Matheson, R. (2018). Transition through the student lifecycle. In R. Matheson, S. Tangney, & M. Sutcliffe (Eds.), *Transition In, through and out of higher education* (pp. 5–16). Routledge.

Miihkinen, A., & Virtanen, T. (2018). Development and application of assessment standards to advanced written assignments. *Accounting Education*, *27*(2), 121–159. https://doi.org/10.1080/09639284.2017.1396480

Musundwa, S. (2025). Implementing constructivist teaching to foster inclusive educational practices in accounting programmes. *Accounting Education*, *34*(6), 913–946. https://doi.org/10.1080/09639284.2024.2367132

Nsor-Ambala, R. (2022). The impact of collaborative learning approaches on assessment outcomes in an accounting theory class. *Accounting Education*, *31*(1), 1–38. https://doi.org/10.1080/09639284.2021.1944232

Oosthuizen, H., De Lange, P., Wilmshurst, T., & Beatson, N. (2021). Teamwork in the accounting curriculum: stakeholder expectations, accounting students' value proposition, and instructors' guidance. *Accounting Education*, *30*(2), 131–158. https://doi.org/10.1080/09639284.2020.1858321

Opdecam, E., & Everaert, P. (2012). Improving Student Satisfaction in a First-Year Undergraduate Accounting Course by Team Learning. *Issues in Accounting Education*, *27*(1), 53–82. https://doi.org/10.2308/iace-10217

Opdecam, E., & Everaert, P. (2019). Choice-based learning: lecture-based or team learning? *Accounting Education*, *28*(3), 239–273. https://doi.org/10.1080/09639284.2019.1570857

Opdecam, E., Everaert, P., van Keer, H., & Buysschaert, F. (2014). Preferences for Team Learning and Lecture-Based Learning Among First-Year Undergraduate Accounting Students. *Research in Higher Education*, 55, 400–432. https://doi.org/10.1007/s11162-013-9315-6

Osgerby, J. (2013). Students' Perceptions of the Introduction of a Blended Learning Environment: An Exploratory Case Study. *Accounting Education*, *22*(1), 85–99. https://doi.org/10.1080/09639284.2012.729341

Page, K. L., & Reynolds, N. (2015). Learning from a wiki way of learning. *Studies in Higher Education*, *40*(6), 988–1013. https://doi.org/10.1080/03075079.2013.865158

Papageorgiou, K., & K. Halabi, A. (2014). Factors contributing toward student performance in a distance education accounting degree. *Meditari Accountancy Research*, *22*(2), 211–223. https://doi.org/10.1108/MEDAR-08-2013-0032

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. https://doi.org/10.1080/09639284.2010.529639

Peters, P., Smith, A., Middledorp, J., Karpin, A., Sin, S., & Kilgore, A. (2013). Learning essential terms and concepts in Statistics and Accounting. *Higher Education Research & Development*, *33*(4), 742–756. https://doi.org/10.1080/07294360.2013.863838

Powell, K. C., & Kalina, C. J. (2009). Cognitive and social constructivism: Developing tools for an effective classroom. *Education*, *130*(2), 241–250.

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. https://doi.org/10.2308/iace-50066

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, *31*(2), 350–371. https://doi.org/10.1111/j.2517-6161.1969.tb00796.x

Reigeluth, C. M. (1979). In Search of a Better Way to Organize Instruction: The Elaboration Theory. *Journal of Instructional Development*, *2*(3), 8–15. https://www.jstor.org/stable/30220561

Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Rott, S., & Weber, E. D. (2013). Preparing Students to use Wiki Software as a Collaborative Learning Tool. *CALICO Journal*, *30*(2), 179–203. https://doi.org/10.11139/cj.30.2.179-203

Salaber, J. (2014). Facilitating student engagement and collaboration in a large postgraduate course using wiki-based activities. *The International Journal of Management Education*, *12*(2), 115–126. https://doi.org/10.1016/j.ijme.2014.03.006

Sangster, A., Stoner, G., & Flood, B. (2020). Insights into accounting education in a COVID-19 world. *Accounting Education*, *29*(5), 431–562. https://doi.org/10.1080/09639284.2020.1808487

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & Quantity*, *52*(4), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8

Seebeck, A., & Wolter, R. M. (2022). Financial education of founders, is it important? A case study of Jacobs Startup Competition. *International Journal of Entrepreneurial Venturing*, *14*(1), 122–143. https://doi.org/10.1504/IJEV.2022.122018

Shawver, T. J. (2020). An experimental study of cooperative learning in advanced financial accounting courses. *Accounting Education*, *29*(3), 247–262. https://doi.org/10.1080/09639284.2020.1736589

Swan, K. (2005). A constructivist model of thinking about learning online. In J. Bourne & J. Moore (Eds.), *Elements of quality online education: Emerging communities* (pp. 13–30). Sloan-C.

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 29–47). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.004

Tan, H. C. (2019). Using a structured collaborative learning approach in a case-based management accounting course. *Journal of Accounting Education*, *49*, 100638. https://doi.org/10.1016/j.jaccedu.2019.100638

Tan, L. M., & Laswad, F. (2008). Impact of prior content and meta-cognitive knowledge on students' performance in an introductory accounting course. *Pacific Accounting Review*, *20*(1), 63–74. https://doi.org/10.1108/01140580810872852

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

Tello, E., Considine, B., Rotaru, K., & Fehrenbacher, D. (2026). Implementing online discussion boards to enhance professional skills and news reading habits in accounting students. *Accounting Education*, *35*(2), 137–169. https://doi.org/10.1080/09639284.2024.2419578

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Open University Press.

Tübbicke, S. (2022). Entropy Balancing for Continuous Treatments. *Journal of Econometric Methods*, *11*(1), 71–89. https://doi.org/10.1515/jem-2021-0002

Turner, M., & Baskerville, R. (2013). The Experience of Deep Learning by Accounting Students. *Accounting Education*, *22*(6), 582–604. https://doi.org/10.1080/09639284.2013.847323

Voshaar, J., Johannsen, F., Mkervalidze, S., & Zimmermann, J. (2024). Unbundling the App Advantage: Evaluating Exam Performance-enhancing Features of Mobile Learning Apps in Accounting. *Proceedings of the International Conference on Information Systems (ICIS) 2024*. https://aisel.aisnet.org/icis2024/learnandiscurricula/learnandiscurricula/4

Voshaar, J., Knipp, M., Loy, T., Zimmermann, J., & Johannsen, F. (2023). The impact of using a mobile app on learning success in accounting education. *Accounting Education*, *32*(2), 222–247. https://doi.org/10.1080/09639284.2022.2041057

Voshaar, J., Wecks, J. O., Plate, B. J., & Zimmermann, J. (2025). Tackling Professorial Expert Bias: The Role of ChatGPT in Simplifying Financial Accounting Exam Texts. *Issues in Accounting Education*, *40*(1), 93–123. https://doi.org/10.2308/ISSUES-2023-091

Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press.

Wecks, J. O., Voshaar, J., Plate, B. J., & Zimmermann, J. (2025). Generative AI Usage and Exam Performance. *Proceedings of the European Conference on Information Systems (ECIS) 2025*. http://dx.doi.org/10.2139/ssrn.4812513

World Medical Association. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*, 310(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Wyness, L., & Dalton, F. (2018). The value of problem-based learning in learning for sustainability: Undergraduate accounting student perspectives. *Journal of Accounting Education*, *45*, 1–19. https://doi.org/10.1016/j.jaccedu.2018.09.001

Zheng, B., Niiya, M., & Warschauer, M. (2015). Wikis and collaborative learning in higher education. *Technology, Pedagogy and Education*, *24*(3), 357–374. https://doi.org/10.1080/1475939X.2014.948041

7. Appendices

7.1. Appendix II-1. Wiki Contribution Guidelines

Wiki contribution guidelines
1. The wiki contribution must have a length between 500 and 800 words overall (including in-text citations, excluding references). The use of embedded media of all kinds is not permitted.
2. Each wiki contribution must include at least 3 academic sources. For example, academic journals and academic books. Investopedia and similar is not an academic source.
3. At the beginning of the wiki contribution, write a stand-alone, source-based, one-sentence definition. Next, provide a more detailed, source-based definition (about 2-3 sentences).
4. Provide at least one example. The example is intended to help the reader understand the relevance of the term. Examples may include accounting records, research questions in published literature, case studies, etc.
5. At least 5 unique links need to be made to related wiki contributions.
6. Each wiki contribution is required to provide one multiple-choice question and one open-ended question prior to the list of references. The multiple-choice question should contain five possible answers, of which one up to all five may be correct. The open-ended question should be capable of being answered with a maximum of 2-3 sentences.
7. The word list term to be described is to be formatted in bold throughout the wiki contribution.
8. Sources used must be cited in-text using "APA Formatting and Style Guide (7th Edition)" following the author-date method.
9. At the end of the wiki contribution, there must be a list of references using "APA Formatting and Style Guide (7th Edition)." Alphabetize entries by authors' last names. If available, the DOI/Link must be provided. Please check finally if the link leads to the corresponding source. All sources cited in the text must be listed in the list of references and vice versa.
10. The wiki contribution must end with: "Back [[Homepage]]."

Notes: The wiki contribution guidelines were made available within the wiki platform and reinforced during an in-person instructional session.

7.2. Appendix II-2. Exemplary Wiki Contribution

Balance sheet

The **balance sheet** is a financial report that presents a corporation's assets, liabilities, and owner's equity at a specific point in time (Phillips et al., 2021).

The **balance sheet** is based on the accounting equation and, as its name suggests, illustrates the relationship between a company's assets, liabilities, and owner's equity. It shows what the business owns (reported as assets) at a specific point in time - typically at the end of the reporting period - and how those assets are financed by creditors (reported as liabilities) and stockholders (reported as owner's equity). The **balance sheet** is one of the four primary financial statements, along with the income statement, cash flow statement, and owner's equity statement (Phillips et al., 2021; Weygandt et al., 2020; White et al., 2005).

The following is an example of a **balance sheet**:

Balance Sheet for FY 01 (in $)

	Assets			**Liabilities and Owner's Equity**	
Current Assets			Current Liabilities		
	Cash	11,874		Accounts payable	8,060
	Accounts receivable			Short-term loans	
	Inventory			Income taxes payable	3,145
	Prepaid expenses			Accrued salaries and wages	
	Short-term investments			Current portion of long-term debt	
	Total current assets	11,874		Total current liabilities	11,205
Non-Current Assets			Long-Term Liabilities		
	Long-term investments	1,208		Long-term debt	3.45
	Property, plant, and equipment	15,340		Deferred income tax	
	(Less accumulated depreciation)	-2,200		Other	
	Intangible assets			Total long-term liabilities	3,450
	Total fixed assets	14,348	Owner's Equity		
Other Assets				Owner's investment	7,178
	Deferred income tax			Retained earnings	4,389
	Other			Other	
	Total Other Assets			Total owner's equity	11,567
Total	**Assets**	26,222	**Total**	**Liabilities and Owner's Equity**	26,222

(continued)

This example illustrates a simplified **balance sheet** for a hypothetical company for the fiscal year ending in FY 01 (in $ thousands). The assets section is divided into current assets, non-current assets, and other assets. Current assets include short-term resources such as cash, which in this case totals $11,874. Non-current assets comprise long-term investments and fixed assets like property, plant, and equipment (PPE), adjusted for accumulated depreciation. Other assets include items such as deferred income tax. Altogether, the total assets amount to $26,222. On the other side, liabilities and owner's equity are also broken down into categories. Current liabilities, which include obligations due within one year like accounts payable and income taxes, total $11,205. Long-term liabilities, such as long-term debt and deferred taxes, come to $3,450. Owner's equity represents the owner's claim on the business and includes both investment and retained earnings, totaling $11,567. In this **balance sheet**, total liabilities and owner's equity equal $26,222, confirming that the sheet is balanced and upholding the fundamental accounting equation: assets are financed either through liabilities or the owner's equity.

Multiple-Choice Question: Company A has $15,000 in assets, $10,000 in owner's equity, and $5,000 in liabilities. The company takes out a $2,000 bank loan. How does this affect the **balance sheet**?

1. Liabilities: debit $5k, assets: debit $5k
2. Owner's Equity: credit $5k, assets: debit $10k
3. Liabilities: credit $5k, assets: debit $5k
4. Owner's Equity: debit $5k, Liabilities: credit $5k
5. None of these

Open-Ended Question: Describe the four main components of a company's financial statements. What makes the **balance sheet** unique among them?

References:

Phillips, F., Clor-Proell, S., Libby, R. & Libby, P. (2021). *Fundamentals of Financial Accounting* (7th ed.). McGraw-Hill.

Weygandt, J. J., Kimmel, P. D., & Mitchell, J. E. (2020). *Accounting Principles* (14th ed.). Wiley.

White, G. I., Sondhi, A. C., & Fried, D. (2002). *The analysis and use of financial statements* (3rd ed.). Wiley.

Back to Homepage

Notes: Exemplary wiki contributions were made available within the wiki platform for student reference.

7.3. Appendix II-3. Review Template

Guideline	Student comment
Does the wiki contribution contain between 500 and 800 words?	Yes / No, …
Does the wiki contribution start with a stand-alone, source-based one-sentence definition?	Yes / No, …
Does the wiki contribution continue with a more detailed, source-based definition?	Yes / No, …
Does the wiki contribution provide an example?	Yes / No, …
Does the wiki contribution provide at least five links to related word list terms? Do the links function correctly?	Yes / No, …
Does the wiki contribution provide one multiple-choice and one open-ended question?	Yes / No, …
Does the wiki contribution provide at least three academic sources?	Yes / No, …
Does the wiki contribution provide in-text citations correctly?	Yes / No, …
Does the wiki contribution provide a list of references according to the guidelines?	Yes / No, …
Does the wiki contribution meet the formatting guideline such as formatting the word list term bold and referring to the homepage at the end of the contribution?	Yes / No, …
What I like about the wiki contribution is that …	…
As a recommendation, I would suggest …	…

Notes: The review template was made available to students as an Excel file and was linked on the wiki platform. Students were instructed to follow a structured review process: (1) locate their assigned review terms in the word list, (2) download the student review template, (3) insert their comments directly into the downloaded Excel file, (4) transfer their completed comments from the Excel file into the comment section of their assigned review term, and (5) refer to the step-by-step instructional video for guidance on steps (1)–(4), in case of uncertainties.

7.4. Appendix II-4. Interview Guide

Central themes	Questions
Student collaboration	How did you experience the individual writing activity on the wiki?
	How did you experience the collaborative writing activities on the wiki?
	How did you experience the peer-review activities on the wiki?
Instructor scaffolding	How did you experience the wiki's guidance resources (e.g., contribution guidelines, exemplary contributions, peer-review template)?
	How did you experience the wiki instructions provided in the beginning of the semester (e.g., in class, by email, within Moodle)?
	How did you experience the wiki support provided during the semester (e.g., IT problems, content feedback, deadline reminders)?
Learning experiences	How did you experience the wiki for your learning experience?
	How did you use the wiki for learning in addition to the lecture and tutorial resources?
	How did you find the wiki learning environment in Moodle?

7.5. Appendix II-5. Variable Definitions

Variables	Definition	Source
ACADEMIC_ PERFORMANCE	Academic performance, measure of performance in the undergraduate introductory accounting course equal to the percentage of points earned out of a maximum of 120. The score is based on the final exam only, which has a theoretical maximum of 120 points. The actual maximum score is 104 points (87%). A minimum of 54 points (45%) is required to pass.	Final exam grading
CONTRIBUTIONS	Wiki contributions, indicating the number of contributions made to the wiki by a student, out of a maximum of 3 wiki contributions.	Wiki platform (i.e., Moodle)
COMPR_CONTRIBUTOR	Indicator variable equal to 1 if a student contributed more than one contribution to the wiki, and 0 otherwise.	Self-constructed based on *CONTRIBUTIONS*
AGE	Student age (years).	University student database
FEMALE	Indicator variable equal to 1 if the student is female, and 0 otherwise.	University student database
GPA	Grade point average, measure of university performance, inversely coded on a best (4.0) to worst (1.0) basis. GPA calculation excludes the course under study.	University student database
PRE_UNI_ PERFORMANCE	Pre-university performance refers to the proportion of university tuition fees covered by merit-based scholarships awarded upon entry. These scholarships are granted based on indicators such as academic achievement in secondary education and extracurricular engagement prior to university admission.	University student database
WORK_EXPERIENCE	Indicator variable equal to 1 if the student has worked in an accounting job prior to or during the semester, and 0 otherwise.	Hand-collected from students
ENGLISH_PROFICIENCY	English proficiency variable coded 0 for non-native students and A-Levels not entirely in English language, 1 for non-native students and A-Levels in English language, and 2 for students native in English language, respectively.	Hand-collected from students
STUDY_PROGRAM	Study program variable coded 1 for "Global Economics and Management," 2 for "Industrial Engineering and Management," 3 for "International Business Administration," and 4 for others, respectively. Used as fixed effect.	University student database
REGION	Region variable coded 1 for "East Asia & Pacific," 2 for "Europe & Central Asia," 3 for "Latin America & Caribbean," 4 for "Middle East & North Africa," 5 for "South Asia," and 6 for "Sub-Saharan Africa," respectively. Used as fixed effect.	University student database

Notes: This table provides definitions of the variables.

8. Statement on Funding

This work was supported by the “Stiftung Innovation in der Hochschullehre” as part of the project “Wiki4ll – Wiki für alle” (Project No. FRFMM-752/2022) within the funding program “Freiraum 2022.”

9. Statement on Declaration of Interest

No potential conflict of interest was reported by the author(s).

10. Statement on the Use of Artificial Intelligence (AI)

During the preparation of this work, the authors used *ChatGPT*, *DeepL Translate*, *DeepL Write*, and *Grammarly* to improve the language and readability, exercising caution. After using these tools, the author reviewed and edited the content as needed, taking full responsibility for the work's content.

11. Statement on Data Availability

The data presented in this study are available on request from the authors. The data are not publicly available due to privacy restrictions.

III. Essay II

Determinants of Corporate Environmental Taxes: Country-Level Empirical Evidence

Lukas Schmidt, Andreas Seebeck, Marius Weiß

Abstract

This study investigates why countries differ in the implementation of corporate environmental taxes, a key instrument for internalizing environmental externalities. Using text mining, web scraping, and multi-stage human coding, we develop a novel index that captures the presence of corporate energy, pollution, resource, and transport taxes across 149 countries. We link cross-country variation in corporate environmental tax implementation to climate risk, corporate tax competition, governance quality, and female political representation using logistic regression. The results show that climate risk does not explain the implementation of corporate environmental taxes. In contrast, greater corporate tax competition significantly decreases the likelihood of implementation, while higher governance quality and stronger female political representation increase it. Additional analyses indicate that the gender effect is likely to operate through parliamentary mechanisms and emerges only once a critical mass of women is present. These findings provide the first global, instrument-specific evidence on the determinants of corporate environmental taxes, highlighting key political and institutional drivers.

Acknowledgements

We thank Adalbert Wilhelm, Frank Hechtner, Tilo Halaszovich, Christoph Lattemann, Philipp Richter, Marvin Nipper, and Inga Schulz (Discussant) for their valuable comments. We also thank the participants of the Virtual Doctoral Tax Seminar Mannheim, the EAA Annual Congress 2024 (Bucharest), as well as 2025 (Rome), and the University of Erlangen-Nürnberg and Constructor University Bremen, for providing the opportunity to present and discuss our work. We sincerely appreciate the valuable comments and suggestions from the Editors and anonymous reviewers of "Accounting Forum," which have greatly contributed to improving this study.

1. Introduction

Corporate activities cause various negative environmental externalities that are not reflected in the market price of goods and services (Chava, 2014). These externalities include the depletion of nonrenewable resources, air and water pollution, and significant contributions to human-induced climate change through corporate greenhouse gas (GHG) emissions (Crowley, 2000; Levitus et al., 2001). Thus, corporations present an intuitive target for implementing environmental policy instruments, such as corporate environmental taxes (Intergovernmental Panel on Climate Change (IPCC), 2022b). In general, these taxes take the form of Pigouvian taxes, designed to internalize environmental and social costs into market prices. Specifically, they address negative environmental externalities by making the polluter pay the full cost of economic activities by integrating a physical unit associated with those externalities into the tax base (Jacob & Zerwer, 2024; Pigou, 2017).

Prior research provides evidence on the effectiveness of corporate environmental taxes in reducing such externalities. For example, Martinsson et al. (2024) find that energy taxes can reduce corporate emissions. Moreover, Brown et al. (2022) find that corporate pollution taxes lead to substantial increases in firms' research and development spending, and Martin et al. (2014) find an adverse effect of energy taxes on corporate energy intensity and electricity use. Consequently, numerous international unions and organizations advocate the implementation of corporate environmental taxes (European Environment Agency (EEA), 2022; International Monetary Fund (IMF), 2019; IPCC, 2022b). However, not all countries have implemented corporate environmental taxes.

Cross-country research on the determinants of environmental policy implementation has examined environmental policy instruments at an aggregate level, often without distinguishing between specific policy types or regulatory mechanisms (e.g., Fredriksson & Neumayer, 2013; Fredriksson & Svensson, 2003). To the best of our knowledge, no prior research has specifically examined the determinants of corporate environmental taxes. This study aims to fill this gap.

To measure cross-country variation in corporate environmental tax implementation, we develop a Corporate Environmental Tax Index (CET-Index) that systematically captures whether countries have implemented corporate energy, pollution, resource, and transport taxes. We build the index through a multi-stage process combining text mining and web scraping of corporate tax legislation and government materials with structured human coding. We first create a validated word list of terms related to corporate environmental taxation to extract all relevant sentences from corporate tax legislation and government materials automatically. Next, we score the presence of each tax dimension. We then employ a logistic regression framework to assess the explanatory power of various country-level determinants for cross-country variation in corporate environmental tax implementation.

Ex ante, in line with public interest theory (Breyer, 1982; Stern, 2008) as well as the theory of focusing events (Birkland, 2006), we posit in our first hypothesis that countries with greater climate risk are more likely to implement corporate environmental taxes. Surprisingly, we do not find evidence that climate risk, whether captured through forward-looking measures of structural exposure and socioeconomic vulnerability or backward-looking indicators such as disaster frequency and temperature change, explains cross-country variation in corporate environmental tax implementation. This null finding may suggest that climate risk alone is insufficient to induce governments to implement preventive fiscal instruments. Structural constraints, limited administrative capacity, and the political prioritization of short-term recovery and adaptation over long-term mitigation likely inhibit the translation of climate hazards into corporate environmental taxation. In addition, institutional inertia and path-dependent policymaking often impede the translation of environmental shocks into regulatory change (Baumgartner & Jones, 1991; IPCC, 2022a, 2022b).

In the second hypothesis, we posit that countries' involvement in corporate tax competition is negatively associated with the implementation of corporate environmental taxes. We build on the basic definition of corporate tax competition as encompassing any form of noncooperative corporate tax setting by independent governments to attract and retain corporate tax bases (Glaeser et al., 2023; Wilson & Wildasin, 2004). We find robust evidence that corporate tax competition significantly reduces the likelihood that countries implement corporate environmental taxes. This pattern is consistent across multiple measures of corporate tax competition and indicates that governments competing for mobile capital tend to refrain from implementing additional cost-increasing tax instruments on firms (Devereux & Loretz, 2013; Glaeser et al., 2023; Wilson & Wildasin, 2004).

Our third hypothesis proposes that countries' governance quality is positively associated with the presence of corporate environmental taxes. Consistent with public interest theory, which argues that policymaking serves as a socially efficient response to market failures, including negative environmental externalities, our findings provide supporting evidence for this relationship. As countries with higher governance quality tend to be more proactive and responsive to the needs and welfare of their citizens, they are more likely to implement corporate environmental taxes, which are designed to internalize environmental and social costs that are not yet reflected in the market price of goods and services (Breyer, 1982; Stern, 2008).

Finally, in line with our fourth hypothesis, we find that countries with greater female political representation are more likely to implement corporate environmental taxes.[35] This finding aligns with gender socialization theory, frequently utilized in accounting research (e.g., Haque et al., 2026), which posits that women's socialized roles and values, emphasizing care and community responsibility, make them more inclined to prioritize social and environmental issues (Eagly, 1987; Eagly & Wood, 1991). Moreover, this finding is consistent with that of previous research on gender differences in legislative behavior, which revealed that women are significantly more likely than men to support environmental legislation and climate mitigation policies (e.g., Dabla-Norris et al., 2023; Ramstetter & Habersack, 2020). We also provide evidence that the observed effect is likely to stem from mechanisms unique to parliamentary representation, such as voting behavior, rather than general societal progressiveness. Furthermore, we find that a critical mass of at least 20–40% female political representation is required before women begin to impact a country's corporate environmental tax framework. Our findings align with critical mass theory, frequently utilized in accounting research (e.g., Chijoke-Mgbame et al., 2020; Haque et al., 2026; Oradi & E-Vahdati, 2021), which posits that a threshold or critical mass of underrepresented group members is required for them to meaningfully influence group dynamics and outcomes (Kanter, 1977a, 1977b).

We acknowledge that cross-country analyses are inherently susceptible to endogeneity concerns due to omitted-variable bias. Therefore, we extend our baseline specification with additional political and economic controls and examine the sensitivity of our core results to unobserved heterogeneity using the Oster (2019) coefficient-stability test. Across all analyses,

[35] We recognize that gender is a social construct that encompasses a spectrum of identities beyond binary categories. We acknowledge that there are broader and multifaceted dynamics related to gender that are beyond the scope of this study.

our results remain stable, providing confidence that endogeneity is unlikely to drive the main findings.

This paper makes several important contributions. First, it adds to the rapidly growing sustainability-related tax literature by providing evidence that corporate tax competition, governance quality, and female political representation are determining factors of countries' corporate environmental taxes. Most sustainability-related tax studies address the determinants and consequences of firms' tax avoidance and corporate social responsibility behavior (e.g., Adrian et al., 2023; Bird & Davis-Nozemack, 2018; Hardeck et al., 2024; Park et al., 2024; Xu & Ren, 2024). While these studies contribute valuable insights into the intersection of taxation and corporate social responsibility, they primarily examine the reactive aspects of firm behavior to sustainability developments and tax policies. However, research on the factors influencing the implementation of corporate environmental taxes at the *country-level* remains limited.

Second, we add to the literature on the determinants of environmental policy instruments (e.g., Bättig & Bernauer, 2009; Gazmararian & Milner, 2024). Most studies in this field focus on the effectiveness of various types of environmental policies. We provide evidence on the determinants of instrument-specific policies, i.e., corporate environmental taxes. Because they affect a wide range of corporations financially, corporate environmental taxes are likely to be more difficult to implement and to attract more public attention than other environmental policy instruments, such as subsidies for green technologies or voluntary emission-reduction programs, making them of particular interest to researchers seeking to understand how political and institutional factors explain the cross-country variation in the implementation of specific environmental policy instruments.

Third, our CET-Index facilitates comparative analyses across countries and regions, enabling the identification of best practices and implementation gaps and distinguishing leaders from laggards. We provide the CET-Index, its four dimensions, and the list of data sources as an exhibit to support cross-country learning and policy transfer. In addition, the data used open new avenues for research on the impact of tax system characteristics, and the CET-Index can be used in future firm-level studies to control for the presence of corporate environmental taxes at the national level.

Finally, our results have direct policy implications for ongoing international efforts to meet climate change and sustainable development goals. They emphasize the importance of addressing international corporate tax competition and increasing gender diversity in legislation to promote corporate environmental taxes as a viable tool for environmental action. Moreover, our results emphasize the importance of holistically considering the trade-offs and co-benefits between multiple sustainable development goals (SDGs), including gender equality (SDG5) and climate action (SDG13) (IPCC, 2022a, 2022b; United Nations Entity for Gender Equality and the Empowerment of Women (UN WOMEN), 2022; UN WOMEN & United Nations Department of Economic and Social Affairs (UN DESA), 2024).

2. Theoretical Framework and Hypotheses Development

2.1. Climate Risk and Corporate Environmental Taxes

Governments are likely to be motivated to implement environmental policy instruments either by anticipating future climate-related risks that threaten long-term economic and social welfare or by reacting to past climate-related shocks that heighten public attention and political pressure for climate action. According to public interest theory, policymaking is a socially efficient response to market failures, such as negative environmental externalities, with policymakers acting in the public interest rather than pursuing personal agendas (Breyer, 1982; Stern, 2008). A forward-looking climate risk perspective reinforces this logic by emphasizing that national governments are likely to respond to projected future hazards. Countries differ in their structural likelihood of experiencing climate-related events, which shapes anticipatory expectations and can motivate preventive measures, including the implementation of corporate environmental taxes. Forward-looking climate risk also encompasses societal vulnerability, e.g., economic and institutional conditions that determine how severely a country would be affected if hazards occur. Together, both exposure and vulnerability provide a lens through which policymakers may view corporate environmental taxes as a proactive instrument to mitigate future climate risks. In this context, recent cross-country research shows that anticipated future climate damages increasingly shape the implementation of mitigation policies (e.g., Gazmararian & Milner, 2024).

The theory of focusing events offers a complementary, backward-looking climate risk perspective by arguing that rare, disruptive, and highly visible disasters can trigger policy change by drawing public attention to underlying vulnerabilities and creating legitimacy for governmental intervention (Birkland, 2006; Bonetti et al., 2024). Such events, including floods, storms, or other extreme hazards, have been shown to reshape policy agendas by raising public environmental concerns and increasing support for mitigation measures (Birkland, 2006; Gazmararian & Milner, 2024; Hoffmann et al., 2022). Cross-country literature shows that individuals exposed to temperature anomalies, heat or drought episodes, and other climate-related hazards tend to exhibit higher environmental concern, greater support for climate policies, and stronger voting preferences for Green parties (Cologna et al., 2025; Dablander, 2025; Hoffmann et al., 2022), patterns that plausibly shape the broader political environment in which corporate environmental taxes are considered.

However, several factors may weaken the link between climate risk and corporate environmental tax implementation. First, environmental taxes are preventive instruments that impose short-term economic costs in exchange for long-term societal benefits, which makes them politically difficult to justify in the immediate aftermath of climate shocks that typically shift attention toward recovery and adaptation rather than mitigation (IPCC, 2022b). Second, countries that are most exposed or vulnerable to climate-related hazards often have limited fiscal and administrative capacity to introduce new and potentially contentious tax measures (IPCC, 2022a). Third, climate exposure is a global, cross-border phenomenon. Highly exposed and vulnerable countries often bear the consequences of emissions produced elsewhere, which can dampen national mitigation ambition, especially when domestic contributions to global climate change have been minimal (IPCC, 2022b). Fourth, policymaking processes may be shaped by institutional inertia and path dependency, which often constrain governments' ability to translate external environmental shocks into regulatory change (Baumgartner & Jones, 1991; IPCC, 2022a, 2022b).

Overall, despite these countervailing considerations, we assume that, in line with public interest theory and the theory of focusing events, greater climate risk heightens the perceived urgency of addressing environmental externalities and strengthens incentives to implement preventive mitigation policies. We therefore expect corporate environmental taxes to be more prevalent in countries with higher climate risk. Accordingly, we hypothesize:

H1: There is a significant positive relationship between a country's climate risk and the presence of corporate environmental taxes.

2.2. Corporate Tax Competition and Corporate Environmental Taxes

When firms and capital are internationally mobile, governments' environmental tax choices are shaped not only by environmental objectives but also by competitive pressures to maintain fiscal attractiveness and avoid deterring investment. We build on the basic definition of corporate tax competition as encompassing any form of noncooperative corporate tax setting by independent governments to attract and retain corporate tax bases (Glaeser et al., 2023; Wilson & Wildasin, 2004). Corporate tax competition occurs when governments reduce the financial and administrative burdens associated with corporate taxation relative to those in other countries to attract corporate capital and income (Devereux et al., 2008). This phenomenon is commonly referred to as the *race to the bottom* and often occurs at the expense of generating sufficient public revenue (Devereux & Loretz, 2013).

Ex ante, we assume that the more a country is involved in corporate tax competition, the more likely it is to prioritize attracting and retaining corporations over implementing corporate environmental taxes. Imposing corporate environmental taxes can increase costs through added tax burdens and increased administrative expenses, potentially prompting firms to relocate to more tax-friendly jurisdictions. In addition, early theoretical work on corporate tax competition has shown that tax competition leads to the underprovision of public goods (Bradford & Oates, 1971; Oates, 1972), which include environmental quality (Rauscher, 1995). In this context, corporate tax competition weakens environmental action, as governments must balance environmental action with economic competitiveness. Moreover, recent theoretical models provide supporting evidence that corporate tax competition increases global carbon emissions (Duan et al., 2024). Finally, corporate tax competition reduces government fiscal capacity, as the *race to the bottom* erodes corporate tax bases, limiting the resources available to introduce and enforce corporate environmental taxes.

However, multiple factors can weaken the connection between international tax competition and the implementation of corporate environmental taxes. First, imposing corporate environmental taxes does not necessarily increase financial or administrative burdens, as it can yield a *double dividend*.[36] Second, international coordination efforts may further reduce the competitive pressures associated with implementing additional tax

[36] The *double dividend* is the idea that environmental taxes simultaneously deliver environmental benefits by reducing pollution and economic benefits by allowing revenues to replace more distortionary taxes that hinder economic growth (Goulder, 1995; Yamazaki, 2022).

instruments. For instance, international coordination may establish a coordinated floor under effective corporate taxation, reducing countries' ability to attract profits through very low tax rates. As Devereux (2023) argues, such coordination weakens incentives for harmful tax undercutting, thereby expanding governments' policy space to implement corporate environmental taxes without jeopardizing international competitiveness. Third, increasing stakeholder emphasis on sustainability places additional pressure on firms and governments to implement environmental policies, even in competitive tax environments. As corporate social responsibility becomes more salient to investors and consumers, firms may prefer jurisdictions with credible environmental standards, thereby indirectly reinforcing the implementation of such taxes. Governments may therefore view corporate environmental taxes as a means to align with stakeholder expectations and preserve investment attractiveness. Moreover, jurisdictions that implement these measures may strengthen their international reputation, whereas those that refrain risk being perceived as tax havens rather than progressive, environmentally responsible actors.

Considering both arguments for and against the relationship between tax competition and corporate environmental tax implementation, we assume that cross-country corporate tax competition has an adverse effect on the decision to implement corporate environmental taxes. Thus, our second hypothesis is as follows:

H2: There is a significant negative relationship between a country's involvement in corporate tax competition and the presence of corporate environmental taxes.

2.3. Governance Quality and Corporate Environmental Taxes

Building on public interest theory, it seems plausible that governments with higher governance quality are more responsive to long-term public interests and thus are more inclined to implement corporate environmental taxes. In line with this notion, previous research indicates that the quality of governance can be a significant driver of a country's environmental policy instruments (e.g., Fredriksson & Neumayer, 2013). High governance quality is also associated with lower levels of corruption, which helps prevent socially suboptimal policy formation often induced by industry lobbying (Damania et al., 2003; Fredriksson & Svensson, 2003). Empirical evidence at the national level further indicates a negative correlation between the level of corruption in a country and its environmental performance (Damania et al., 2003). Conversely, higher citizen participation, which correlates positively with governance quality, is associated with stronger environmental commitment (Fredriksson & Neumayer, 2013). Moreover, democracies are more likely to implement environmental policies than nondemocracies are (Bättig & Bernauer, 2009).

However, high governance quality does not inherently imply the presence of corporate environmental taxes, as governments often weigh short-term economic growth against long-term environmental goals. Corporate environmental taxes may be seen as detrimental to economic competitiveness, leading some governments to forgo their implementation. In addition, countries may opt for alternative environmental policy tools, such as market-based instruments (e.g., environmental tax incentives) or command-and-control measures (e.g., emission-reduction targets). Policymakers must also balance mitigation and adaptation strategies when addressing climate change. While corporate environmental taxes act as mitigation instruments by reducing the impact of future environmental degradation and climate change, adaptation instruments, such as climate-resilient infrastructure investments, focus on enhancing the ability to cope with the unavoidable impacts of climate change. These complexities may make it less straightforward to assume a positive association between governance quality and the presence of corporate environmental taxes. Nevertheless, the literature generally suggests that higher governance quality enhances environmental governance because policymakers are more benevolent and public-oriented. Thus, we propose our third hypothesis as follows:

H3: There is a significant positive relationship between the governance quality in a country and the presence of corporate environmental taxes.

2.4. Female Political Representation and Corporate Environmental Taxes

Differences in gender-specific policy priorities may influence a country's environmental tax implementation by shaping legislators' voting behavior on fiscal environmental policies. Gender socialization theory assumes that societal expectations shape human behavior differently on the basis of gender. Women are often encouraged to develop community-supporting traits, such as compassion, cooperation, and nurturance. In contrast, men are socialized to embrace agentic values and are often taught that conforming to masculine norms entails being ambitious and acting in a self-directed way (Eagly, 1987). Accordingly, it is assumed that women tend to exhibit greater tendencies toward social facilitation and concern for the welfare of others than men do (Eagly & Karau, 1991). Assuming politicians implement policies aligned with their preferences, research finds that women's political representation can influence political decision-making.

Prior research documents systematic gender differences in legislative voting on environmental issues. Mavisakalyan and Tarverdi (2019) show that higher female political representation is associated with more stringent climate policies, consistent with women's stronger pro-environmental preferences. Ramstetter and Habersack (2020) provide evidence that female parliamentarians vote more pro-environmentally than their male counterparts. Similarly, May et al. (2021) find that women exhibit greater support for market-based environmental instruments, while Dabla-Norris et al. (2023) and Fredriksson and Wang (2011) show that female legislators are more supportive of climate mitigation and stricter environmental policies. Together, these findings suggest that gender-specific policy priorities translate into systematic differences in legislative voting behavior, increasing support for environmental policy measures. Finally, Al Mamun et al. (2024) show that female political empowerment promotes green finance through higher environmental tax burdens, likely driven by more pro-environmental voting by female parliamentarians.

However, opposing views suggest several reasons why female political representation may not translate into systematic differences in corporate environmental tax implementation. First, prior research points to minimal gender differences in leadership traits (Adams & Funk, 2012; García-Lara et al., 2017). Women aspiring to leadership may reject traditional stereotypes and exhibit values and styles similar to those of men (Powell, 1990), and Eagly et al. (1995) find no overall gender difference in leadership effectiveness. Second, even when female legislators hold stronger pro-environmental preferences, these preferences may be expressed through support for alternative policy instruments rather than corporate environmental taxes, such as regulatory standards, subsidies, or public investment. Third, observed associations between female political representation and environmental taxation may reflect broader institutional or societal characteristics. Countries with higher female representation often exhibit stronger governance quality and more progressive political environments, which may independently facilitate the implementation of corporate environmental taxes.

Most studies suggest notable gender differences in values, perceptions, and beliefs (e.g., Dabla-Norris et al., 2023; May et al., 2021; Ramstetter & Habersack, 2020). We assume that female political representation explains cross-country variation in corporate environmental tax implementation because female legislators shape fiscal environmental policymaking through voting behavior. Our fourth hypothesis is as follows:

H4: There is a significant positive relationship between female political representation in a country and the presence of corporate environmental taxes.

3. Development of the Corporate Environmental Tax Index (CET-Index)

3.1. General Approach

The Corporate Environmental Tax Index (CET-Index) focuses on the national presence of Pigouvian environmental taxes that are imposed on corporations as of December 31, 2022.[37] Environmental taxes are taxes whose tax base is a physical unit (or a proxy of it) of something that has a proven, specific negative impact on the environment (IMF, 2024). To capture the different facets of corporate environmental taxes, we consider the four commonly distinguished environmental tax dimensions: energy taxes, pollution taxes, resource taxes, and transport taxes (IMF, 2024; Organization for Economic Co-operation and Development (OECD), 2024). We construct the CET-Index using a four-step process. First, we compile a word list containing corporate environmental tax terminology. Second, we extract and process textual data from corporate tax legislation and government materials using text mining and web scraping. Third, we independently score the presence of corporate environmental tax dimensions through multiple coding rounds to ensure reliability and accuracy. Fourth, we validate the index through external comparisons with third-party tax databases and internal assessments of the coherence and structure of the four dimensions.

3.2. Word List Development and Validation

Two researchers independently constructed a comprehensive word list capturing terminology relevant to the four corporate environmental tax dimensions. The compilation was informed by recent peer-reviewed environmental tax research and documentation of datasets. The independently developed word lists were subsequently compared, and inconsistencies

[37] The CET-Index excludes subnational policies, green tax incentives, and non-tax instruments. Subnational policies are omitted because consistent and accessible documentation is unavailable across regional jurisdictions for a global sample. Green tax incentives are excluded as they differ conceptually from Pigouvian taxes, being income- or expenditure-based rather than linked to environmental externalities. Non-tax instruments, such as subsidies or emission trading systems, differ in their regulatory and economic mechanisms and are beyond our focus on fiscal measures. See Appendix III-1 for an introductory overview of different forms of environmental policy instruments.

were reconciled through discussions between the two researchers. Any remaining divergences were adjudicated in consultation with a third researcher to ensure inter-coder reliability and methodological rigor. The finalized word list comprises 177 distinct terms associated with corporate environmental taxation and its four dimensions.[38]

3.3. Collecting Data on Corporate Environmental Taxes

To construct the CET-Index, we collected information from corporate tax legislation in force at the time of data collection and related materials published online by national governments. For each country, the relevant documents were obtained either as downloadable PDFs or through official government websites. Using text mining and web scraping, we extracted textual data from these sources, translating non-English materials as necessary. The extracted texts were then pre-processed to remove formatting artifacts and converted to lowercase. Subsequently, regular expression matching was performed using our corporate environmental tax word list. For each country and source, this procedure produced a comprehensive set of sentences containing at least one of the predefined environmental tax terms.[39]

3.4. Index Construction

Two researchers independently screened the sentences related to corporate environmental taxes for each country and source, indicating whether each of the four tax dimensions was present. Each researcher conducted two independent scoring rounds to minimize potential bias from single scoring attempts and enhance reliability. For each country and source, this process produced two independent assessments of the presence or absence of corporate environmental tax dimensions. The results were compared, and any discrepancies were resolved through discussion between the two researchers. Remaining inconsistencies were addressed through additional independent scoring by a third researcher, followed by joint discussions among all authors to ensure accuracy and methodological rigor. The resulting scores for the four tax dimensions constitute the CET-Index, which comprises six measures: the four individual dimensions, as well as two aggregate indicators that capture whether at least one corporate environmental tax is present and how many of the four dimensions are implemented.[40]

[38] See Appendix III-2 for the corporate environmental tax terminology word list.
[39] See Appendix III-3 for the sources used for data collection.
[40] See Appendix III-4 for the CET-Index scores.

3.5. Index Validation

To assess the external validity of the CET-Index, we compare our country-level classifications with three authoritative third-party sources commonly used in tax research and index construction (e.g., Schanz et al., 2017): The International Bureau of Fiscal Documentation (IBFD) Country Tax Guides, the Worldwide Tax Summaries Online, hosted by PricewaterhouseCoopers (PwC), and the Worldwide Corporate Tax Guide, published by Ernst & Young (EY). As no global database systematically documents the presence of corporate environmental taxes under the specific definition employed in this study, these sources represent the closest available benchmarks for validation. For each of the four tax dimensions, we verify that countries coded as "1" in our index (indicating the presence of a corporate environmental tax) are also reported as having the corresponding tax instrument in at least one of the external sources. Conversely, we confirm that observations coded as "0" are not contradicted by these sources. Where discrepancies arise, we apply a structured resolution procedure. In cases where the CET-Index indicates a tax dimension's presence, but external sources do not, we manually re-examine the underlying corporate tax legislation to exclude potential false positives from the text-mining and web-scraping stages. When external sources report an environmental tax dimension that is absent in our coding, we engage in professional consultation with government officials and country-specific tax specialists to clarify whether the tax instrument applies to corporations under the definitions used in this paper. These consultations ensure that the classification aligns with both legal practice and conceptual scope. Despite these extensive validation efforts, we acknowledge that a residual risk of misclassification cannot be entirely eliminated due to heterogeneous reporting standards across countries and the absence of a dedicated global dataset on corporate environmental taxes.

Although the CET-Index is conceptualized as a formative composite intended to capture four distinct dimensions of corporate environmental taxation, it is nonetheless informative to examine whether the indicators exhibit coherent empirical behavior. To this end, we conduct a set of internal diagnostics grounded in Classical Test Theory. We first compute Cronbach's alpha (Cronbach, 1951) using the four binary tax indicators for energy, pollution, transport, and resource taxes. The resulting reliability coefficient of 0.65 (untabulated) indicates moderate internal consistency, which is notable given the formative nature of the index and dichotomous scaling of the components. This diagnostic suggests that the four components exhibit sufficiently coherent empirical behavior while still capturing distinct aspects of corporate environmental taxation.

Second, we examine whether the four binary tax indicators load onto a common latent structure by conducting an Exploratory Factor Analysis based on their tetrachoric correlation matrix. The tetrachoric correlations, which range from 0.19 to 0.77, indicate moderate associations among most item pairs and justify a factor-analytic approach for dichotomous variables. Extracting factors via principal-component factoring, the first factor yields an eigenvalue of 2.52 and explains approximately 63% of the total variance, while all subsequent eigenvalues fall well below 1.0. Consistent with a unidimensional structure, all items exhibit positive loadings on the first factor, ranging from 0.56 for resource taxes to 0.89 for transport taxes. This pattern suggests that the four corporate environmental tax dimensions share a common underlying propensity. Countries that implement one type of corporate environmental tax are, on average, more likely to implement others. The variation in loadings confirms that the indicators capture distinct facets of the broader construct rather than reflecting a redundant measurement structure. Taken together, these diagnostics support treating the CET-Index as a coherent set of related but non-interchangeable indicators and justify our empirical strategy of using both the aggregated index and its underlying dimensions in subsequent analyses.

4. Data and Research Design

4.1. Sample Selection

Our initial sample comprises the 193 UN member states as of December 31, 2022. Countries with insufficient sources for the construction of the CET-Index were excluded. Due to limited data availability, the initial sample size was reduced to 149 countries. This final sample still accounts for approximately 93% of the total gross domestic product (GDP) and GHG emissions of UN member states between 2017 and 2021. However, certain countries remain excluded from our sample, presenting an inherent limitation. Table III-1 summarizes our sample selection.

Table III-1. Sample Selection

	Observations
Member states of the UN as of December 31, 2022	193
Less: Countries with insufficient sources for CET-Index construction	3
Less: Missing complete data on independent and control variables	41
Final sample	149

Notes: This table summarizes the sample selection procedures.

4.2. Empirical Model

To test our research hypotheses, we estimate the following logistic regression model:

$$
\begin{aligned}
&Prob(ENV\_TAX = 1)_{i,t} \\
&\qquad = \beta_0 + \beta_1 CLI\_EXP_{i,p} + \beta_2 TAX\_COM_{i,p} + \beta_3 GOV\_QUA_{i,p} + \beta_4 FEM\_PAR_{i,p} \\
&\qquad + \beta_j CONTROLS_{i,p} + \varepsilon_{i,t}
\end{aligned}
\tag{III-1}
$$

where i is the country, t is the time of assessment of the dependent variables, p is the period-average of the independent variables and control variables j, and $\varepsilon_{i,t}$ is a logistically distributed error term.

4.3. Dependent Variable

We use *ENV_TAX*, defined as the presence of corporate environmental taxes, as the dependent variable in Equation (III-1). It is equal to "1" if the country has implemented at least one of the four corporate environmental tax dimensions and "0" otherwise.

4.4. Independent Variables

To test Hypothesis 1, following prior country-level studies in the accounting literature (e.g., Mouti et al., 2026), we measure a country's climate risk as the probability of physical exposure to climate-related hazards (*CLI_EXP*). For Hypothesis 2, we capture corporate tax competition through the inverse transformation of the taxes and mandatory contributions borne by the business as a percentage of commercial profit (*TAX_COM*). In line with prior country-level studies in the accounting literature (e.g., Kaya & Seebeck, 2019; Ramanna & Sletten, 2014), Hypothesis 3 is tested using aggregated country-level governance indicators to proxy national governance quality (*GOV_QUA*). Finally, in line with previous cross-country studies on the determinants of climate change policies (e.g., Mavisakalyan & Tarverdi, 2019), we test Hypothesis 4 using the proportion of parliamentary seats held by women (*FEM_PAR*) as a proxy for a country's female political representation.

4.5. Control Variables

We control for a wide range of factors, including a country's legal origin, economic development, emission levels, industry lobbying, population pressures, and general fiscal orientation, all of which are likely to shape corporate environmental tax legislation. First, to capture differences in institutional and regulatory traditions, we include *COM_LAW*, which equals "1" if a country follows a common law system and "0" if it follows a civil law system. Common law systems, which emphasize dispute resolution and market outcomes, tend to be less interventionist than civil law systems that are more policy-implementing and stakeholder-oriented (Kock & Min, 2016; La Porta et al., 2008; Liang & Renneboog, 2017). Second, we include per capita values for GDP (*GDP_PCA*) to capture a country's economic development. Third, we control for a country's negative contribution to environmental pollution by including the country's per capita GHG emissions (*GHG_PCA*). Fourth, we consider manufacturing value added as a percentage of GDP (*MAN_VAD*) to account for the relative lobbying strength of sectors typically engaged in environmentally harmful activities, which are motivated to avoid additional costs from environmental taxes (e.g., Von Stein, 2022). Fifth, we follow prior research and account for population pressures (e.g., Shi, 2003), such as resource scarcity, by measuring population density as the midyear population divided by the land area in square kilometers (*POP_DEN*). Sixth, we include *TAX_DTY*, which is equal to "1" if the country's constitution refers to the duty to pay taxes, and "0" if it does not. This accounts for a country's general tendency to use taxes as a policy instrument. To address reverse causality issues, we use lagged averages over 2017–2021 for the independent and control variables.[41]

5. Results

5.1. Descriptive Statistics and Pairwise Correlations

Table III-2 presents the descriptive statistics for the dependent, independent, and control variables included in Equation (III-1). Among the 149 countries included in our final sample, 117 (78.5%) countries have implemented at least one of the four dimensions of corporate environmental taxes, with 90 (60.4%) using energy taxes, 76 (51.0%) using transport taxes, 56 (37.6%) using resource taxes, and 49 (32.9%) using pollution taxes. On average, countries have implemented 1.8 corporate environmental tax dimensions.[42]

[41] Please note that the *TAX_COM* data is the average for the period 2015–2019. The *TAX_INCOMPLEX* data is the average for the years 2016, 2018 and 2020. The *GOV_CBL* data is the average for the period 2017–2020. The *FEM_MIN* data is the average for the period 2018–2020.

[42] See Appendix III-5 for descriptive statistics on corporate environmental taxes by geographic region; See Appendix III-6 for sub-sample analysis for non-EU and non-OECD member states. Our main results are robust to excluding EU and OECD member states, alleviating concerns that harmonized standards or policy convergence drive the findings.

Table III-2. Descriptive Statistics

Variables	Observations	Mean	Std. Dev	Min	Max
ENV_TAX	149	0.785	0.412	0	1
ENV_TAX_4	149	1.819	1.366	0	4
ENV_TAX_INC	149	0.852	0.356	0	1
ENE_TAX	149	0.604	0.491	0	1
POL_TAX	149	0.329	0.471	0	1
RES_TAX	149	0.376	0.486	0	1
TRA_TAX	149	0.510	0.502	0	1
CLI_EXP	149	3.577	1.791	0.100	8.200
TAX_COM	149	61.058	14.852	0.000	91.500
GOV_QUA	149	0.025	0.836	-1.630	1.755
FEM_PAR	149	23.461	11.946	0.000	61.250
COM_LAW	149	0.362	0.482	0	1
GDP_PCA	149	9.556	1.079	6.902	11.715
GHG_PCA	149	1.706	0.853	-0.267	4.268
MAN_VAD	149	11.830	6.174	0.502	33.056
POP_DEN	149	4.342	1.401	0.741	8.969
TAX_DTY	149	0.403	0.492	0	1

Notes: This table presents the descriptive statistics for all dependent, independent, and control variables used in the main logistic regression model. See Appendix III-1 for variable definitions.

We report the Pearson correlation coefficients in Table III-3. Our primary dependent variable, *ENV_TAX*, is statistically and significantly related to all variables of interest except for *CLI_EXP*. All bivariate correlations between the control variables are below 0.6, except for the correlations between *GDP_PCA* and *GOV_QUA* and *GHG_PCA*, respectively. In line with prior cross-country studies, we address potential multicollinearity issues by calculating variance inflation factors (VIFs) (e.g., De Groote et al., 2023; Di Pietro & Masciarelli, 2022). In our main model, the VIFs for *GOV_QUA*, *GDP_PCA*, and *GHG_PCA* are 3.52, 5.79, and 3.32, respectively. Additionally, the mean VIFs across all regression models range from 1.68 to 2.82. Thus, these values are below the typical cutoff of 10 (Ryan, 2008), suggesting that multicollinearity is unlikely to compromise the robustness of our findings.

Table III-3. Pearson Correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) *ENV_TAX*	1.000										
(2) *CLI_EXP*	0.053	1.000									
(3) *TAX_COM*	-0.248 ***	-0.133	1.000								
(4) *GOV_QUA*	0.170 **	-0.440 ***	0.229 ***	1.000							

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(5) *FEM_PAR*	0.367 ***	0.004	-0.149 *	0.254 ***	1.000						
(6) COM_LAW	-0.252 ***	0.012	0.307***	0.112	-0.236 ***	1.000					
(7) *GDP_PCA*	0.038	-0.334 ***	0.170 **	0.742 ***	0.205 **	-0.112	1.000				
(8) *GHG_PCA*	-0.052	-0.131	0.153 *	0.363 ***	0.080	-0.114	0.704 ***	1.000			
(9) *MAN_VAD*	0.154 *	0.246 ***	-0.102	-0.020	0.105	-0.222 ***	0.184 **	0.198 **	1.000		
(10) *POP_DEN*	-0.078	-0.160 *	0.106	0.105	-0.103	0.109	0.094	-0.296 ***	0.110	1.000	
(11) *TAX_DTY*	0.129	0.150*	-0.115	-0.355 ***	-0.103	-0.277 ***	-0.180 **	-0.050	0.103	-0.106	1.000

Notes: This table presents the Pearson correlations and their two-tailed p-values for the key variables used in the main logistic regression model. See Appendix III-1 for variable definitions. ***p<0.01, **p<0.05, *p<0.1.

5.2. Regression Results

5.2.1. H1: Climate Risk and Corporate Environmental Taxes

Model (1) in Table III-4 reports the logistic regression results for Equation (III-1). Contrary to our expectation, the forward-looking climate risk proxy measuring structural exposure to climate-related hazards shows no statistically significant association with corporate environmental tax implementation (p>0.1). To assess robustness, we estimate alternative specifications using different climate risk proxies. Model (2) replaces hazard exposure with socioeconomic vulnerability (forward-looking). Model (3) incorporates recent disaster frequency (backward-looking), consistent with the focusing-events mechanism. Model (4) uses mean temperature change, capturing experiential climate conditions (backward-looking). Across all specifications, the coefficients remain statistically insignificant (p>0.1). These results indicate that neither proactive (forward-looking) nor reactive (backward-looking) climate risk proxies explain cross-country variation in corporate environmental tax implementation.

The surprising absence of significant effects across all climate risk proxies suggests that structural and political constraints may prevent climate risk from translating into the implementation of corporate environmental taxes. For example, it may be difficult to justify implementing environmental taxes immediately after climate shocks. Pigouvian taxes are preventive instruments that impose short-term economic costs in exchange for long-term societal benefits (Jacob & Zerwer, 2024; Pigou, 2017). After climate shocks, governments may allocate financial resources toward recovery and adaptation rather than mitigation (IPCC, 2022b). Implementing environmental taxes requires fiscal and administrative capacity, which highly exposed and vulnerable countries often lack (IPCC, 2022a). Moreover, highly exposed and vulnerable countries usually bear the consequences of emissions produced elsewhere. Since climate exposure is a global, cross-border phenomenon, it can reduce national ambition to mitigate climate change, especially when a country's contributions to global climate change have been minimal (IPCC, 2022b). Finally, governments often lack the ability to translate external environmental shocks into regulatory change due to policymaking processes shaped by institutional inertia and path dependency (Baumgartner & Jones, 1991; IPCC, 2022a, 2022b).

Table III-4. Climate Risk and Corporate Environmental Taxes

Variables	*ENV_TAX*			
	Model (1)	Model (2)	Model (3)	Model (4)
CLI_EXP	**0.194**			
	(0.906)			
CLI_VUL		**0.373**		
		(1.283)		
CLI_DIS			**0.014**	
			(0.122)	
CLI_TMP				**0.087**
				(0.117)
TAX_COM	-0.060**	-0.059**	-0.065***	-0.058**
	(-2.488)	(-2.387)	(-2.716)	(-2.512)
GOV_QUA	2.582***	2.697***	2.324***	2.253***
	(4.183)	(4.112)	(4.290)	(4.213)
FEM_PAR	0.092**	0.096***	0.123***	0.096***
	(2.493)	(2.656)	(3.357)	(2.687)
COM_LAW	-1.268**	-1.247*	-0.754	-1.063
	(-2.029)	(-1.891)	(-1.129)	(-1.639)
GDP_PCA	-1.262***	-0.953*	-1.287***	-1.162**
	(-2.625)	(-1.773)	(-2.623)	(-2.378)
GHG_PCA	0.207	0.256	0.222	0.100
	(0.391)	(0.471)	(0.336)	(0.190)
MAN_VAD	0.060	0.073*	0.063	0.071*
	(1.610)	(1.917)	(1.599)	(1.836)
POP_DEN	0.045	0.017	0.021	-0.042
	(0.185)	(0.078)	(0.073)	(-0.183)
TAX_DTY	1.491**	1.526**	1.556**	1.369**
	(2.328)	(2.299)	(2.177)	(2.154)
Constant	13.628***	9.880*	14.078***	13.402***
	(3.203)	(1.757)	(3.185)	(3.108)

Variables	ENV_TAX			
	Model (1)	Model (2)	Model (3)	Model (4)
Observations	149	149	138	144
Pseudo R^2	0.361	0.365	0.381	0.350
Wald χ^2	43.05***	39.91***	39.06***	37.64***
Hosmer-Lemeshow χ^2	8.661	9.433	6.646	6.123
Prob>Hosmer-Lemeshow χ^2	0.372	0.307	0.575	0.633
VIF mean (max)	2.21 (5.79)	2.44 (6.66)	2.11 (5.52)	2.27 (6.05)

Notes: This table presents the logistic regression results for *ENV_TAX* as the dependent variable with available independent and control variables. Model (1) provides the results for the main model. Models (2)-(4) present results for alternative model specifications using alternative independent variables for *CLI_EXP*. Z values are reported in parentheses, based on robust standard errors. In all our models, the joint significance of the predictors in explaining the outcome variable is indicated by the Wald $\chi2$ test statistic. In all our models, the Hosmer-Lemeshow $\chi2$ test statistic provides no strong evidence against an acceptable fit to the data. See Appendix III-1 for variable definitions. ***p<0.01, **p<0.05, *p<0.1.

5.2.2. H2: Corporate Tax Competition and Corporate Environmental Taxes

Next, consistent with Hypothesis 2, the coefficient estimate of *TAX_COM* in our main model, presented in Table III-4, is negative and statistically significant (p<0.05). The average marginal effect of *TAX_COM* is -0.0063 (untabulated). Thus, on average, a one-unit increase in *TAX_COM*, which is the inverse transformation of the taxes and mandatory contributions borne by the business as a percentage of commercial profit, decreases the probability that the country imposes a corporate environmental tax by approximately 0.63 percentage points. This suggests that a country's involvement in corporate tax competition negatively affects a national government's decision to implement corporate environmental taxes. Such behavior aligns with strategies aimed at attracting and retaining corporations by minimizing both tax burdens and administrative costs.

In Table III-5, we replace *TAX_COM* with alternative proxies. Model (1) uses *TAX_COM_GDP* (p<0.05), capturing deviations from countries with similar GDP levels, and Model (2) uses *TAX_COM_REG* (p<0.01), capturing deviations from regional averages. Model (3) uses *TAX_INCOMPLEX* (p<0.01), which is the inverse transformation of the complexity of a country's corporate income tax system as faced by multinational corporations (Hoppe et al., 2023). Across all specifications, the coefficient estimates remain negative and statistically significant, supporting Hypothesis 2 that corporate tax competition negatively affects the implementation of corporate environmental taxes.

Table III-5. Corporate Tax Competition and Corporate Environmental Taxes

Variables	*ENV_TAX*		
	Model (1)	Model (2)	Model (3)
CLI_EXP	0.200	0.162	-0.111
	(1.021)	(0.731)	(-0.263)
TAX_COM_GDP	**-0.056****		
	(-2.565)		
TAX_COM_REG		**-0.066*****	
		(-2.602)	
TAX_INCOMPLEX			**-0.077*****
			(-3.145)
GOV_QUA	2.650***	2.558***	3.671***
	(4.516)	(4.098)	(3.063)
FEM_PAR	0.087**	0.101**	0.195***
	(2.370)	(2.527)	(2.950)
Constant	10.539***	10.738***	6.538
	(2.817)	(2.718)	(0.712)
Observations	149	149	84
Controls	YES	YES	YES
Pseudo R^2	0.357	0.367	0.607
Wald χ^2	50.28***	43.41***	19.72**
Hosmer-Lemeshow χ^2	8.469	9.258	0.468
Prob>Hosmer-Lemeshow χ^2	0.389	0.321	1.000
VIF mean (max)	2.24 (5.84)	2.21 (5.83)	2.82 (8.05)

Notes: This table presents the logistic regression results for *ENV_TAX* as the dependent variable with available independent and control variables. Compared to our main model results, Models (1)-(3) present results for alternative model specifications using alternative independent variables for *TAX_COM*. Z values are reported in parentheses, based on robust standard errors. In all our models, the joint significance of the predictors in explaining the outcome variable is indicated by the Wald $\chi 2$ test statistic. In all our models, the Hosmer-Lemeshow $\chi 2$ test statistic provides no strong evidence against an acceptable fit to the data. See Appendix III-1 for variable definitions. ***p<0.01, **p<0.05, *p<0.1.

5.2.3. H3: Governance Quality and Corporate Environmental Taxes

Third, consistent with Hypothesis 3 and public interest theory, the coefficient estimate of *GOV_QUA* in our main model, presented in Table III-4, is positive and statistically significant (p<0.01). The average marginal effect of *GOV_QUA* is 0.2712 (untabulated). This means that, on average, a one-unit increase in governance quality increases the probability that a country implements corporate environmental taxes by approximately 27.12 percentage points.[43] This suggests that when governance structures are robust, policymakers may feel more confident in the enforceability and effectiveness of corporate environmental taxes. This finding shows that governance quality is an important factor explaining the cross-country variation in corporate environmental tax implementation.

[43] In our final sample, the mean values of governance quality (*GOV_QUA*) vary significantly across different regions. For instance, on average, countries in "Europe and Central Asia" have a score of 0.589, while countries in "Sub-Saharan Africa" have a significantly lower average score of -0.538.

In Table III-6, we substitute *GOV_QUA* with alternative proxies to address the potential oversimplification inherent in our aggregated measure based on the six World Bank governance indicators.[44] Model (1) relies on the overall freedom status (*GOV_FDM*) from *Freedom House* ($p<0.01$), Model (2) employs electoral democracy and civil liberties (*GOV_DEM*) from *Our World in Data* ($p<0.05$), and Model (3) uses checks and balances on executive power (*GOV_CBL*) from the *Database of Political Institutions* ($p<0.01$). Across all specifications, the coefficients remain statistically significant, supporting Hypothesis 3.

Table III-6. Governance Quality and Corporate Environmental Taxes

Variables	*ENV_TAX*		
	Model (1)	Model (2)	Model (3)
CLI_EXP	0.007	-0.047	-0.112
	(0.041)	(-0.278)	(-0.599)
TAX_COM	-0.045**	-0.055*	-0.033
	(-1.972)	(-1.947)	(-1.547)
GOV_FDM	**0.038*****		
	(3.438)		
GOV_DEM		**3.332****	
		(2.214)	
GOV_CBL			**0.635*****
			(2.801)
FEM_PAR	0.095***	0.133***	0.101***
	(3.043)	(3.664)	(3.344)
Constant	1.231	-0.008	-2.032
	(0.401)	(-0.002)	(-0.653)
Observations	149	138	139
Controls	YES	YES	YES
Pseudo R^2	0.323	0.391	0.325
Wald χ^2	34.86***	45.23***	32.57***
Hosmer-Lemeshow χ^2	9.119	6.417	6.258
Prob>Hosmer-Lemeshow χ^2	0.332	0.601	0.618
VIF mean (max)	1.86 (4.02)	1.86 (4.10)	1.68 (3.21)

Notes: This table presents the logistic regression results for *ENV_TAX* as the dependent variable with available independent and control variables. Compared to our main model results, Models (1)-(3) present results for alternative model specifications using alternative independent variables for *GOV_QUA*. Z values are reported in parentheses, based on robust standard errors. In all our models, the joint significance of the predictors in explaining the outcome variable is indicated by the Wald $\chi2$ test statistic. In all our models, the Hosmer-Lemeshow $\chi2$ test statistic provides no strong evidence against an acceptable fit to the data. See Appendix III-1 for variable definitions. ***$p<0.01$, **$p<0.05$, *$p<0.1$.

[44] We note that when estimating each of the six World Bank Governance Indicators separately, we find that all dimensions exhibit significantly positive coefficients. This consistency across individual indicators suggests that aggregating them into a single governance quality measure (*GOV_QUA*) does not mask heterogeneous or conflicting effects.

5.2.4. H4: Female Political Representation and Corporate Environmental Taxes

Finally, consistent with Hypothesis 4, the coefficient estimate of *FEM_PAR* in our main model, presented in Table III-4, is positive and statistically significant ($p<0.05$) with an average marginal effect of 0.0097 (untabulated). This implies that a one-unit increase in female political representation increases the probability that a country implements a corporate environmental tax by approximately 0.97 percentage points. Thus, female political representation emerges as another critical factor explaining the cross-country variation in corporate environmental tax implementation. Our finding suggests that as gender diversity increases in legislative bodies, the likelihood of environmental taxes being considered a viable environmental policy instrument increases. This finding supports the idea that female legislators, possibly due to socialization and greater sensitivity to social and environmental issues, may be more supportive of corporate accountability through taxation. Increasing women's parliamentary representation could thus contribute to the implementation of more environmentally progressive policies that align fiscal measures with sustainable development goals.

In Table III-7, we replace *FEM_PAR* with three alternative proxies. Model (1) uses women's political rights (*FEM_GOV*) and Model (2) women's relative political power (*FEM_PRT*). Both remain positive and statistically significant ($p<0.01$). Model (3) uses female ministerial representation (*FEM_MIN*), which is statistically insignificant ($p>0.1$). This divergence is informative. While *FEM_PAR*, *FEM_GOV*, and *FEM_PRT* capture women's representation in the legislative branch, where fiscal legislation is debated, amended, and ultimately approved through collective voting, *FEM_MIN* reflects representation in the executive branch, which is primarily involved in agenda-setting and ministry coordination. If our main results merely reflected a generally progressive societal climate, female ministerial representation should exhibit a positive association.[45] The absence of such an association suggests that women's influence on corporate environmental tax implementation operates primarily through their formal voting power in the legislature, rather than through broader gender equality norms or executive- or ministerial-level representation per se.

[45] Untabulated results show that *FEM_MIN* is positively and significantly correlated with female political representation (*FEM_PAR*; $\rho=0.61$, $p<0.01$), women's political rights (*FEM_GOV*; $\rho=0.62$, $p<0.01$), and women's relative political power (*FEM_PRT*; $\rho=0.47$, $p<0.01$).

Table III-7. Female Political Representation and Corporate Environmental Taxes

Variables	*ENV_TAX*		
	Model (1)	Model (2)	Model (3)
CLI_EXP	0.265	0.099	0.301
	(1.260)	(0.413)	(1.445)
TAX_COM	-0.051**	-0.064**	-0.056**
	(-2.207)	(-2.047)	(-2.547)
GOV_QUA	2.652***	1.747**	2.595***
	(3.981)	(2.239)	(3.748)
FEM_GOV	**1.329*** **		
	(2.592)		
FEM_PRT		**7.714*** **	
		(4.188)	
FEM_MIN			**0.014**
			(0.708)
Constant	11.521**	6.688	12.879**
	(2.464)	(1.244)	(2.479)
Observations	148	138	149
Controls	YES	YES	YES
Pseudo R^2	0.322	0.437	0.284
Wald χ^2	35.18***	45.79***	35.75***
Hosmer-Lemeshow χ^2	15.14	8.088	2.936
Prob>Hosmer-Lemeshow χ^2	0.0566	0.425	0.938
VIF mean (max)	2.21 (5.72)	2.29 (6.01)	2.26 (5.80)

Notes: This table presents the logistic regression results for *ENV_TAX* as the dependent variable with available independent and control variables. Compared to our main model results, Models (1)-(3) present results for alternative model specifications using alternative independent variables for *FEM_PAR*. Z values are reported in parentheses, based on robust standard errors. In all our models, the joint significance of the predictors in explaining the outcome variable is indicated by the Wald $\chi 2$ test statistic. In all our models, the Hosmer-Lemeshow $\chi 2$ test statistic provides no strong evidence against an acceptable fit to the data. See Appendix III-1 for variable definitions. ***p<0.01, **p<0.05, *p<0.1.

5.3. Addressing Endogeneity Issues

5.3.1. Additional Control Variables

A common limitation in global cross-country studies is the availability of data for control variables. Nevertheless, we control for a broad set of country characteristics in Equation (III-1) that are likely to explain cross-country variation in corporate environmental tax implementation. Prior research highlights additional factors, such as political systems, trade openness, and ruling-party ideology, as relevant determinants of environmental policy. However, including these variables substantially reduces the final sample size, thereby weakening the global representativeness of the sample. To avoid this trade-off in our main specification, we examine these additional factors in separate robustness analyses. First, we add the dummy variable *PRESIDENTIAL*, capturing whether executive power is concentrated in a directly elected president with strong veto authority and weaker reliance on legislative coalitions. Such systems may face higher barriers to negotiated fiscal-environmental reforms.

Second, we include *TRADE*, measured as imports plus exports over GDP, based on the expectation that more open economies tend to cooperate more on global environmental issues (Neumayer, 2002). Third, we control for the ruling party's ideology (*PARTY_IDEOLOGY*), as right-leaning governments (*RIGHT*) have been shown to be less inclined to advance climate policy than centrist (*CENTER*) and left-leaning governments (*LEFT*) (Gazmararian & Milner, 2024; Schulze, 2021). Table III-8 reports the results. We include each control variable separately in Models (1) to (3), and then together in Model (4).[46] We find that the additional controls behave essentially as expected. Presidential systems are associated with lower probabilities of corporate environmental tax implementation. At the same time, trade openness and left- or center-leaning governments are positively associated. Our baseline findings in Model (1) of Table III-4 remain qualitatively unchanged, thereby mitigating concerns that our results are driven by omitted-variable bias.

Table III-8. Additional Control Variables

Variables	*ENV_TAX*			
	Model (1)	Model (2)	Model (3)	Model (4)
CLI_EXP	0.143	0.277	0.676	0.116
	(0.587)	(1.217)	(1.613)	(0.418)
TAX_COM	-0.065**	-0.071***	0.004	-0.096**
	(-2.181)	(-2.611)	(0.092)	(-2.413)
GOV_QUA	2.096***	2.354***	5.521***	1.720**
	(3.221)	(3.501)	(2.864)	(2.167)
FEM_PAR	0.089**	0.105***	0.125**	0.109***
	(2.450)	(2.811)	(2.178)	(3.265)
PRESIDENTIAL	**-1.142***			**-1.924****
	(-1.681)			**(-2.463)**
TRADE		**0.014****		**0.011****
		(1.972)		**(2.072)**
CENTER.PARTY_IDEOLOGY			**2.893***	
			(1.790)	
LEFT.PARTY_IDEOLOGY			**3.206****	
			(2.167)	
Constant	13.552***	14.366***	14.235	15.125**
	(2.722)	(2.968)	(1.315)	(2.243)
Observations	141	136	79	131
Controls	YES	YES	YES	YES
Pseudo R^2	0.382	0.438	0.555	0.480
Wald χ^2	39.70***	38.95***	27.14***	46.39***
Hosmer-Lemeshow χ^2	4.674	11.05	3.383	8.703
Prob>Hosmer-Lemeshow χ^2	0.792	0.199	0.908	0.368
VIF mean (max)	2.23 (5.96)	2.20 (6.05)	2.57 (7.12)	2.24 (6.15)

Notes: This table presents the logistic regression results for *ENV_TAX* as the dependent variable, with available independent variables, control variables included in Equation (III-1), and additional control variables. Z values

[46] *PARTY_IDEOLOGY* is omitted in Model (4) due to perfect prediction in the full specification.

are reported in parentheses, based on robust standard errors. In all our models, the joint significance of the predictors in explaining the outcome is indicated by the Wald χ^2 test statistic. In all our models, an adequate fit to the data is indicated by the Hosmer-Lemeshow χ^2 test statistic. See Appendix III-1 for variable definitions. ***p<0.01, **p<0.05, *p<0.1.

5.3.2. Oster Test: Robustness of the *FEM_PAR* Effect to Omitted Variables

A central concern in interpreting the positive association between female political representation and corporate environmental tax implementation is that the estimated effect may be biased by unobserved factors that correlate with both variables. In line with recent literature (e.g., Chen et al., 2026; Guo et al., 2026; Lee et al., 2026; Lu et al., 2026; Mayberry et al., 2025), we apply the Oster (2019) test to assess the sensitivity of our estimated effect of female political representation to potential omitted-variable bias. We proceed in two steps. First, we fix the proportional selection parameter at δ=1 and set the hypothetical maximum explanatory power R_{max} to 1.3 times the R-squared of the fully controlled model (R_{max}=0.392). Using these inputs, the bias-adjusted coefficient $\beta^*(R_{max}, \delta=1)$ equals 0.0054, which lies within the 95 percent confidence interval of the original fully controlled estimate for *FEM_PAR* [0.0027; 0.0139]. Second, keeping R_{max} fixed at 0.392, we compute the value of δ that would reduce the coefficient on *FEM_PAR* to zero. The resulting δ=2.058 exceeds the conventional threshold of $|\delta|>1$, indicating that selection on unobservables would need to be more than twice as strong as selection on observables to eliminate the estimated effect. Taken together, these two diagnostics suggest that our findings are robust to plausible levels of omitted variable endogeneity.

5.4. Robustness Checks and Additional Analyses

5.4.1. Alternative Dependent Variables

In Table III-9, we present additional tests using alternative dependent variables derived from our CET-Index to further evaluate the robustness of our findings. First, we use *ENV_TAX_4*, which is the number of environmental tax dimensions implemented. An ordinal response model is used to account for the variable's ordinal nature (Greene, 2019; Verbeek, 2017). Although the CET-Index is formative, the number of implemented dimensions still reflects meaningful variation in the breadth of a country's corporate environmental tax policy mix. Examining this ordered outcome, therefore, provides a useful robustness check on whether our determinants also explain differences in how extensive corporate environmental taxation is. The results are presented in Model (1) and broadly support the main conclusions.

Notably, *TAX_COM* turns insignificant, indicating that the influence of corporate tax competition does not extend uniformly across all four tax dimensions.

The result presented in Model (1) is confirmed when using the four environmental tax dimensions individually in Models (2) to (5). Specifically, corporate tax competition is negatively associated with energy taxes at conventional levels of significance ($p<0.05$). Pollution taxes exhibit only weak evidence of an association ($p<0.1$), while the coefficients for resource and transport taxes are statistically insignificant ($p>0.1$). Thus, the effect of corporate tax competition appears to be concentrated primarily in the energy tax dimension. Several mechanisms may explain why corporate tax competition primarily explains cross-country variation in energy tax implementation, whereas the other environmental tax dimensions are not systematically shaped by corporate tax competition. Pollution taxes primarily target waste streams generated by fixed production facilities, which may limit firms' ability to avoid them through cross-border relocation. Resource taxes apply to sectors such as mining, forestry, or fishing, where production is tied to immobile natural endowments, making these taxes largely insulated from competitive pressures. Transport taxes are similarly anchored in domestic infrastructure and local service networks, which may reduce relocation risks. By contrast, energy taxes materially increase operating costs for a wide range of internationally mobile firms, offering significant leverage for tax-induced relocation. In addition, energy taxes account for the majority of total environmental tax revenue.[47] This may make energy taxes uniquely sensitive to corporate tax competition relative to the other three dimensions. Additionally, *GOV_QUA* is strongly and positively associated with *ENE_TAX* and *TRA_TAX* ($p<0.01$), while its effect is weak for *POL_TAX* ($p<0.1$) and absent for *RES_TAX* ($p>0.1$). This pattern is consistent with the expectation that countries with higher governance quality prioritize the most consequential mitigation instruments. Energy and transport taxes are most directly tied to carbon mitigation, making them the natural focus of governance-driven policy action. Pollution and resource taxes, by contrast, are less closely aligned with carbon externalities, which may explain why governance quality does not systematically predict their implementation within our global sample.

[47] For instance, in 2023, energy taxes made up roughly 76% of all environmental tax revenue in the EU, far exceeding transport (19%) and pollution/resource taxes (5%) (EEA, 2025).

Finally, we use *ENV_TAX_INC*, which equals "1" if a country has implemented at least one corporate environmental tax or at least one of the following six corporate income tax-based environmental incentives: tax holidays, reduced tax rates, loss carryovers, investment allowances, tax credits, or accelerated depreciation, and "0" otherwise. Compared to our main results, Model (6) remains unchanged when we expand the primary CET-Index measure (*ENV_TAX*) to include corporate income tax-based environmental incentives.

Table III-9. Alternative Dependent Variables

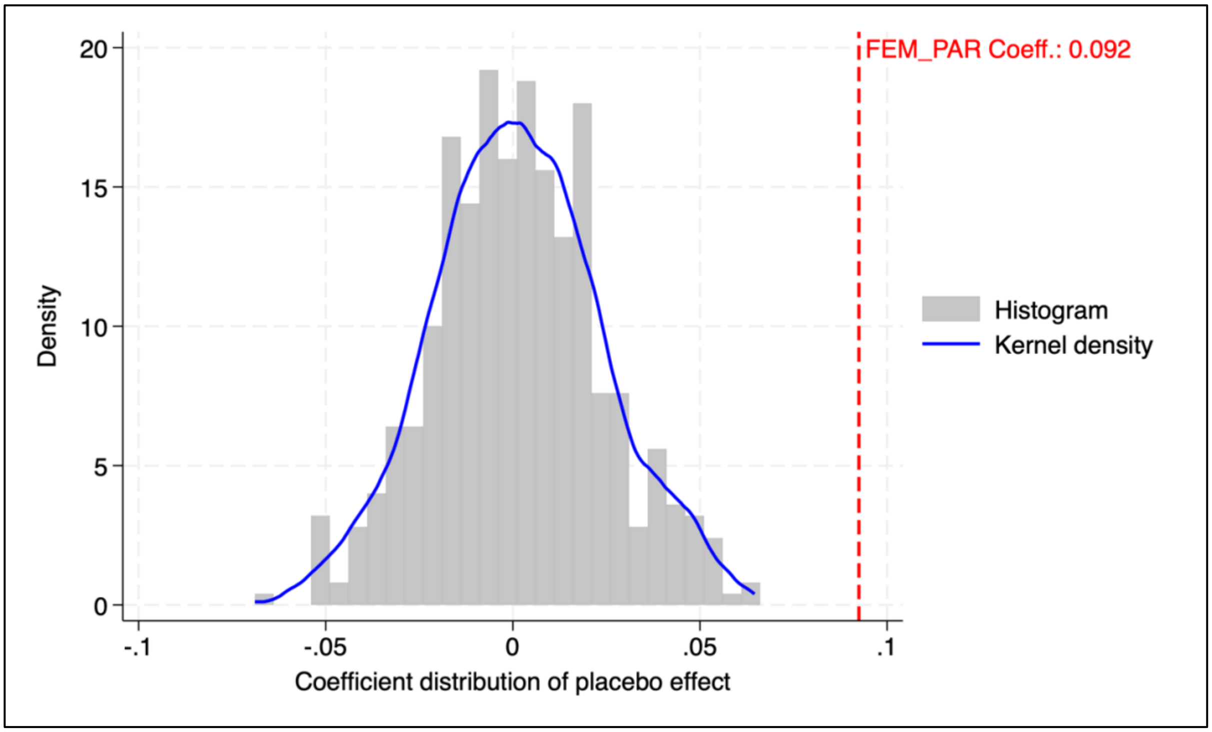
Variables	***ENV_ TAX_4***	***ENE_ TAX***	***POL_ TAX***	***RES_ TAX***	***TRA_ TAX***	***ENV_ TAX_INC***
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
CLI_EXP	-0.027	0.197	-0.224	-0.062	-0.033	0.498
	(-0.272)	(1.265)	(-1.378)	(-0.501)	(-0.267)	-1.501
TAX_COM	-0.016	-0.035**	-0.027*	0.009	-0.009	-0.043*
	(-1.487)	(-2.230)	(-1.742)	(0.667)	(-0.606)	(-1.909)
GOV_QUA	1.357***	2.427***	0.892*	0.065	1.391***	2.861***
	(4.068)	(4.162)	(1.944)	(0.169)	(3.214)	-3.944
FEM_PAR	0.042**	0.032	0.042*	0.034*	0.041**	0.113**
	(2.522)	(1.643)	(1.905)	(1.810)	(2.201)	-2.112
Constant		7.700**	0.030	1.223	5.384	11.951**
		(2.114)	(0.007)	(0.378)	(1.536)	-2.362
Observations	149	149	149	149	149	149
Controls	YES	YES	YES	YES	YES	YES
Pseudo R^2	0.132	0.245	0.260	0.148	0.182	0.478
Wald χ^2	68.77***	33.98***	32.97***	24.24***	31.57***	53.49***
Hosmer-Lemeshow χ^2	26.357	5.884	8.410	9.937	12.15	4.523
Prob>Hosmer-Lemeshow χ^2	0.854	0.660	0.394	0.270	0.144	0.807
VIF mean (max)	2.21 (5.79)					

Notes: This table presents the regression results for six alternative dependent variables with available independent and control variables. Model (1) provides the ordered logistic regression result using *ENV_TAX_4*. Models (2) to (5) provide the logistic regression results using *ENE_TAX*, *POL_TAX*, *RES_TAX*, and *TRA_TAX*. Model (6) provides the logistic regression result using *ENV_TAX_INC*. Z values are reported in parentheses, based on robust standard errors. In all our models, the joint significance of the predictors in explaining the outcome variables is indicated by the Wald χ^2 test statistic. In all (ordered) logistic regression models, an adequate fit to the data is indicated by the (ordinal) Hosmer-Lemeshow χ^2 test statistic. In Model (1), there is no single constant term, but cut-points for the different outcomes of *ENV_TAX_4*, which are not reported for the sake of brevity. See Appendix III-1 for variable definitions. ***p<0.01, **p<0.05, *p<0.1.

5.4.2. Placebo Test: Female Political Representation

To address the concern that the estimated effect of female political representation may reflect spurious correlation rather than a substantive empirical relationship, we follow Wu et al. (2025) by conducting a placebo test. Specifically, we randomly reshuffle *FEM_PAR* across countries 500 times, re-estimating our main specification in each iteration. This procedure preserves the joint distribution of all covariates while breaking any systematic link between female representation and corporate environmental tax implementation. As presented in Figure III-1, the resulting distribution of placebo coefficients is centered around zero and approximately symmetric, consistent with the null hypothesis of no effect. Importantly, the actual coefficient (0.092) lies far in the right tail of the placebo distribution. This indicates that the estimated relationship is not an artifact of random assignment and provides additional support for a meaningful empirical association between female political representation and corporate environmental tax implementation.

Figure III-1. Placebo Test: Robustness of the *FEM_PAR* Effect to Spurious Correlation



Notes: This figure shows the distribution of *FEM_PAR* coefficients obtained from 500 placebo regressions with randomly permuted female representation, along with the actual coefficient from our main model.

5.4.3. Critical Mass Test: Female Political Representation

Kanter's critical mass theory suggests that a certain threshold or critical mass of underrepresented group members is necessary for these individuals to influence group dynamics and outcomes meaningfully (Kanter, 1977a, 1977b). Prior to reaching this critical mass, women are perceived as mere tokens, controlled by the dominant majority group. However, once critical mass is achieved, gender composition becomes less of a factor of influence by itself because women are no longer seen as outsiders or tokens and are fully integrated into the decision-making process. In our research setting, we assume that women in parliament need to reach a critical mass to impact corporate environmental tax policymaking. However, this expected positive effect is likely to diminish in nearly balanced or fully balanced political bodies.

Following the seminal work of Kanter (1977a, 1977b) as well as applications of the critical mass theory (e.g., Joecks et al., 2013; Pandey et al., 2020; Seebeck & Vetter, 2022), we construct different categories of groups according to their gender composition: skewed groups, tilted groups, and balanced groups.[48] Kanter (1977a) emphasizes that relative numbers determine group dynamics and influence. Accordingly, our application of the critical mass theory relies on proportional representation.[49] Specifically, Kanter's critical mass theory suggests that a male-dominated, "skewed" group has a male-female ratio of perhaps 85:15 (Kanter, 1977a, p. 208); a "tilted" group of perhaps 65:35 (Kanter, 1977a, p. 209); and a "balanced" group at about 60:40, down to 50:50 (Kanter, 1977a, p. 209). Building on these theoretical thresholds, accounting research commonly operationalizes Kanter's framework by defining groups with up to 20% women as "skewed," those with 20–40% women as "tilted," and those with 40–60% women as "balanced" (e.g., Joecks et al., 2013; Pandey et al., 2020; Seebeck & Vetter, 2022).

Model (1) in Table III-10 shows that *SKEWED* (≤ 20% women) has a negative and significant coefficient ($p<0.01$), indicating that women, as a small minority, lack influence on corporate environmental tax policymaking. In contrast, Model (2) shows a positive and significant effect for *TILTED* (20–40% women; $p<0.01$), suggesting that once female representation in national parliaments reaches a critical mass, women collectively begin to shape policy outcomes. The effect becomes insignificant in Model (3) for *BALANCED* (40–60% women; $p>0.1$), implying that as gender parity approaches, other factors take precedence. These results support Kanter's (1977a, 1977b) critical mass theory and align with prior evidence that similar thresholds enhance group effects (e.g., Joecks et al., 2013; Pandey et al.,

[48] In Kanter's seminal work, she distinguishes four group types: uniform, skewed, tilted, and balanced. In uniform groups, all members share the same characteristic. In our sample, only two observations are entirely male, providing insufficient within-group variation for statistical estimation. Therefore, we exclude these two observations (Kanter, 1977a, 1977b); One observation with *FEM_PAR*=61.25%, exceeds the commonly used thresholds, albeit slightly. Note that this observation was placed in the *BALANCED* category, given its proximity to the upper limit of the balanced group.

[49] Due to cross-country variation in national parliament size, *FEM_PAR* captures the percentage of female parliamentary members rather than the absolute number of parliamentary seats.

2020; Seebeck & Vetter, 2022). Thus, a critical mass of female parliamentarians is necessary before they can have a favorable impact on corporate environmental tax policy.[50] These results reinforce that substantive influence arises from a critical mass of women in the legislature, rather than from token female representation in executive or ministerial positions.

Table III-10. Critical Mass Theory and Female Political Representation

Variables	*ENV_TAX*		
	Skewed	Tilted	Balanced
	Model (1)	Model (2)	Model (3)
CLI_EXP	0.173	0.209	0.268
	(0.812)	(0.993)	(1.275)
TAX_COM	-0.079***	-0.084***	-0.062***
	(-3.027)	(-3.314)	(-2.608)
GOV_QUA	2.722***	2.943***	2.618***
	(4.125)	(4.393)	(3.746)
SKEWED	**-1.796*** **		
	(-3.105)		
TILTED		**1.787*** **	
		(3.101)	
BALANCED			**0.338**
			(0.314)
Constant	17.868***	17.300***	14.416***
	(3.626)	(3.384)	(2.764)
Observations	147	147	147
Controls	YES	YES	YES
Pseudo R^2	0.355	0.351	0.284
Wald χ^2	50.11***	46.87***	34.44***
Hosmer-Lemeshow χ^2	8.628	5.670	5.344
Prob>Hosmer-Lemeshow χ^2	0.375	0.684	0.720
VIF mean (max)	2.19 (5.72)	2.19 (5.79)	2.20 (5.75)

Notes: This table presents the logistic regression results for *ENV_TAX* as the dependent variable on dummy variables for different types of political bodies, other independent variables, and control variables. In Model (1), skewed political bodies consist of up to 20% women. In Model (2), tilted political bodies consist of more than 20% up to 40% women. In Model (3), balanced political bodies consist of more than 40% up to 60% women. Z values are reported in parentheses, based on robust standard errors. In all our models, the joint significance of the predictors in explaining the outcome is indicated by the Wald χ^2 test statistic. In all our models, an adequate fit to the data is indicated by the Hosmer-Lemeshow χ^2 test statistic. See Appendix III-1 for variable definitions. ***p<0.01, **p<0.05, *p<0.1.

6. Conclusion and Limitations

Our cross-country study is motivated by the need to understand why some countries implement corporate environmental taxes, an instrument widely endorsed for internalizing environmental externalities, while others do not. To address this gap, we developed a novel index (CET-Index) that systematically captures the presence of corporate energy, pollution,

[50] Albeit the slight differences in range specification, we test the thresholds originally proposed by Kanter (1977a, 1977b) and find that our results remain qualitatively unchanged.

resource, and transport taxes through a rigorous multi-stage process combining text mining, web scraping, and structured human coding. Using the CET-Index, we conduct a series of cross-country logistic regression analyses to examine the influence of climate risk, corporate tax competition, governance quality, and female political representation on the likelihood of corporate environmental tax implementation. The results show that climate risk does not explain the implementation of corporate environmental taxes. In contrast, greater corporate tax competition significantly decreases the likelihood of implementation, while higher governance quality and stronger female political representation increase it. Additional analyses indicate that the gender effect operates through parliamentary mechanisms and requires a critical mass of female parliamentarians. Overall, the study contributes the first instrument-specific evidence on the determinants of corporate environmental taxes, introduces a novel index to support comparative policy research, and highlights the importance of political, institutional, and gender-related factors for the implementation of corporate environmental taxes.

We acknowledge several limitations. First, despite conducting several robustness checks, we cannot completely rule out the possibility that our results are biased due to the omission of variables. Second, the CET-Index reflects the status quo as of December 31, 2022, resulting in a cross-sectional research design. While this design enables explaining cross-country variations in corporate environmental tax implementation, it limits the ability to draw causal inferences, as temporal dynamics and policy changes over time cannot be observed. Third, to enable instrument-specific cross-country analysis, our CET-Index focuses on national-level corporate environmental taxes. This scope necessarily omits environmental tax incentives, individual-level environmental taxes, subnational tax measures, and other instruments such as subsidies or emission trading systems, meaning it does not fully capture the breadth of countries' environmental policy frameworks. Fourth, despite rigorous scoring, including web scraping, text mining, multiple scorers, and third-party validation, some misclassification risk remains. To ensure transparency, the CET-Index, its four dimensions, and its underlying sources are provided as an exhibit.

7. References

Adams, R. B., & Funk, P. (2012). Beyond the Glass Ceiling: Does Gender Matter? *Management Science*, *58*(2), 219–235. https://doi.org/10.1287/mnsc.1110.1452

Adrian, C., Garg, M., Pham, A. V., Phang, S.-Y., & Truong, C. (2023). Do Natural Disasters Affect Corporate Tax Avoidance? The Case of Drought. *Journal of Business Ethics*, *186*(1), 105–135. https://doi.org/10.1007/s10551-022-05250-7

Al Mamun, M., Boubaker, S., Hossain, M. Z., & Manita, R. (2024). Female political empowerment and green finance. *Energy Economics*, *131*, 107370. https://doi.org/10.1016/j.eneco.2024.107370

Ballard, C. L., & Medema, S. G. (1993). The marginal efficiency effects of taxes and subsidies in the presence of externalities. *Journal of Public Economics*, *52*(2), 199–216. https://doi.org/10.1016/0047-2727(93)90020-T

Bättig, M. B., & Bernauer, T. (2009). National Institutions and Global Public Goods: Are Democracies More Cooperative in Climate Change Policy? *International Organization*, *63*(2), 281–308. https://doi.org/10.1017/S0020818309090092

Baumgartner, F. R., & Jones, B. D. (1991). Agenda Dynamics and Policy Subsystems. *Journal of Politics*, *53*(4), 1044–1074. https://doi.org/10.2307/2131866

Baumol, W. J., & Oates, W. E. (1988). *The Theory of Environmental Policy* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139173513

Bird, R., & Davis-Nozemack, K. (2018). Tax Avoidance as a Sustainability Problem. *Journal of Business Ethics*, *151*(4), 1009–1025. https://doi.org/10.1007/s10551-016-3162-2

Birkland, T. A. (2006). *Lessons of Disaster: Policy Change after Catastrophic Events*. Georgetown University Press. http://www.jstor.org/stable/j.ctt2tt2sn

Bonetti, P., Cho, C. H., & Michelon, G. (2024). Environmental Disclosure and the Cost of Capital: Evidence from the Fukushima Nuclear Disaster. *European Accounting Review*, *33*(5), 1693–1721. https://doi.org/10.1080/09638180.2023.2203410

Bradford, D. F., & Oates, W. E. (1971). The Analysis of Revenue Sharing in a New Approach to Collective Fiscal Decisions. *Quarterly Journal of Economics*, *85*(3), 416–439. https://doi.org/10.2307/1885931

Breyer, S. (1982). *Regulation and Its Reform*. Harvard University Press. http://www.jstor.org/stable/j.ctv1smjtgr

Brown, J. R., Martinsson, G., & Thomann, C. (2022). Can Environmental Policy Encourage Technical Change? Emissions Taxes and R&D Investment in Polluting Firms. *Review of Financial Studies*, *35*(10), 4518–4560. https://doi.org/10.1093/rfs/hhac003

Chava, S. (2014). Environmental Externalities and Cost of Capital. *Management Science*, *60*(9), 2223–2247. https://doi.org/10.1287/mnsc.2013.1863

Chen, Y., Ofosu, E., & Pittman, J. (2026). The Cost of Investor Protection: Bank Loan Contracting During SEC Investigations. *Accounting Review*, *101*(1), 203–234. https://doi.org/10.2308/TAR-2024-0405

Chijoke-Mgbame, A. M., Boateng, A., & Mgbame, C. O. (2020). Board gender diversity, audit committee and financial performance: evidence from Nigeria. *Accounting Forum*, *44*(3), 262–286. https://doi.org/10.1080/01559982.2020.1766280

Cologna, V., Meiler, S., Kropf, C. M., Lüthi, S., Mede, N. G., Bresch, D. N., Lecuona, O., Berger, S., Besley, J., Brick, C., Joubert, M., Maibach, E. W., Mihelj, S., Oreskes, N., Schäfer, M. S., Linden, S. van der, Aziz, N. I. A., Abdulsalam, S., Shamsi, N. A., … Zenklusen, A. (2025). Extreme weather event attribution predicts climate policy support across the world. *Nature Climate Change*, *15*(7), 725–735. https://doi.org/10.1038/s41558-025-02372-4

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Crowley, T. J. (2000). Causes of Climate Change Over the Past 1000 Years. *Science*, *289*(5477), 270–277. https://doi.org/10.1126/science.289.5477.270

Dabla-Norris, E., Helbling, T., Khalid, S., Khan, H., Magistretti, G., Sollaci, A., & Srinivasan, K. (2023). Public Perceptions of Climate Mitigation Policies: Evidence from Cross-Country Surveys. Staff Discussion Note SDN2023/002. Washington, DC, February. https://www.imf.org/-/media/Files/Publications/SDN/2023/English/SDNEA2023002.ashx

Dablander, F. (2025). Climate hazard experience linked to increased climate risk perception worldwide. *Environmental Research Letters*, *20*(11), 114010. https://doi.org/10.1088/1748-9326/ae0ae9

Damania, R., Fredriksson, P. G., & List, J. A. (2003). Trade liberalization, corruption, and environmental policy formation: theory and evidence. *Journal of Environmental Economics and Management*, *46*(3), 490–512. https://doi.org/10.1016/S0095-0696(03)00025-1

De Groote, S., Bruynseels, L., & Gaeremynck, A. (2023). Are All Directors Treated Equally? Evidence from Director Turnover Following Opportunistic Insider Selling. *Journal of Business Ethics*, *185*(1), 185–207. https://doi.org/10.1007/s10551-022-05127-9

Devereux, M. P. (2023). International Tax Competition and Coordination with A Global Minimum Tax. *National Tax Journal*, *76*(1), 145–166. https://doi.org/10.1086/723198

Devereux, M. P., & Loretz, S. (2013). What do we know about corporate tax competition? *National Tax Journal*, *66*(3), 745–773. https://doi.org/10.17310/ntj.2013.3.08

Devereux, M. P., Lockwood, B., & Redoano, M. (2008). Do countries compete over corporate tax rates? *Journal of Public Economics*, *92*(5–6), 1210–1235. https://doi.org/10.1016/j.jpubeco.2007.09.005

Di Pietro, F., & Masciarelli, F. (2022). The Effect of Local Religiosity on Financing Cross-Regional Entrepreneurial Projects Via Crowdfunding (Local Religiosity and Crowdfinancing). *Journal of Business Ethics*, *178*(2), 429–443. https://doi.org/10.1007/s10551-021-04805-4

Duan, Y., Zhang, Z., Li, Y., Wang, S., Yang, C., & Lu, Y. (2024). Global corporate tax competition challenges climate change mitigation. *Nature Climate Change*, *14*(4), 353–356. https://doi.org/10.1038/s41558-024-01952-0

Eagly, A. H. (1987). *Sex Differences in Social Behavior*. Psychology Press. https://doi.org/10.4324/9780203781906

Eagly, A. H., & Karau, S. J. (1991). Gender and the emergence of leaders: A meta-analysis. *Journal of Personality and Social Psychology*, *60*(5), 685–710. https://doi.org/10.1037/0022-3514.60.5.685

Eagly, A. H., & Wood, W. (1991). Explaining Sex Differences in Social Behavior: A Meta-Analytic Perspective. *Personality and Social Psychology Bulletin*, *17*(3), 306–315. https://doi.org/10.1177/0146167291173011

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin*, *117*(1), 125–145. https://doi.org/10.1037/0033-2909.117.1.125

European Environment Agency (EEA) (2022). *The role of (environmental) taxation in supporting sustainability transitions,* Publications Office of the European Union. https://doi.org/10.2800/24487

European Environment Agency (EEA) (2025, October 3). *Share of environmental taxes in total tax revenues in Europe*. https://www.eea.europa.eu/en/analysis/indicators/share-of-environmental-taxes-in

Fredriksson, P. G., & Neumayer, E. (2013). Democracy and climate change policies: Is history important? *Ecological Economics*, *95*, 11–19. https://doi.org/10.1016/j.ecolecon.2013.08.002

Fredriksson, P. G., & Svensson, J. (2003). Political instability, corruption and policy formation: the case of environmental policy. *Journal of Public Economics*, *87*(7–8), 1383–1405. https://doi.org/10.1016/S0047-2727(02)00036-1

Fredriksson, P. G., & Wang, L. (2011). Sex and environmental policy in the U.S. House of Representatives. *Economics Letters*, *113*(3), 228–230. https://doi.org/10.1016/j.econlet.2011.07.019

García-Lara, J. M., García Osma, B., Mora, A., & Scapin, M. (2017). The monitoring role of female directors over accounting quality. *Journal of Corporate Finance*, *45*, 651–668. https://doi.org/10.1016/j.jcorpfin.2017.05.016

Gazmararian, A. F., & Milner, H. v. (2024). Political Cleavages and Changing Exposure to Global Warming. *Comparative Political Studies*. *58*(9), 1965–1999. https://doi.org/10.1177/00104140241283014

Glaeser, S. A., Olbert, M., & Werner, A.-C. (2023). Tax Competition and Employment. *Accounting Review*, *98*(5), 267–296. https://doi.org/10.2308/TAR-2020-0575

Goulder, L. H. (1995). Environmental taxation and the double dividend: A reader's guide. *International Tax and Public Finance*, *2*(2), 157–183. https://doi.org/10.1007/BF00877495

Greene, W. (2019). *Econometric Analysis* (8th ed.). Pearson International.

Guo, C., Luo, J., & Shi, J. (2026). Navigating Ethical Waters: Trade Credit Provisions for Related-Party Customers. *Journal of Business Ethics*, *204*, 479–504. https://doi.org/10.1007/s10551-025-06100-y

Haque, F., Adjei-Mensah, G., Nguyen, T. H. H., & Ntim, C. G. (2026). Does gender diversity in corporate boards and executive management teams influence carbon performance? Evidence from Europe. *Accounting Forum*, *50*(2), 215–244. https://doi.org/10.1080/01559982.2024.2423989

Hardeck, I., Hechtner, F., Seebeck, A., & Weiß, M. (2024). Does Public Tax Sustainability Disclosure Deter Corporate Tax Avoidance? Evidence from GRI 207 Reporting. Available at SSRN: https://ssrn.com/abstract=4910591

Hoffmann, R., Muttarak, R., Peisker, J., & Stanig, P. (2022). Climate change experiences raise environmental concerns and promote Green voting. *Nature Climate Change*, *12*(2), 148–155. https://doi.org/10.1038/s41558-021-01263-8

Hoppe, T., Schanz, D., Sturm, S., & Sureth-Sloane, C. (2023). The Tax Complexity Index – A Survey-Based Country Measure of Tax Code and Framework Complexity. *European Accounting Review*, *32*(2), 239–273. https://doi.org/10.1080/09638180.2021.1951316

Intergovernmental Panel on Climate Change (IPCC) (2022a). *Climate Change 2022: Impacts, Adaptation and Vulnerability Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. https://www.ipcc.ch/report/ar6/wg2/downloads/report/IPCC_AR6_WGII_FullReport.pdf

Intergovernmental Panel on Climate Change (IPCC) (2022b). *Climate Change 2022 Mitigation of Climate Change Working Group III contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. https://www.ipcc.ch/report/ar6/wg3/downloads/report/IPCC_AR6_WGIII_FullReport.pdf

International Monetary Fund (IMF) (2024, November 28). *Environmental Taxes*. https://climatedata.imf.org/datasets/3fb1ed30d3394574b3145246846023b1_0/about

International Monetary Fund (IMF) (2019). *Fiscal Monitor: How to Mitigate Climate Change*. Washington, DC, October. https://www.imf.org/-/media/Files/Publications/fiscal-monitor/2019/October/English/text.ashx

Ivanov, I. T., Kruttli, M. S., & Watugala, S. W. (2024). Banking on Carbon: Corporate Lending and Cap-and-Trade Policy. *Review of Financial Studies*, *37*(5), 1640–1684. https://doi.org/10.1093/rfs/hhad085

Jacob, M., & Zerwer, K. L. (2024). Emission Taxes and Capital Investments: The Role of Tax Incidence. *Accounting Review*, *99*(5), 247–278. https://doi.org/10.2308/TAR-2023-0053

Joecks, J., Pull, K., & Vetter, K. (2013). Gender Diversity in the Boardroom and Firm Performance: What Exactly Constitutes a “Critical Mass?”. *Journal of Business Ethics*, *118*(1), 61–72. https://doi.org/10.1007/s10551-012-1553-6

Kanter, R. M. (1977a). *Men and women of the corporation*. Basic Books.

Kanter, R. M. (1977b). Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women. *American Journal of Sociology*, *82*(5), 965–990. https://doi.org/10.1086/226425

Kaya, D., & Seebeck, A. (2019). The Dissemination of Firm Information via Company Register Websites: Country-Level Empirical Evidence. *Journal of Accounting & Organizational Change*, *15*(3), 382–429. https://doi.org/10.1108/JAOC-03-2018-0023

Kock, C. J., & Min, B. S. (2016). Legal Origins, Corporate Governance, and Environmental Outcomes. *Journal of Business Ethics*, *138*(3), 507–524. https://doi.org/10.1007/s10551-015-2617-1

Kremer, M., & Willis, J. (2016). Guns, Latrines, and Land Reform: Dynamic Pigouvian Taxation. *American Economic Review*, *106*(5), 83–88. https://doi.org/10.1257/aer.p20161096

La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2008). The Economic Consequences of Legal Origins. *Journal of Economic Literature*, *46*(2), 285–332. https://doi.org/10.1257/jel.46.2.285

Lee, J., Ng, S., Yoo, I. S., & Zhang, L. (2026). Freedom of Expression Protection and Corporate Concealment of Bad News: Evidence from State Anti-SLAPP Laws. *Journal of Accounting Research*, *64*(1), 365–415. https://doi.org/10.1111/1475-679X.70012

Levitus, S., Antonov, J. I., Wang, J., Delworth, T. L., Dixon, K. W., & Broccoli, A. J. (2001). Anthropogenic Warming of Earth’s Climate System. *Science*, *292*(5515), 267–270. https://doi.org/10.1126/science.1058154

Liang, H., & Renneboog, L. (2017). On the Foundations of Corporate Social Responsibility. *Journal of Finance*, *72*(2), 853–910. https://doi.org/10.1111/jofi.12487

Lu, Y., Wu, J. S., & Yang, Y. (John). (2026). Does Foreign Investors’ Information Access Vary with Geopolitical Tensions? Evidence from Corporate Conference Calls. *Journal of Accounting Research*, *64*(1), 417–475. https://doi.org/10.1111/1475-679X.70009

Martin, R., de Preux, L. B., & Wagner, U. J. (2014). The impact of a carbon tax on manufacturing: Evidence from microdata. *Journal of Public Economics*, *117*, 1–14. https://doi.org/10.1016/j.jpubeco.2014.04.016

Martinsson, G., Sajtos, L., Strömberg, P., & Thomann, C. (2024). The Effect of Carbon Pricing on Firm Emissions: Evidence from the Swedish CO2 Tax. *Review of Financial Studies*, *37*(6), 1848–1886. https://doi.org/10.1093/rfs/hhad097

Mavisakalyan, A., & Tarverdi, Y. (2019). Gender and climate change: Do female parliamentarians make difference? *European Journal of Political Economy*, *56*, 151–164. https://doi.org/10.1016/j.ejpoleco.2018.08.001

May, A. M., McGarvey, M. G., Gustafson, C. R., & Mieno, T. (2021). Gender, environmental issues and policy: An examination of the views of male and female economists. *Ecological Economics*, *182*, 106877. https://doi.org/10.1016/j.ecolecon.2020.106877

Mayberry, M., Nipper, M., & Weiß, M. (2025). *Personal Income Taxes and Corporate Pollution*. Available at SSRN: https://ssrn.com/abstract=5717802

Mouti, K., Kilincarslan, E., & Li, J. (2026). Environmental Credit Risk, Climate Change and Bank Performance: Evidence From a Global Panel of Banks. *Business Strategy and the Environment*, *35*(2), 1646–1666. https://doi.org/10.1002/bse.70259

Neumayer, E. (2002). Does Trade Openness Promote Multilateral Environmental Cooperation? *World Economy*, *25*(6), 815–832. https://doi.org/10.1111/1467-9701.00464

Oates, W. E. (1972). *Fiscal Federalism*. Harcourt Brace Jovanovich.

Oradi, J., & E-Vahdati, S. (2021). Female directors on audit committees, the gender of financial experts, and internal control weaknesses: evidence from Iran. *Accounting Forum*, *45*(3), 273–306. https://doi.org/10.1080/01559982.2021.1920127

Organization for Economic Co-operation and Development (OECD) (2024). *Environmentally related tax revenue - Dataset documentation*. https://stats.oecd.org/wbos/fileview2.aspx?IDFile=4f959e36-628a-4196-9012-b422b2e0ad40

Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics*, *37*(2), 187–204. https://doi.org/10.1080/07350015.2016.1227711

Pandey, R., Biswas, P. K., Ali, M. J., & Mansi, M. (2020). Female directors on the board and cost of debt: evidence from Australia. *Accounting & Finance*, *60*(4), 4031–4060. https://doi.org/10.1111/acfi.12521

Park, H., Lee, J., & Shin, J. (2024). How Does Taxation Affect Corporate Social Responsibility? Evidence from a Korean Tax Reform. *Journal of Business Ethics*, *192*(4), 745–774. https://doi.org/10.1007/s10551-023-05536-4

Pigou, A. (2017). *The economics of welfare*. Routledge.

Powell, G. N. (1990). One More Time: Do Female and Male Managers Differ? *Academy of Management Perspectives*, *4*(3), 68–75. https://doi.org/10.5465/ame.1990.4274684

Ramanna, K., & Sletten, E. (2014). Network Effects in Countries' Adoption of IFRS. *Accounting Review*, *89*(4), 1517–1543. https://doi.org/10.2308/accr-50717

Ramstetter, L., & Habersack, F. (2020). Do women make a difference? Analysing environmental attitudes and actions of Members of the European Parliament.

Environmental Politics, *29*(6), 1063–1084. https://doi.org/10.1080/09644016.2019.1609156

Rauscher, M. (1995). Environmental regulation and the location of polluting industries. *International Tax and Public Finance*, *2*(2), 229–244. https://doi.org/10.1007/BF00877499

Ryan, T. P. (2008). *Modern regression methods*. John Wiley & Sons.

Schanz, D., Keller, S., Dinkel, A., Fritz, J., & Grosselfinger, C. (2017). The Tax Attractiveness Index: Methodology. Available at SSRN: https://doi.org/10.2139/ssrn.3013603

Schulze, K. (2021). Policy Characteristics, Electoral Cycles, and the Partisan Politics of Climate Change. *Global Environmental Politics*, *21*(2), 44–72. https://doi.org/10.1162/glep_a_00593

Seebeck, A., & Vetter, J. (2022). Not Just a Gender Numbers Game: How Board Gender Diversity Affects Corporate Risk Disclosure. *Journal of Business Ethics*, *177*(2), 395–420. https://doi.org/10.1007/s10551-020-04690-3

Shi, A. (2003). The impact of population pressure on global carbon dioxide emissions, 1975–1996: evidence from pooled cross-country data. *Ecological Economics*, *44*(1), 29–42. https://doi.org/10.1016/S0921-8009(02)00223-9

Stern, N. (2008). The Economics of Climate Change. *American Economic Review*, *98*(2), 1–37. https://doi.org/10.1257/aer.98.2.1

Sterner, T., Barbier, E. B., Bateman, I., van den Bijgaart, I., Crépin, A.-S., Edenhofer, O., Fischer, C., Habla, W., Hassler, J., Johansson-Stenman, O., Lange, A., Polasky, S., Rockström, J., Smith, H. G., Steffen, W., Wagner, G., Wilen, J. E., Alpízar, F., Azar, C., … Robinson, A. (2019). Policy design for the Anthropocene. *Nature Sustainability*, *2*(1), 14–21. https://doi.org/10.1038/s41893-018-0194-x

United Nations Entity for Gender Equality and the Empowerment of Women (UN WOMEN) & United Nations Department of Economic and Social Affairs (UN DESA). (2024). *The Progress on the Sustainable Development Goals: The Gender Snapshot 2024*. https://unstats.un.org/sdgs/gender-snapshot/2024/GenderSnapshot2024.pdf

United Nations Entity for Gender Equality and the Empowerment of Women (UN WOMEN). (2022, February 27). *Explainer: How gender inequality and climate change are interconnected*. https://wrd.unwomen.org/explore/insights/explainer-how-gender-inequality-and-climate-change-are-interconnected

Verbeek, M. (2017). *A guide to modern econometrics*. John Wiley & Sons.

Von Stein, J. (2022). Democracy, Autocracy, and Everything in Between: How Domestic Institutions Affect Environmental Protection. *British Journal of Political Science*, *52*(1), 339–357. https://doi.org/10.1017/S000712342000054X

Weitzman, M. L. (1974). Prices vs. Quantities. *Review of Economic Studies*, *41*(4), 477–491. https://doi.org/10.2307/2296698

Wilson, J. D., & Wildasin, D. E. (2004). Capital tax competition: bane or boon. *Journal of Public Economics*, *88*(6), 1065–1091. https://doi.org/10.1016/S0047-2727(03)00057-4

Wu, D., Xie, Y., & Pei, Y. (2025). Moving from Unethical Competition to Sustainability: Understanding the Impact of Competitive Ethics Review on Carbon Reduction of Digital Enterprise. *Journal of Business Ethics*. Advance online publication. https://doi.org/10.1007/s10551-025-06135-1

Xu, R., & Ren, L. (2024). Natural Disaster, Tax Avoidance, and Corporate Pollution Emissions: Evidence from China. *Journal of Business Ethics*, *197*, 195–217. https://doi.org/10.1007/s10551-024-05716-w

Yamazaki, A. (2022). Environmental taxes and productivity: Lessons from Canadian manufacturing. *Journal of Public Economics*, *205*, 104560. https://doi.org/10.1016/j.jpubeco.2021.104560

8. Appendices

8.1. Appendix III-1. Environmental Policy Instruments

Broadly, government responses to negative environmental externalities caused by corporate activities can be categorized into two categories: command-and-control instruments and market-based instruments (Bohm & Russel, 1985). Command-and-control instruments typically impose direct rules or standards that firms must follow. Such regulations are prescriptive, requiring corporations to follow specific procedures or meet set thresholds, regardless of their individual costs or circumstances. Well-known examples include emission-reduction targets or the mandatory installation and use of sustainable technologies.

In contrast, market-based instruments seek to address the market failure of negative environmental externalities by influencing firms' behavior through economic incentives. Such instruments help align financial incentives with environmental goals, thus encouraging innovation and enabling corporations to find the most efficient solutions (Baumol & Oates, 1988). Market-based instruments can be implemented either as quantity-based or as price-based instruments (Sterner et al., 2019; Weitzman, 1974).

A common quantity-based instrument is a cap-and-trade system that sets a maximum quantity, e.g., on emissions (Ivanov et al., 2024). This approach provides flexibility and cost efficiency by allowing firms to trade emission allowances, which incentivizes reductions, but can be complex to administer and may lead to uneven economic impacts if not properly regulated.

Price-based instruments are typically environmental taxes and environmental tax incentives. Environmental taxes impose a cost on activities that harm the environment, providing a financial incentive for firms to reduce their negative environmental impacts and adopt more sustainable practices. In general, these taxes appear in the form of Pigouvian taxes, which are designed to internalize the costs of negative environmental externalities (Jacob & Zerwer, 2024; Pigou, 2017). For example, carbon taxes impose a cost on each unit of carbon dioxide emitted, encouraging corporations to lower their carbon emissions and adopt cleaner energy solutions. In contrast, environmental tax incentives aim to encourage the adoption of environmentally friendly technologies or practices through profit- and expenditure-based incentives, such as tax holidays, reduced corporate income tax rates, and investment allowances. Tax incentives are typically embedded in broader corporate tax legislation. They can lead to market distortions, as corporations may rely on government support rather than making long-term investments in cleaner technologies or operations (Kremer & Willis, 2016). Moreover, tax incentives tend to be more volatile than environmental taxes, as they are often subject to changes in government policies or budget constraints.

Since we are interested in long-term oriented climate change mitigation, Pigouvian taxes provide a more stable and predictable mechanism for incentivizing firms' environmental responsibility and are generally considered more market-efficient than environmental tax incentives (Ballard & Medema, 1993), such that they are better suited for the purpose of our study. Specifically, different dimensions of corporate environmental taxation are aimed at reducing different types of negative environmental externalities (e.g., GHG emissions, air and water pollution, depletion of natural resources, vehicle use, and traffic congestion). Consequently, we focus on the four most used environmental taxes: energy taxes, pollution taxes, resource taxes, and transport taxes.

8.2. Appendix III-2. Corporate Environmental Tax Terminology Word List

air	ecological taxes	forest	kilometers	petrol	solar
atmosphere	efficiency	forestry	kilometre	petroleum	solid
automobile	efficient	fossil	kilometres	plastic	substances
automotive	electric	fossils	kilos	pollutant	sulfur
biodiesel	electricity	fuel	kilowatt	polluted	sulphur
car	emission	fuels	kilowatts	polluter	temperature
carbon	emissions	gallon	kwh	polluting	tetrachloride
carbon tax	emitter	gallons	landfill	pollution	timber
carbon taxation	emitting	gas	landfilling	pollution tax	timbering
carbon taxes	energy	gases	liter	pollution taxation	toxic
cars	energy tax	gasoline	liters	pollution taxes	transport
ccm	energy taxation	ghg	megawatt	railway	transport tax
chemical	energy taxes	gravel	megawatts	recycle	transport taxation
chemicals	environment	green	methane	recycling	transport taxes
chlorofluorocarbons	environmental	green tax	minerals	registration	transportation
chlorofluoromethanes	environmental tax	green taxation	mining	renew	transportation tax
clean	environmental taxation	green taxes	mobility	renewable	transportation taxation
cleaner	environmental taxes	greener	monoxide	renewables	transportation taxes
climate	environmentally	greenhouse gas	motor	resource	unit
co2	environmentally related tax	greenhouse gases	motorized	resource tax	units
contamination	environmentally related taxation	hazardous	natural resources	resource taxation	vehicle
cubic	environmentally related taxes	hunt	nitrogen	resource taxes	vehicles
deforestation	excavating	hunting	ocean	resources	waste
diesel	excavation	hybrid	oil	road	wastewater
dioxide	exploitation	hydrocarbon	oils	roads	water
eco-tax	exploiting	iron	ore	royalties	watt
eco-taxation	fertilizer	joule	ores	royalty	watts
eco-taxes	fish	joules	ozone	sand	
ecological tax	fisheries	kilo	packaging	soil	
ecological taxation	fishing	kilometer	pesticides	soiling	

Notes: This table shows the self-developed word list used for data collection, specifically manual screening, web scraping, and text mining.

8.3. Appendix III-3. Data Sources for Data Collection

ISO	Ministry of Finance	National Tax Authority	IBFD[a]	PwC[b]	EY[c]	Disc. [d]
AFG	https://mof.gov.af	https://ard.gov.af	x	n/a	n/a	x
AGO	https://minfin.gov.ao	https://agt.minfin.gov.ao	x	x	x	
ALB	https://financa.gov.al	https://tatime.gov.al	x	x	x	
ARE	https://mof.gov.ae	https://tax.gov.ae	x	x	x	
ARG	https://argentina.gob.ar	https://afip.gob.ar	x	x	x	
ARM	https://minfin.am	n/a	x	x	x	
ATG	https://ab.gov.ag	https://ird.gov.ag	x	n/a	n/a	
AUS	https://finance.gov.au	https://ato.gov.au	x	x	x	
AUT	https://bmf.gv.at	https://bmf.gv.at	x	x	x	
AZE	https://maliyye.gov.az	https://taxes.gov.az	x	x	x	x
BEL	https://finance.belgium.be	https://finance.belgium.be	x	x	x	x
BEN	https://finances.bj	https://impots.bj	x	n/a	n/a	
BFA	https://finances.gov.bf	https://impots.gov.bf	x	n/a	n/a	
BGD	http://www.mof.gov.bd	https://nbr.gov.bd	x	n/a	x	
BHR	https://mofne.gov.bh	https://nbr.gov.bh	x	x	x	
BHS	https://bahamas.gov.bs	https://inlandrevenue.finance.gov.bs	x	n/a	x	
BIH	http://www.fmf.gov.ba/	n/a	x	x	n/a	x
BLR	http://www.minfin.gov.by	https://nalog.gov.by	x	x	n/a	
BLZ	https://mof.gov.bz	https://bts.gov.bz	x	n/a	n/a	
BOL	https://economiayfinanzas.gob.bo	https://economiayfinanzas.gob.bo	x	x	x	x
BRB	n/a	https://bra.gov.bb	x	x	x	
BTN	https://mof.gov.bt	https://mof.gov.bt	x	n/a	n/a	
BWA	https://finance.gov.bw	https://burs.org.bw	x	x	n/a	x
CAF	n/a	https://impots.gouv.cf	x	n/a	n/a	x
CAN	https://canada.ca	https://canada.ca	x	x	x	x
CHE	https://efd.admin.ch	https://estv.admin.ch	x	x	x	x
CHN	http://www.mof.gov.cn	http://www.chinatax.gov.cn	x	x	x	
CIV	n/a	https://guce.gouv.ci	x	x	x	x
CMR	https://dgtcfm.cm	https://impots.cm	x	x	x	x
COG	https://finances.gouv.cg	https://finances.gouv.cg	x	x	x	
COL	https://irc.gov.co	https://dian.gov.co	x	x	x	x
CPV	https://mf.gov.cv	https://mf.gov.cv	x	x	x	
CRI	https://hacienda.go.cr	https://hacienda.go.cr	x	x	x	x
CYP	http://www.mof.gov.cy	https://mof.gov.cy	x	x	x	
CZE	https://mfcr.cz	http://www.financnisprava.cz	x	x	x	x
DEU	https://bundesfinanzministerium.de	https://bzst.de	x	x	x	
DJI	https://economie.gouv.dj	https://icaew.com	x	n/a	n/a	
DMA	https://finance.gov.dm	https://ird.gov.dm	x	n/a	n/a	
DNK	https://en.fm.dk	https://sktst.dk	x	x	x	x
DOM	n/a	n/a	x	x	x	
DZA	https://mf.gov.dz	https://mfdgi.gov.dz	x	x	x	x
ECU	https://finanzas.gob.ec	n/a	x	x	x	
EGY	https://mof.gov.eg	https://egypt.gov.eg	x	x	x	
ESP	https://hacienda.gob.es	https://hacienda.gob.es	x	x	x	
EST	https://fin.ee	https://fin.ee	x	x	x	x
ETH	https://mofed.gov.et	n/a	x	x	n/a	
FIN	https://um.fi	http://www.vero.fi	x	x	x	x
FJI	n/a	https://frcs.org.fj	x	x	x	x
FSM	https://dofa.gov.fm	https://dofa.gov.fm	x	n/a	n/a	
GAB	n/a	n/a	x	x	n/a	

ISO	Ministry of Finance	National Tax Authority	IBFD[a]	PwC[b]	EY[c]	Disc. [d]
GBR	https://gov.uk	https://gov.uk	x	x	x	x
GEO	https://mof.ge	https://matsne.gov.ge	x	x	x	x
GHA	https://mofep.gov.gh	https://gra.gov.gh	x	x	x	
GIN	https://devex.com	https://invest.gov.gn	x	n/a	x	
GMB	n/a	https://gra.gm	x	n/a	n/a	
GNB	https://mef.gw	https://kontaktu.mef.gw	n/a	n/a	n/a	
GNQ	https://minhacienda-gob.com/	n/a	x	x	n/a	
GRC	https://minfin.gr	https://gsis.gr	x	x	x	
GRD	https://finance.gd	https://ird.gd	x	n/a	n/a	
GTM	https://minfin.gob.gt	https://leyestributariasguatemala.com	x	x	x	
GUY	https://finance.gov.gy	https://gra.gov.gy	x	x	x	
HND	http://www.sefin.gob.hn	https://sar.gob.hn	x	x	x	x
HRV	https://mfin.gov.hr	https://porezna-uprava.hr	x	x	x	x
IDN	https://kemenkeu.go.id	https://pajak.go.id	x	x	x	
IND	n/a	https://incometaxindia.gov.in	x	x	x	
IRL	https://gov.ie	https://revenue.ie	x	x	x	
IRN	https://irangov.ir	n/a	x	n/a	n/a	x
IRQ	http://mof.gov.iq	http://mof.gov.iq	n/a	x	x	x
ISL	https://government.is	https://rsk.is	x	x	x	x
ITA	https://mef.gov.it	https://agenziaentrate.gov.it	x	x	x	
JAM	https://mof.gov.jm	https://jamaicatax.gov.jm	x	x	x	
JPN	https://mof.go.jp	https://nta.go.jp	x	x	x	
KAZ	https://gov.kz	https://kgd.gov.kz	x	x	x	
KEN	https://treasury.go.ke	https://kra.go.ke	x	x	x	x
KGZ	https://minfin.kg	n/a	x	x	n/a	x
KHM	http://www.mef.gov.kh	http://www.tax.gov.kh	x	x	x	
KNA	https://mof.gov.kn	https://sknird.com	x	n/a	n/a	
KOR	http://www.mof.go.kr	https://nts.go.kr	x	x	x	
KWT	https://mof.gov.kw	https://mof.gov.kw	x	x	x	
LAO	https://mof.gov.la	https://rd.go.th	x	x	x	
LBN	https://finance.gov.lb	n/a	x	x	x	
LCA	https://finance.gov.lc	n/a	x	x	x	
LKA	https://treasury.gov.lk	http://www.ird.gov.lk	x	x	x	x
LSO	http://www.finance.gov.ls	http://www.rsl.org.ls	x	n/a	x	
LTU	https://finmin.lrv.lt	https://vmi.lt	x	x	x	x
LUX	https://mfin.gouvernement.lu	https://impotsdirects.public.lu	x	x	x	
LVA	https://fm.gov.lv	https://vid.gov.lv	x	x	x	x
MAR	https://finances.gov.ma	https://finances.gov.ma	x	x	x	
MDA	https://mf.gov.md	https://iota-tax.org	x	x	x	x
MDV	https://finance.gov.mv	http://www.mira.gov.mv	x	x	x	x
MEX	https://gob.mx	n/a	x	x	x	
MKD	https://finance.gov.mk	http://www.ujp.gov.mk	x	x	x	
MLI	https://finances.ml	n/a	x	n/a	n/a	
MLT	n/a	https://cfr.gov.mt	x	x	x	x
MNG	https://mof.gov.mn	http://en.mta.mn	x	x	x	
MOZ	https://mef.gov.mz	https://at.gov.mz	x	x	x	
MRT	https://finances.gov.mr	n/a	x	x	x	
MUS	https://mof.govmu.org	https://mra.mu	x	x	x	
MWI	https://www.finance.gov.mw	https://www.mra.mw	x	x	x	
MYS	https://mof.gov.my	https://hasil.gov.my	x	x	x	
NAM	https://mfpe.gov.na	https://namra.org.na	x	x	x	x
NGA	https://finance.gov.ng	https://firs.gov.ng	x	x	x	x
NIC	http://www.hacienda.gob.ni	n/a	x	x	x	
NLD	https://government.nl	https://belastingdienst.nl	x	x	x	x

ISO	Ministry of Finance	National Tax Authority	IBFD[a]	PwC[b]	EY[c]	Disc.[d]
NOR	https://regjeringen.no	https://skatteetaten.no	x	x	x	
NPL	https://mof.gov.np	http://nepal.gov.np	x	n/a	n/a	
NZL	https://treasury.govt.nz	https://ird.govt.nz	x	x	x	x
OMN	https://mof.gov.om	https://tms.taxoman.gov.om	x	x	x	x
PAK	https://finance.gov.pk	https://fbr.gov.pk	x	x	x	
PAN	https://mef.gob.pa	n/a	x	x	x	x
PER	https://mef.gob.pe	https://ww3.sunat.gob.pe	x	x	x	x
PHL	https://dof.gov.ph	https://bir.gov.ph	x	x	x	
PNG	https://treasury.gov.pg	https://irc.gov.pg	x	x	x	x
POL	https://gov.pl	https://gov.pl	x	x	x	
PRT	https://portugal.gov.pt	https://info.portaldasfinancas.gov.pt	x	x	x	x
PRY	https://hacienda.gov.py	https://set.gov.py	x	x	x	
QAT	https://mofa.gov.qa	https://gta.gov.qa	x	x	x	
RUS	http://government.ru	https://nalog.gov.ru	x	x	n/a	
RWA	https://minecofin.gov.rw	https://rra.gov.rw	x	x	x	x
SAU	https://mof.gov.sa	https://zatca.gov.sa	x	x	x	
SGP	https://mof.gov.sg/	https://iras.gov.sg/	x	x	x	
SLB	https://solomons.gov.sb	http://www.ird.gov.sb	x	n/a	n/a	
SLE	https://mofsl.gov.sl	https://nra.gov.sl	x	n/a	n/a	
SLV	https://mh.gob.sv	https://transparenciafiscal.gob.sv	x	x	x	
SUR	https://gov.sr	https://dna.sr	x	n/a	x	
SVK	https://mfsr.sk	https://financnasprava.sk	x	x	x	x
SVN	https://gov.si	https://fu.gov.si	x	x	x	x
SWE	https://government.se	https://skatteverket.se	x	x	x	x
SWZ	https://gov.sz	https://ers.org.sz	x	x	x	x
SYC	http://www.finance.gov.sc	https://src.gov.sc	x	n/a	n/a	
TGO	https://finances.gouv.tg	https://numerique.gouv.tg	x	n/a	n/a	
THA	http://www2.mof.go.th	https://rd.go.th	x	x	x	
TJK	n/a	https://andoz.tj	x	x	n/a	x
TON	http://www.finance.gov.to	https://revenue.gov.to	x	n/a	n/a	
TTO	https://finance.gov.tt	https://ird.gov.tt	x	x	x	
TUR	https://en.hmb.gov.tr	https://mevzuat.gov.tr	x	x	x	
TZA	https://mof.go.tz	https://tra.go.tz	x	x	x	
UGA	https://finance.go.ug	https://ura.go.ug	x	x	x	
UKR	https://mof.gov.ua	https://tax.gov.ua	x	x	x	
URY	https://gub.uy	https://dgi.gub.uy	x	x	x	
USA	https://home.treasury.gov	https://irs.gov	x	x	x	
UZB	https://mf.uz	https://tax.uz	x	x	x	
VCT	http://www.finance.gov.vc	http://finance.gov.vc	x	n/a	n/a	
VNM	https://mof.gov.vn	https://mof.gov.vn	x	x	x	x
VUT	https://doft.gov.vu	https://customsinlandrevenue.gov.vu	x	n/a	n/a	
WSM	https://mof.gov.ws	https://revenue.gov.ws	x	n/a	n/a	
ZAF	https://treasury.gov.za	https://sars.gov.za	x	x	x	x
ZMB	https://mofnp.gov.zm	https://zra.org.zm	x	x	x	
ZWE	n/a	https://zimra.co.zw	x	x	x	x

Notes: This table shows the sources used to compile the CET-Index scores. [a] International Bureau of Fiscal Documentation (IBFD) Country Tax Guides; [b] Worldwide Tax Summaries Online, hosted by PricewaterhouseCoopers (PwC); [c] Worldwide Corporate Tax Guide, published by Ernst & Young (EY); [d] Supplementary professional discussions were only conducted in case of ambiguities; If “n/a”, no data available.

8.4. Appendix III-4. CET-Index

ISO	*ENV_TAX*	*ENV_TAX_4*	*ENE_TAX*	*POL_TAX*	*RES_TAX*	*TRA_TAX*	*ENV_TAX_INC*
AFG	1	1	0	0	1	0	1
AGO	1	3	1	0	1	1	1
ALB	1	2	1	0	0	1	1
ARE	0	0	0	0	0	0	0
ARG	1	2	1	1	0	0	1
ARM	1	4	1	1	1	1	1
ATG	0	0	0	0	0	0	0
AUS	1	1	1	0	0	0	1
AUT	1	4	1	1	1	1	1
AZE	1	2	0	0	1	1	1
BEL	1	3	1	1	0	1	1
BEN	1	2	1	0	0	1	1
BFA	1	3	1	1	0	1	1
BGD	0	0	0	0	0	0	0
BHR	0	0	0	0	0	0	0
BHS	0	0	0	0	0	0	0
BIH	1	1	1	0	0	0	1
BLR	1	3	1	1	0	1	1
BLZ	1	2	1	1	0	0	1
BOL	1	3	1	0	1	1	1
BRB	1	2	1	0	0	1	1
BTN	0	0	0	0	0	0	1
BWA	0	0	0	0	0	0	0
CAF	1	4	1	1	1	1	1
CAN	1	2	1	1	0	0	1
CHE	1	3	1	1	0	1	1
CHN	1	3	1	1	0	1	1
CIV	1	4	1	1	1	1	1
CMR	1	3	1	0	1	1	1
COG	1	2	0	0	1	1	1
COL	1	2	1	0	0	1	1
CPV	1	1	0	1	0	0	1
CRI	1	1	0	1	0	0	1
CYP	1	1	1	0	0	0	1
CZE	1	2	1	0	0	1	1
DEU	1	2	1	0	0	1	1
DJI	1	3	1	0	1	1	1
DMA	1	2	1	0	0	1	1
DNK	1	4	1	1	1	1	1
DOM	1	3	1	0	1	1	1
DZA	1	3	1	1	1	0	1
ECU	1	4	1	1	1	1	1
EGY	1	1	0	0	1	0	1
ESP	1	3	1	1	0	1	1
EST	1	4	1	1	1	1	1
ETH	1	1	0	0	0	1	1
FIN	1	3	1	1	0	1	1
FJI	1	1	0	0	1	0	1
FSM	1	1	1	0	0	0	1
GAB	1	1	0	0	1	0	1
GBR	1	4	1	1	1	1	1
GEO	0	0	0	0	0	0	1
GHA	1	1	1	0	0	0	1
GIN	1	1	0	0	1	0	1

ISO	*ENV_TAX*	*ENV_TAX_4*	*ENE_TAX*	*POL_TAX*	*RES_TAX*	*TRA_TAX*	*ENV_TAX_INC*
GMB	0	0	0	0	0	0	0
GNB	0	0	0	0	0	0	0
GNQ	1	1	0	1	0	0	1
GRC	1	3	1	1	0	1	1
GRD	0	0	0	0	0	0	0
GTM	1	1	0	0	1	0	1
GUY	1	2	0	1	0	1	1
HND	1	2	0	0	1	1	1
HRV	1	4	1	1	1	1	1
IDN	1	1	1	0	0	0	1
IND	0	0	0	0	0	0	1
IRL	1	3	1	1	0	1	1
IRN	0	0	0	0	0	0	1
IRQ	0	0	0	0	0	0	1
ISL	1	4	1	1	1	1	1
ITA	1	3	1	1	0	1	1
JAM	0	0	0	0	0	0	0
JPN	1	2	1	0	0	1	1
KAZ	1	4	1	1	1	1	1
KEN	0	0	0	0	0	0	1
KGZ	0	0	0	0	0	0	1
KHM	1	2	1	0	0	1	1
KNA	1	1	1	0	0	0	1
KOR	1	2	1	0	0	1	1
KWT	0	0	0	0	0	0	0
LAO	1	1	0	0	1	0	1
LBN	0	0	0	0	0	0	0
LCA	1	1	1	0	0	0	1
LKA	1	2	1	0	0	1	1
LSO	1	1	0	0	1	0	1
LTU	1	4	1	1	1	1	1
LUX	1	2	1	0	0	1	1
LVA	1	4	1	1	1	1	1
MAR	1	1	1	0	0	0	1
MDA	1	3	0	1	1	1	1
MDV	0	0	0	0	0	0	0
MEX	1	2	1	0	0	1	1
MKD	1	2	1	0	0	1	1
MLI	1	1	0	0	1	0	1
MLT	1	3	1	1	0	1	1
MNG	1	4	1	1	1	1	1
MOZ	1	1	0	0	1	0	1
MRT	1	1	1	0	0	0	1
MUS	1	2	1	0	0	1	1
MWI	1	1	1	0	0	0	1
MYS	1	2	0	0	1	1	1
NAM	1	2	1	0	0	1	1
NGA	0	0	0	0	0	0	0
NIC	1	1	1	0	0	0	1
NLD	1	4	1	1	1	1	1
NOR	1	4	1	1	1	1	1
NPL	1	1	0	0	0	1	1
NZL	1	2	1	0	0	1	1
OMN	0	0	0	0	0	0	0
PAK	1	2	1	0	0	1	1
PAN	0	0	0	0	0	0	1

ISO	*ENV_TAX*	*ENV_TAX_4*	*ENE_TAX*	*POL_TAX*	*RES_TAX*	*TRA_TAX*	*ENV_TAX_INC*
PER	1	1	0	0	1	0	1
PHL	1	2	1	0	1	0	1
PNG	0	0	0	0	0	0	0
POL	1	4	1	1	1	1	1
PRT	1	4	1	1	1	1	1
PRY	0	0	0	0	0	0	0
QAT	0	0	0	0	0	0	0
RUS	1	4	1	1	1	1	1
RWA	1	2	1	0	0	1	1
SAU	0	0	0	0	0	0	0
SGP	1	2	1	0	1	0	1
SLB	0	0	0	0	0	0	0
SLE	0	0	0	0	0	0	0
SLV	0	0	0	0	0	0	1
SUR	0	0	0	0	0	0	1
SVK	1	4	1	1	1	1	1
SVN	1	4	1	1	1	1	1
SWE	1	4	1	1	1	1	1
SWZ	0	0	0	0	0	0	0
SYC	1	3	1	1	0	1	1
TGO	1	2	1	0	1	0	1
THA	1	3	1	0	1	1	1
TJK	1	3	1	0	1	1	1
TON	1	1	0	0	0	1	1
TTO	1	1	0	0	0	1	1
TUR	1	1	0	0	1	0	1
TZA	1	4	1	1	1	1	1
UGA	1	3	1	1	1	0	1
UKR	1	1	0	1	0	0	1
URY	1	2	1	0	0	1	1
USA	1	2	1	1	0	0	1
UZB	1	2	1	0	1	0	1
VCT	1	2	1	0	0	1	1
VNM	1	2	0	1	1	0	1
VUT	1	1	1	0	0	0	1
WSM	1	2	1	0	0	1	1
ZAF	1	4	1	1	1	1	1
ZMB	1	3	1	0	1	1	1
ZWE	1	1	1	0	0	0	1

Notes: This table shows the scores for the presence of the corporate environmental tax dimensions as well as the aggregated scores for a unique global sample of 149 countries as of December 31, 2022. See Appendix III-1 for variable definitions.

8.5. Appendix III-5. Corporate Environmental Taxes by Geographic Regions

Regions	East Asia and Pacific	Europe and Central Asia	Latin America and Caribbean	Middle East and North Africa	North America	South Asia	Sub-Saharan Africa	EU	Non-EU
Variables					Mean				
ENV_TAX	0.909	0.951	0.714	0.308	1.000	0.500	0.800	1.000	0.746
ENV_TAX_4	1.727	2.927	1.321	0.615	2.000	0.750	1.657	3.304	1.548
ENV_TAX_INC	0.909	1.000	0.821	0.462	1.000	0.750	0.829	1.000	0.825
ENE_TAX	0.636	0.878	0.500	0.231	1.000	0.250	0.543	1.000	0.532
POL_TAX	0.182	0.683	0.179	0.077	1.000	0.000	0.257	0.826	0.238
RES_TAX	0.409	0.537	0.214	0.231	0.000	0.125	0.429	0.522	0.349
TRA_TAX	0.500	0.829	0.429	0.077	0.000	0.375	0.429	0.957	0.429
Observations	22	41	28	13	2	8	35	23	126

Notes: This table presents the mean of the dependent variables by geographical region as defined by the World Bank for analytical and operational purposes. Additionally, it includes mean values separated by EU and non-EU countries. See Appendix III-1 for variable definitions.

8.6. Appendix III-6. Sub-sample Analysis

Variables	*ENV_TAX*	
	Without EU member states	**Without OECD member states**
	Model (1)	Model (2)
CLI_EXP	0.237	0.192
	(1.104)	(0.889)
TAX_COM	-0.058**	-0.053**
	(-2.436)	(-2.216)
GOV_QUA	2.432***	2.276***
	(3.923)	(3.370)
FEM_PAR	0.091***	0.087**
	(2.606)	(2.411)
Constant	14.212***	13.130***
	(3.272)	(3.096)
Observations	126	115
Controls	YES	YES
Pseudo R^2	0.330	0.292
Wald χ^2	38.71***	34.45***
Hosmer-Lemeshow χ^2	9.637	10.56
Prob>Hosmer-Lemeshow χ^2	0.291	0.228
VIF mean (max)	2.06 (4.75)	1.97 (4.09)

Notes: This table presents the logistic regression results for *ENV_TAX* as the dependent variable with available independent and control variables. Model (1) provides the regression results excluding EU member states from our final sample. Model (2) provides the regression results excluding OECD member states from our final sample. Z values are reported in parentheses, based on robust standard errors. In all our models, the joint significance of the predictors in explaining the outcome is confirmed by the Wald χ^2 test statistic. In all our models, an adequate fit to the data is confirmed by the Hosmer-Lemeshow χ^2 test statistic. See Appendix III-1 for variable definitions. ***p<0.01, **p<0.05, *p<0.1.

8.7. Appendix III-7. Variable Definitions

Variables	Definition	Source
Dependent Variables		
ENV_TAX	is equal to "1" if the country has implemented at least one of the following four dimensions of corporate environmental taxes and "0" otherwise: energy taxes, pollution taxes, resource taxes, and transport taxes.	Self-constructed as of December 31, 2022
ENV_TAX_4	is the number of four dimensions of corporate environmental taxes present in a country.	Self-constructed as of December 31, 2022
ENV_TAX_INC	is equal to "1" if the country has implemented at least one of the four corporate environmental tax dimensions or at least one of the six corporate income tax-based environmental incentives (tax holidays, reduced tax rates, loss carryovers, investment allowances, tax credits, or accelerated depreciation) and "0" otherwise.	Self-constructed as of December 31, 2022
ENE_TAX	is equal to "1" if the country has implemented a corporate energy tax and "0" otherwise, including taxes on carbon emissions; fossil fuels, such as petrol and diesel; electricity consumption; and energy products.	Self-constructed as of December 31, 2022
POL_TAX	is equal to "1" if the country has implemented a corporate pollution tax and "0" otherwise, including taxes on wastewater discharge, SOx and NOx emissions, disposal of solid waste, and plastic packaging.	Self-constructed as of December 31, 2022
RES_TAX	is equal to "1" if the country has implemented a corporate resource tax and "0" otherwise, including taxes on water extraction, forest products, mining royalties, and hunting and fishing licenses.	Self-constructed as of December 31, 2022
TRA_TAX	is equal to "1" if the country has implemented a corporate transport tax and "0" otherwise, including taxes on vehicle miles travelled, vehicle registration or ownership, import taxes on transport equipment, and recurrent taxes on motor vehicle usage.	Self-constructed as of December 31, 2022
Independent Variables (H1)		
CLI_EXP	is the probability of physical exposure associated with specific climate-driven hazards.	IMF's Climate Change Indicators Dashboard (avg. 2017–2021)
CLI_VUL	is the economic, political, and social characteristics of the community that can be destabilized in case of a hazard event.	IMF's Climate Change Indicators Dashboard (avg. 2017–2021)
CLI_DIS	is the frequency of climate-related disasters.	IMF's Climate Change Indicators Dashboard (avg. 2017–2021)
CLI_TMP	is the mean surface temperature change measured with respect to a baseline temperature, corresponding to the period 1951–1980.	IMF's Climate Change Indicators Dashboard (avg. 2017–2021)
Independent Variables (H2)		
TAX_COM	is the inverse transformation of the taxes and mandatory contributions borne by the business in the second year of	Doing Business (avg. 2015–2019)

Variables	Definition	Source
	operation (percentage of commercial profit). For original values up to 100 percent, the inverse is defined as 100 minus the reported percentage. For values above 100 percent, the inverse is set to 0 to avoid negative values and maintain interpretability.	
TAX_COM_GDP	is the difference between the country's *TAX_COM* value and the average *TAX_COM* value for countries within the same GDP group (20 groups).	Doing Business (avg. 2015–2019)
TAX_COM_REG	is the difference between the country's *TAX_COM* value and the average *TAX_COM* value for countries in the same World Bank region (7 regions).	Doing Business (avg. 2015–2019)
TAX_INCOMPLEX	is the inverse transformation of the complexity of a country's corporate income tax system as faced by multinational corporations, comprising the two dimensions tax code complexity and tax framework complexity. The transformation is performed by subtracting each original value from the sum of the maximum and minimum values of the original variable. The resulting measure is then linearly rescaled to range from 0 to 100 using a min–max normalization.	Tax Complexity Index (avg. 2016, 2018, 2020)
Independent Variables (H3)		
GOV_QUA	is the aggregate score of the following six World Bank Worldwide Governance Indicators, ranging from -2.5 to +2.5: voice and accountability, political stability, governance effectiveness, regulatory quality, rule of law, and control of corruption.	World Bank (avg. 2017–2021)
GOV_CBL	is the score of a country's checks and balances on a country's executive's policymaking powers in terms of electoral competition and political majorities.	Database of Political Institutions (avg. 2017–2020)
GOV_DEM	is the extent to which political leaders are elected under comprehensive suffrage in free and fair elections and the extent to which freedoms of association and expression are guaranteed. The index ranges from 0 to 1 (most democratic).	Our World in Data (avg. 2017–2021)
GOV_FDM	is the state of freedom based on a series of 25 indicators categorized into groups, such as political rights and civil liberties, ranging from 0 to 100.	Freedom House (avg. 2017–2021)
Independent Variables (H4)		
FEM_PAR	is the percentage of parliamentary seats in a single or lower chamber held by women.	World Bank (avg. 2017–2021)
FEM_GOV	is the score of women's political rights, including the rights to vote, run for political office, hold elected and appointed government positions, join political parties, and petition government officials.	Quality of Government Standard Dataset (avg. 2017–2021)
FEM_PRT	is the extent to which women are represented in the legislature and have an equal share of political power.	Our World in Data (avg. 2017–2021)
FEM_MIN	is the percentage of ministerial-level positions held by women.	World Bank (avg. 2018–2020)
SKEWED	is equal to "1" if the percentage of parliamentary seats in a single or lower chamber held by women (*FEM_PAR*) is > 0 and ≤ 20 and "0" otherwise.	World Bank (avg. 2017–2021)
TILTED	is equal to "1" if the percentage of parliamentary seats in a single or lower chamber held by women (*FEM_PAR*) is > 20 and ≤ 40 and "0" otherwise.	World Bank (avg. 2017–2021)

Variables	Definition	Source
BALANCED	is equal to "1" if the percentage of parliamentary seats in a single or lower chamber held by women (*FEM_PAR*) is > 40 and ≤ 60 and "0" otherwise.	World Bank (avg. 2017–2021)
Control Variables		
COM_LAW	is equal to "1" if the country's legal system is based on common law and "0" if civil law-based.	La Porta et al. (2008)
GDP_PCA	is the natural logarithm of the country's per capita values for GDP expressed in current international dollars converted by purchasing power parity (PPP) conversion factor.	World Bank (avg. 2017–2021)
GHG_PCA	is the natural logarithm of the country's greenhouse gas emissions measured in tons per person of carbon dioxide-equivalents.	Our World in Data (avg. 2017–2021)
MAN_VAD	is the value added of the country's manufacturing industry as a percentage of GDP. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs.	World Bank (avg. 2017–2021)
PARTY_IDEOLOGY	is the economic-policy orientation of the executive party. Values range from 1 to 3, where 1 indicates right-leaning (conservative or Christian democratic) (*RIGHT*), 2 indicates centrist orientation (*CENTER*), and 3 indicates left-leaning (socialist or social democratic) (*LEFT*).	Database of Political Institutions (avg. 2017–2020)
POP_DEN	is the natural logarithm of the midyear population divided by land area in square kilometers.	World Bank (avg. 2017–2021)
PRESIDENTIAL	is equal to "1" if the country has a presidential political system, meaning the chief executive is directly elected or appointed with strong presidential powers (e.g., veto authority, ability to appoint/dismiss the prime minister, or dissolve parliament) and "0" otherwise.	Database of Political Institutions (avg. 2017–2020)
TAX_DTY	is equal to "1" if the country's constitution refers to the duty to pay taxes and "0" otherwise.	Quality of Government Standard Dataset (avg. 2017–2021)
TRADE	is the sum of exports and imports of goods and services. This indicator is expressed as a percentage of GDP which is the total income earned through the production of goods and services in an economic territory during an accounting period.	World Bank (avg. 2017–2021)

Notes: This table presents the variable definitions.

9. Statement on Funding

Financial support was received from the German Federal Ministry of Research, Technology and Space (BMFTR) under the reference number 01FP24036. Responsibility for the contents of this publication lies with the authors. No additional financial resources, grants, or assistance were received.

10. Statement on Declaration of Interest

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. This is an observational study. In accordance with the Ethics and Good Scientific Practice Guidelines of the Bremen International Graduate School of Social Sciences (BIGSSS), no ethical approval is required.

11. Statement on the Use of Artificial Intelligence (AI)

During the preparation of this work, the authors used *ChatGPT*, *DeepL Translate*, *DeepL Write*, and *Grammarly* to improve the language and readability, exercising caution. After using these tools, the author reviewed and edited the content as needed, taking full responsibility for the work's content.

12. Statement on Data Availability

The data presented in this study are available on request from the authors.

IV. Essay III

How to design and employ Specialized Large Language Models for Accounting and Tax Research: The Example of TaxBERT

Frank Hechtner, Lukas Schmidt, Andreas Seebeck, Marius Weiß

Abstract

Large Language Models (LLMs) are emerging as the new gold standard for accounting researchers utilizing Natural Language Processing (NLP) techniques to analyze the characteristics of financial and non-financial disclosures. This study provides an "A-to-Z" description of how to design and employ specialized Bidirectional Encoder Representations of Transformers (BERT) models that are suitable for practical implementation by accounting and tax researchers and are environmentally sustainable. We begin by highlighting some NLP-related key challenges, pitfalls, and shortcomings. Next, we provide a user's guide to LLMs and BERT models. Using the case-based approach of TaxBERT, a domain-specific LLM pretrained for analyzing qualitative corporate tax disclosures, we provide a detailed discussion on the efficient design, evaluation, and application of specialized BERT models tailored to specific domains. We show that TaxBERT leads to significant performance improvements over generic LLMs, FinBERT, and traditional bag-of-words approaches. By showcasing the development as well as the potential of domain-specific models, we aim to inspire researchers to develop and apply specialized LLMs in their research.

Acknowledgements

We are grateful to all researchers who provided valuable feedback throughout the process. In particular, we would like to thank Amir Amel-Zadeh, Kay Blaufus, Inga Hardeck, Robert Faff, Niels Johannesen, Reining Petacchi, Nicolas Webersinke, and Jennifer Wu Tucker for valuable comments. We also thank participants at the EAA Annual Congress (2024, 2025), Diginomics Summer Conference (2025), Annual MannheimTaxation Conference (2025), PSP-Doctoral Seminar (2025), Arqus-Annual Meeting (2025), AuBaNüDu-Doctoral Seminar (2025), and accounting workshops at the University of Erlangen-Nuremberg and Constructor University in Bremen. This paper received the Dr.-Michael-Munkert Research Award 2024. We sincerely appreciate the valuable comments and suggestions from the Editors and anonymous reviewers of "European Accounting Review," which will greatly contribute to improving this study.

1. Introduction

Over the past two decades, textual analysis has played an increasingly important role in accounting and tax research, enabling scholars to extract valuable insights from the vast and complex body of financial and non-financial disclosure texts. For instance, researchers have frequently employed textual analysis to detect topics (Dyer et al., 2017), quantify content (Lin et al., 2024), assess sentiment (Henry & Leone, 2016), evaluate readability (Rennekamp, 2012), measure specificity (Seebeck & Kaya, 2023), and identify similarities (Bozanic & Thevenot, 2015) and differences (Bochkay et al., 2023) of financial and non-financial disclosures.

Early studies predominantly relied on Natural Language Processing (NLP) techniques that operate under traditional dictionary-based approaches (Loughran & McDonald, 2011) and the bag of words framework (e.g., Huang et al., 2018; Lang & Stice-Lawrence, 2015; Li, 2010), where text is treated as an unordered collection of words. Many of these studies relied on the assumption that such approaches perform as well as methods that account for a document's internal structure (e.g., Friedman et al., 2001).[51] However, this assumption was only reasonable given the limitations of NLP technology at the time. Significant advancements in language models have since demonstrated that modern Large Language Models (LLMs) can substantially outperform earlier NLP algorithms across a range of tasks, including language translation, named entity recognition, and sentiment analysis (e.g., Azimi & Agrawal, 2021; Devlin et al., 2019; Siano & Wysocki, 2021).

Consequently, the focus of textual analysis in accounting and tax research is increasingly shifting toward advanced language models that use sophisticated machine learning techniques (e.g., Balakrishnan et al., 2023; Lai et al., 2017; Rajpurkar et al., 2016; Wang et al., 2019), in particular LLMs. The key advantage of LLMs is that they directly consider the context of a token by capturing syntactic and semantic relations among words, which has positioned them as the emerging gold standard for analyzing complex financial and non-financial disclosures.

In particular, the *Bidirectional Encoder Representations from Transformers* (*BERT*) model and its derivatives are popular among researchers (e.g., Baker et al., 2024; Cao et al.,

[51] While the bag-of-words approach can provide useful information about the frequency and presence of certain words, it does not consider the word context and the internal structure of documents (Huang et al., 2023).

2023; Huang et al., 2023; Siano & Wysocki, 2021). *BERT* models offer a bidirectional understanding of language that enables them to interpret text with great accuracy. By simultaneously considering both preceding and succeeding words in a sentence, *BERT's* Transformer architecture allows for a deeper comprehension of context and semantics (Devlin et al., 2019). This makes *BERT* particularly effective in handling complex narrative structures and extracting nuanced information across a wide range of applications, including financial and non-financial disclosures.

However, the application of generic LLMs may fail to fully capture the nuances of domain-specific contexts, as they are pretrained on general language (Beltagy et al., 2019; Huang et al., 2023). This limitation poses significant challenges for accounting and tax researchers and regulators, who frequently work with highly specialized, context-dependent texts. In financial and non-financial disclosures, terms and phrases often carry unique meanings that depend on the regulatory framework, financial context, or narrative structure of the disclosure. For instance, words like "impairment," "provision," or "adjustment" may have precise, technical interpretations that generic LLMs struggle to understand. For researchers, this can lead to inaccuracies in analyzing disclosures, ultimately compromising the validity and reliability of their findings.

Prior studies find that domain-adaptive pretraining can significantly improve the performance of LLMs by fine-tuning them on specialized corpora, enhancing their ability to understand and generate context-specific language. Examples include *Med-BERT* (Rasmy et al., 2021), *SciBERT* (Beltagy et al., 2019), *LegalBERT* (Chalkidis et al., 2020), *FinBERT* (Huang et al., 2023), and *ClimateBERT* (Webersinke et al., 2022).

While *FinBERT*, which is pretrained on financial texts, provides a significant improvement for the analysis of financial and non-financial disclosures, it still falls short in addressing the unique complexities of specialized accounting subdisciplines such as tax, auditing, or corporate governance. For example, tax-related language includes intricate legal terminology (e.g., "transfer pricing," "tax treaties," and "controlled foreign corporations"), jurisdiction-specific rules (e.g., FIN 48 in the U.S. and ATAD in the EU), and nuanced phrasing (e.g., "aggressive tax planning," "base erosion," and "profit shifting") that both generic *BERT* models and *FinBERT* may misinterpret. Similarly, auditing and corporate governance involve technical terms related to assurance, internal controls, and compliance that require a deep understanding of the domain. Addressing these challenges requires the development of domain-specific, specialized LLMs. As the narrative requirements for disclosures become increasingly multifaceted (Fiechter et al., 2022), with the continuous evolution of highly specialized fields such as biodiversity and human rights disclosures, there is an increase in the need for domain-specific LLMs.

This paper presents a comprehensive guide on how to design, evaluate, and employ domain-specific, specialized LLMs that are environmentally sustainable and practically feasible for accounting and tax researchers *analyzing* corporate disclosures. To enhance clarity and usability, we use a case-based approach in which each step is illustrated through the concrete example of *TaxBERT*, enabling readers to grasp the process better and apply these insights to their research endeavors.[52] The respective codes and architecture, including training code and model weights, are available at GitHub and Hugging Face repositories.[53]

In our guide, we adhere to practical limitations regarding computing power, energy efficiency,[54] time feasibility, and cost. Specifically, domain-specific *BERT* models that follow the approach presented in this study can be executed using commonly available computational resources, such as energy-efficient mid-range GPUs, rather than requiring access to expensive, high-energy-consuming, high-performance computing clusters. Moreover, the time feasibility of our approach ensures that the model training and fine-tuning do not take prohibitively long periods, which might delay research timelines or render the approach impractical for iterative studies.

Our contribution is twofold. First, we address the challenges of reproducibility and transparency in accounting and tax research that leverages NLP by confronting the black-box nature of LLMs (Hail et al., 2020). Commercial LLMs like *ChatGPT* are not open-source, which means their code and underlying embeddings are not accessible to the public. These models lack clear explanations of how specific outputs are derived from given inputs. By contrast, our guidelines and the case-based approach of this study can help researchers across

[52] *TaxBERT* is designed for the *analysis* of corporate tax disclosures rather than text generation. Its primary purpose is to extract and interpret information from textual data to support empirical research.

[53] The training code and model weights for all LLMs used in this study are available at the following GitHub repository: https://github.com/TaxBERT/TaxBERT and Hugging Face: https://huggingface.co/mariusweiss/TaxBERT.

[54] We acknowledge that energy efficiency is crucial in machine learning and large-scale computing, as it supports sustainability efforts by reducing the substantial electricity consumption and environmental impact of training and deploying large language models.

disciplines develop methodologically diverse and open *BERT*-based models with transparent architectures. To this end, we provide detailed documentation of model parameters and processing steps. This enables other researchers to replicate models and studies and to validate their results.

Second, we introduce *TaxBERT*, a domain-specific language model for corporate tax research that allows for transparency and reproducibility of research findings in future tax studies. We provide an environmentally sustainable, cost-effective, and accessible solution for researchers and practitioners working with large volumes of tax-related texts. This approach democratizes the use of advanced NLP tools in the tax field.

In Section 2, we begin by highlighting some NLP-related key challenges, pitfalls, and shortcomings. In Section 3, we provide a user's guide to LLMs and *BERT* models for accounting and tax researchers. Based on the case-based approach of *TaxBERT*, we then discuss how to efficiently design, evaluate, and employ specialized *BERT* models for various use cases in the accounting and tax domain in Section 4. Section 5 concludes the essay.

2. NLP-related Challenges, Pitfalls, and Shortcomings

The application of NLP in accounting and tax research has grown significantly (e.g., Bochkay et al., 2023; Li, 2010; Loughran & McDonald, 2016), offering researchers various tools to analyze complex financial and non-financial disclosures as well as narrative texts. However, despite the advancements brought by LLMs, significant challenges, pitfalls, and limitations persist in this field, with some even stemming directly from their application.

2.1. Challenges

First, as discussed above, accounting and tax texts often contain highly technical and context-dependent terminology that generic NLP models struggle to interpret accurately (Huang et al., 2023). Variations across jurisdictions and industries further complicate the analysis. Relying on generic language models (e.g., *GPT* or *BERT*) without domain-specific fine-tuning may lead to potentially less accurate results.

Second, financial and non-financial disclosures lack consistent formatting and structure, making it challenging to preprocess and analyze such texts effectively (Bochkay et al., 2023). Prior research has studied disclosures and other accounting documents published in various formats, including PDF (e.g., Hardeck et al., 2024; Nguyen et al., 2024), HTML (e.g., Dyer et al., 2017), XBRL (e.g., Blankespoor, 2019; Hoitash & Hoitash, 2018), or plain text (e.g., Loughran & McDonald, 2011), requiring different extraction techniques. Moreover, they often contain embedded tables and multi-column layouts that make extracting text in the correct order difficult, and may break the flow of text, complicating sequential text analysis.

Third, financial and non-financial disclosures often include lengthy, complex sentences with multiple clauses, which are difficult for NLP models to parse and interpret correctly. For instance, long sentences in firms' reports often contain embedded clauses, conjunctions, and prepositional phrases that create intricate syntactic structures. NLP models must untangle these structures to interpret the sentence accurately, which can be computationally challenging and prone to errors.

Fourth, sentences in firms' disclosures often derive meaning from broader contexts (e.g., previous disclosures or related reports), which many NLP models fail to capture. For instance, accounting texts often reference related sections, notes, or reports for detailed explanations.[55] Generic NLP models, which usually process individual sentences or paragraphs in isolation, struggle to incorporate this referenced information into their analysis.

Fifth, high-quality, labeled datasets are crucial for machine learning (Sheng et al., 2008). Annotating such financial and non-financial disclosures requires expertise in both NLP and accounting, making the process costly and time-consuming. Without sufficient and good-quality labeled data, NLP models fine-tuned for accounting tasks may fail to generalize effectively to real-world applications, resulting in suboptimal performance – a concept known as "garbage in – garbage out" (Geiger et al., 2021). On the other hand, limited labeled datasets increase the risk of overfitting and bias, as models may learn patterns specific to the training data but not reflective of broader accounting contexts. This may lead to unreliable or skewed results in accounting research.

Sixth, the vast number of parameters makes LLMs resource-intensive to train and deploy, requiring substantial computational power, storage, energy, extensive amounts of data, and time (Bender et al., 2021). Thus, researchers should consider using traditional NLP approaches when the research context involves straightforward text analysis or when only small datasets are available (Schwartz et al., 2019). Nevertheless, LLMs are generally superior in performance and flexibility, and should be the default consideration. Researchers are encouraged to evaluate their applicability first before opting for less sophisticated approaches.

[55] A common sentence in 10-K filings is: "For more information on '*tax topic X*,' refer to Note #."

Seventh, reduced reproducibility of the findings has become a major problem in accounting research (Hail et al., 2020). The black-box nature of commercial large language models like *ChatGPT* poses significant challenges for reproducibility and transparency. These models often make it difficult for researchers to understand how specific outputs are derived from given inputs. This opacity hinders the ability to replicate studies and validate results, as other researchers cannot easily assess the influence of model parameters or processing steps (Castelvecchi, 2016). Consequently, relying on such opaque models may undermine the scientific rigor and credibility essential in our field.

2.2. Pitfalls

A common pitfall is that NLP approaches fail to integrate numerical and visual information alongside textual narratives effectively. Controlling for financial performance in multivariate regressions helps to isolate the effect of textual narratives on the dependent variable and to prevent the risk of spurious correlation. However, it does not fully address the deeper integration challenges of multimodal data in NLP approaches. To fully address this issue, more sophisticated approaches are needed. These include multimodal models that combine numerical data, visuals, and textual embeddings in machine learning models to analyze how they interact. A recent working paper by Dinh et al. (2025) presents an initial attempt at such an approach, using machine learning techniques to match content from visuals with narratives in sustainability reports. However, to the best of our knowledge, there is no widely adopted, standardized approach specifically designed for numerical-aware word embeddings available that would allow researchers to incorporate numerical information into text representation.

Moreover, NLP tools used in accounting research often lack standardized evaluation metrics or benchmarks, making it difficult to assess and compare their performance across studies. Additionally, similar NLP measures are often defined differently across studies, hindering meaningful comparisons and consistency. For instance, one study might define sentiment analysis as distinguishing between positive and negative disclosures (e.g., Tetlock, 2007), while another might include neutral or risky statements as separate categories (e.g., Kravet & Muslu, 2013). These inconsistencies make direct performance comparisons between studies problematic.[56]

[56] In the next section of this study, we present some common evaluation metrics for LLMs.

Finally, a common pitfall in accounting and tax studies is reduced construct validity, which refers to the challenge of ensuring that NLP-based measures accurately reflect the theoretical concept or construct that the researcher aims to study. For instance, when analyzing linguistic sentiment in analyst questions during earnings calls, sentiment can represent multiple overlapping constructs, and it is often unclear which one is being measured. On the one hand, analysts might use positive or negative sentiment to signal their expectations about the firm's future performance. On the other hand, it could simply reflect analysts' mood or overall outlook, not necessarily their expectations about the firm. Finally, the sentiment may also reflect how challenging or critical the questions are, indicating a more confrontational or in-depth approach rather than optimism or pessimism. Often, these constructs are intertwined, making it difficult to isolate one from another. If the measure (e.g., sentiment score) does not clearly distinguish between these constructs or account for their interactions, this leads to measurement errors and potentially incorrect results and conclusions. To address this pitfall, researchers must carefully define their constructs and consider using more granular measures or models that can disentangle overlapping constructs, employing qualitative validation to ensure the measure aligns with the intended construct, and supplementing linguistic sentiment analysis with additional proxies or methodologies to isolate the desired construct more effectively.

2.3. Shortcomings

While NLP offers promising solutions to study important research questions, neither traditional approaches nor LLMs are a panacea (Loughran & McDonald, 2016). One of the major shortcomings of accounting and tax studies that use LLMs is their lack of reproducibility and transparency. LLMs often function as "black boxes," as their complexity makes it difficult to discern precisely how they generate outputs or the significance of specific inputs, especially when access to model specifications and architecture is limited or unavailable. (Castelvecchi, 2016; Lauriola et al., 2022). Thus, this makes it difficult for researchers to validate results or explain how specific predictions are made.

Moreover, many studies that develop customized approaches and models or use proprietary datasets do not share or standardize their datasets. This lack of transparency makes it harder to replicate results or validate findings against other methodologies. We encourage researchers to share datasets, code, and model architectures to enable replication and comparison across studies.

Finally, ethical and regulatory challenges can arise when deploying LLMs in sensitive contexts, where outputs might influence decisions with substantial financial implications. Accounting research must proactively address issues such as data privacy, accountability for predictions, and the unintended consequences of automated analysis to ensure the responsible use of LLMs. For example, biases related to gender, race, or country could unfairly influence results, making it important to detect and reduce such biases while ensuring transparency in how models are trained and used (Gupta et al., 2022).

3. User's Guide to LLMs and BERT Models

In the following case-based "A-to-Z" guide, we offer practical solutions for designing, evaluating, and employing domain-specific LLMs to advance accounting and tax research while maintaining efficiency, validity, reliability, and transparency of the research approach. We use *TaxBERT* as an example, which is a domain-specific language model for qualitative corporate tax disclosure. We start by providing some technical background on LLMs and *BERT* models, including the existing financial and non-financial disclosure domain-specific *BERT* models *FinBERT* and *ClimateBERT*.

3.1. What are Large Language Models (LLMs)?

LLMs are a subset of deep learning-based NLP models but are distinguished by their large scale, both in terms of the number of parameters (often billions) and the volume of data used for training. The most well-known examples include *ELMo*, *OpenAI GPT*, and *BERT* (Devlin et al., 2019; Peters et al., 2018; Radford et al., 2018). Due to their massive number of parameters, they are capable of learning complex relationships between words by analyzing large amounts of text. Consequently, they can capture the context and meaning more effectively than traditional methods, and thus, should be considered as the preferred approach in accounting studies that analyze complex linguistic patterns within financial and non-financial reports.

LLMs achieve their performance through pretraining objectives. These tasks help language models develop a deep understanding of syntax, semantics, and discourse structure. Many LLMs, such as *BERT* (Devlin et al., 2019), *RoBERTa* (Liu et al., 2019), and *ALBERT* (Lan et al., 2019), are trained using Masked Language Modeling (MLM), where words are randomly masked, and the model is tasked with predicting them based on context. For example, in the sentence: "The auditor issued a [MASK] opinion." the model learns to predict the missing word ("qualified," "unqualified," etc.) by understanding the context provided by the surrounding words.

Additionally, these models are pretrained by using tasks such as Next Sentence Prediction (NSP) (Devlin et al., 2019), Sentence Order Prediction (SOP) (Liu et al., 2019), or Span Corruption (Lan et al., 2019). NSP is a task where the model is given two sentences and asked to predict whether the second sentence follows the first in the original text. For example: Sentence A: "The company issued a restatement." Sentence B: "Its stock price significantly dropped immediately." The model learns to classify whether Sentence B actually follows Sentence A or is randomly sampled from the corpus. The goal is to help the model understand sentence-level relationships. Alternatively, SOP can be used to capture inter-sentence coherence. In this task, the model is given two consecutive sentences from the same document. It must predict whether the order of the two sentences is correct or swapped. This tests the model's ability to understand sentence-to-sentence flow and logical progression. Finally, in Span Corruption tasks, the model is given an input where contiguous spans of tokens are masked or corrupted, and the model is trained to reconstruct the original text. This differs from MLM, which typically masks individual random tokens. Since spans often include phrases or multiple words, the model must understand longer dependencies and contextual relationships.

Another foundational pretraining objective, primarily used in decoder-only models like *GPTs,* is Causal Language Modeling (CLM), also known as autoregressive language modeling (Radford et al., 2018). It trains a model to predict the next token in a sequence, using only the previous tokens as context. That is, the model generates text from left to right, one token at a time. This one-directional training objective enables the model to learn contextual dependencies, word order, and patterns in natural language. While CLM only sees the tokens that come before the current position, it aligns well with generative tasks such as text completion, summarization, and dialogue generation.

3.2. Why should accounting researchers consider using LLMs?

Accounting researchers should consider using LLMs because they open up new analytical possibilities and address several long-standing limitations in textual and qualitative research. The following example shows why LLMs are superior, for instance, in sentiment analysis. Consider the following phrase in a financial report: "While the company's revenue declined slightly, profitability improved significantly due to cost-saving measures." Although there is a slight decline in revenue, the overall sentiment of the sentence is positive because profitability improved significantly.

A bag-of-words model would treat the text as an unordered collection of words, focusing on the presence or frequency of words like "declined" and "improved." "Declined" might be interpreted as negative, while "improved" might be seen as positive, potentially leading to an overall mixed or unclear sentiment, as context is ignored.

By contrast, LLMs would consider the relationships between words and their context within the sentence. For instance, *BERT* would process the sentence using multiple layers of self-attention, which allows it to understand which words are most relevant to each other (e.g., linking "profitability" with "improved significantly." Also, it captures nuanced meaning, such as contrast signaled by "while," which indicates a shift from a negative to a more important positive point. Finally, it produces contextual embeddings for each word. For example, the word "declined" has a different meaning depending on whether it is followed by a mitigating phrase like "slightly" or emphasized by stronger terms. This enables LLMs like *BERT* to more correctly interpret the sentiment, which is especially important in accounting contexts, where subtle phrasing can drastically change meaning.

3.3. Why are BERT models particularly suitable for Accounting Research?

The *BERT* model (*Bidirectional Encoder Representations from Transformers*) is a subtype of LLMs that was first introduced by Google AI in 2018 (Devlin et al., 2019). It is pretrained using a large corpus of general-domain text, such as Wikipedia and BooksCorpus. *BERT* and its successor *RoBERTa* (Liu et al., 2019) represent a significant advancement in NLP mainly due to three reasons. First, *BERT* models consider the context of words in a sentence bidirectionally, meaning it considers both the words before and after a given word, which improves its ability to capture nuanced meanings and relationships in text. This contrasts with other LLMs like *GPT* models, which are unidirectional (processing text left-to-right). The unidirectional nature of *GPT* models can limit the ability to understand the context fully, making predictions harder to interpret. While *GPT* models excel at text generation, *BERT's* bidirectional nature makes it superior for text analysis, an area of great importance in accounting research.

Second, *BERT* employs a self-attention mechanism, which assigns different weights to words in a sentence based on their relevance to one another. This mechanism enables the model to focus on critical words or phrases while downplaying less relevant ones. For example, in the sentence: "Although profits decreased, the firm's future outlook remains strong." *BERT's* self-attention helps the model recognize that "future outlook" and "strong" are more important for determining the overall sentiment than "decreased."

Third, *BERT* is open source and designed to be easily fine-tuned on domain-specific data. This fine-tuning process is straightforward and involves adjusting the model to specialized datasets through additional training. This stepwise adaptation allows researchers to control and observe how the model integrates new domain-specific information, providing an additional layer of interpretability. Thereby, it also addresses some aspects of the "black box" shortcoming associated with deep learning models. Fine-tuning *BERT* models is important because prior research shows that general *BERT* models show inferior performance when applied to niche topics like clinical discharge summaries (Zhu et al., 2018), clinical notes (Si et al., 2019), or biomedicine (Lee et al., 2019).

3.4. Existing domain-specific BERT Models particularly suitable for the Analysis of financial and non-financial Disclosures

Prior research has developed domain-specific LLMs based on the *BERT* architecture. Specifically, they used domain-adaptive pretraining to improve the performance of language models by fine-tuning them on specialized corpora, enhancing their ability to understand context-specific language. Two prominent examples in the broader accounting domain are *ClimateBERT* (Webersinke et al., 2022) and *FinBERT* (Huang et al., 2023).

ClimateBERT (Webersinke et al., 2022) is frequently employed in sustainability and climate-related accounting research. It is pretrained on climate-focused texts such as sustainability reports, climate policy documents, scientific articles, and news related to environmental issues. Further, it is fine-tuned for tasks like sentiment analysis, climate risk identification, and topic classification in sustainability contexts (Bingler et al., 2022). *ClimateBERT* captures the nuanced language and terminology specific to environmental and climate-related discussions, including terms like "carbon offset" or "net-zero emissions." As a result, it outperforms generic models in tasks involving climate disclosures, sustainability reporting, and environmental impact analyses (Webersinke et al., 2022).

FinBERT (Huang et al., 2023),[57] which is frequently used in accounting research (e.g., Baker et al., 2024), is pretrained on financial documents such as 10-K and 10-Q reports, analyst reports, earnings call transcripts, and financial news articles. It is fine-tuned for tasks like sentiment analysis, financial risk detection, and topic classification. It captures finance-specific language and context, such as terms like "impairment" or "leveraged buyout." Consequently, it performs better than generic models on tasks involving financial disclosures and market sentiment (Huang et al., 2023). While generally effective for analyzing financial and non-financial disclosures, *FinBERT* is not specialized for subdomains. For instance, our results presented in Section 4 indicate that *FinBERT* does not sufficiently address the highly specialized tax disclosure language.

3.5. Developing a domain-specific BERT Model

In this section, we present a comprehensive guide for creating specialized *BERT* models tailored for use by accounting and tax researchers. Our approach prioritizes not only achieving exceptional model accuracy but also ensuring high resource efficiency, effective practical deployment, an intuitive user experience, and scalability for diverse applications.

3.5.1. Step 1: Understanding the domain-specific Corpora

In this study, we use the case-based approach of *TaxBERT*, a tool specialized in the analysis of corporate qualitative tax disclosures, to explain the steps in developing domain-specific LLMs.

Before developing or applying a specialized LLM, researchers should first evaluate whether such a model is truly necessary for their specific research objectives. Key considerations include the complexity of the task, the availability of domain-specific data, the potential for significant performance improvement over general models, and the trade-offs in terms of development time, computational resources, and cost. Therefore, it is important to understand the domain-specific characteristics of the corpora.

Looking at the example of *TaxBERT*: The past decade has seen a significant increase in the volume and complexity of private and public qualitative tax disclosures (Hoopes et al., 2024; Müller et al., 2020), in both mandatory (e.g., UK Tax Strategy) and voluntary (e.g., Australian Tax Transparency Code, GRI 207 tax sustainability) disclosure formats, underlining

[57] We note that another *FinBERT* model with domain-adaptive pretraining exists (Araci, 2019).

the need for using sophisticated tools for profound textual analysis. Using general LLMs in tax research, however, remains challenging for several reasons. First, tax-specific disclosures are characterized by the complexity of legal jargon, which stems from a volatile regulatory environment and the need to address different tax authorities. Consequently, general LLMs may lack the depth required for specialized tasks in the tax domain. Further, the intricacy of corporate tax planning strategies, often involving multiple tax jurisdictions, complicates the tax-related disclosures and their automated analysis. Moreover, the lack of standardized terminology and the use of diverse reporting formats across different firms and jurisdictions create additional obstacles for consistent and accurate text interpretation. Overall, a specialized LLM for qualitative tax disclosure is essential to address these challenges by providing tailored linguistic capabilities. To the best of our knowledge, there are no such LLMs for tax disclosure yet.

After defining the need for a specialized model, researchers should also consider data availability and accessibility. While it is true that *BERT* models perform well even with smaller training datasets (Huang et al., 2023), we recommend using as much training data as possible.

Given the limited availability of private tax disclosure data, *TaxBERT* is designed using only publicly available corporate tax disclosures.[58] There is a wide variety of mandatory and voluntary public qualitative corporate tax disclosures. For example, firms have to provide relevant tax information within footnotes (Luo et al., 2024) and the Management Discussion and Analysis (MD&A) sections (Cho & Muslu, 2021; Durnev & Mangen, 2020). Well-known examples of voluntary tax disclosures include GRI 207 from the Global Reporting Initiative and the Tax Transparency Code (TTC) from the Australian Taxation Office. Moreover, public tax reporting includes third-party tax disclosures, for instance the mandatory public tax return disclosures and tax-related SEC comment letters (Acito & Nessa, 2022; Hoopes et al., 2018; Kubick et al., 2016), and voluntary third-party disclosures, such as analysts' earnings forecasts, media and NGO reports, as well as whistleblower revelations (Mauler, 2019; Nesbitt et al., 2023; Wilde, 2017). Additionally, expanded auditor reporting requirements mandate that auditors have to disclose tax-related information in their audit reports when tax matters are identified as Key Audit Matters (KAMs). Consequently, the massive amount of these various

[58] We note that the *TaxBERT* model developed in this study does not consider private corporate tax disclosure corpora, as these are generally available only to tax authorities and selected recipients.

qualitative tax disclosures provides a fruitful ground for training a specialized LLM. Next, we have to decide what data to use for the pretraining of the domain-specific model.

3.5.2. Step 2: Defining the Input Data

Pretraining *BERT* models on a domain-specific corpus requires several steps. The first step is to gather a diverse set of domain-specific documents that encompass the broad range of language and contexts used in the research field. Researchers should ensure that the documents are heterogeneous in order to effectively capture the most significant aspects of the domain.

We include the following data sources in the tax corpus for *TaxBERT*, covering a wide range of regions and jurisdictions: tax-related sections of 10-K filings, GRI 207 reports, UK Tax Strategy reports, Australian TTCs, Tax KAMs, and corporate tax news. We thereby ensure that our corpus captures all relevant public tax disclosure formats: mandatory and voluntary, by firm and by third party. We exclude tax policy and regulatory documents from the corpus as these documents lack the strategic or risk management perspectives found in corporate disclosures. We also do not include social media data, as we assume that these texts are too noisy.[59] Appendix IV-1 describes the specific composition of our tax corpus.

Table IV-1 reports a large heterogeneity between our data sources in terms of text length, readability, and sentiment. Australian TTC disclosures, Tax KAMs, GRI 207 reports, as well as UK Tax Strategy documents, tend to be longer and are easier to read than tax disclosures in 10-K filings. Overall, our corpus encompasses 28,180 tax-related documents with approximately 20 million tokens.

Table IV-1. Tax Corpora

Source	Documents	Sentences	Tokens	Mean Readability (Gunning FOG)	Mean Polarity (Textblob)
10-K (XBRL)	21,643	653,287	15,101,685	16.298	0.050
Tax News	1,176	40,082	793,745	13.179	0.068
Australian TTC	657	92,881	1,673,115	14.383	0.080
GRI 207	643	48,149	717,153	15.632	0.080
UK Tax Strategy	476	20,667	492,550	15.830	0.122
Tax KAMs	3,585	39,262	974,157	18.328	0.091
Tax Corpus	**28,180**	**894,328**	**19,752,405**	**16.360**	**0.060**

Notes: This table presents the sample selection process for the different tax corpora used to pretrain *TaxBERT*.

[59] Tan et al. (2025) find that *TaxBERT* shows significant performance improvements over general language models on legal tax text analysis tasks, suggesting that it can also be useful for the analysis of other tax-related documents.

Following Gururangan et al. (2020), we also analyze the similarity between the different data sources. Table IV-2 reports that the vocabulary overlap between the different corpora ranges from 35.86% between UK Tax Strategy and Tax KAMs to 50.54% between GRI 207 and Australian TTC. This variety in the vocabulary underscores the importance of considering a broad range of data sources for pretraining.

Table IV-2. Vocabulary Overlaps between Domains

		1	2	3	4	5	6
1	10-K (XBRL)	1.000					
2	Tax News	0.451	1.000				
3	Australian TTC	0.422	0.451	1.000			
4	GRI 207	0.410	0.429	0.505	1.000		
5	UK Tax Strategy	0.382	0.458	0.459	0.431	1.000	
6	Tax KAMs	0.392	0.376	0.399	0.418	0.359	1.000

Notes: This table presents the overlap of the domain-specific vocabulary in the different tax corpora used to pretrain TaxBERT (%).

3.5.3. Step 3: Choice of an optimal base Model

The third step in pretraining a *BERT* model on a domain-specific corpus is the choice of an optimal base model that was pretrained on a general domain. Appendix IV-4 shows an overview of the most common *BERT* models that are typically considered for such tasks.

For many researchers, especially those in academic or smaller institutional settings, computational resources are limited and expensive to acquire. Thus, in line with Webersinke et al. (2022), we recommend using *DistilRoBERTa* as the starting point for training. It is a distilled version of *RoBERTa*, which is an optimized version of the *BERT* model that enhances performance across various NLP tasks.[60] *DistilRoBERTa* inherits these optimizations while offering a more lightweight architecture through the approach of *DistilBERT*, with roughly 60% of *RoBERTa's* parameters. While it uses only 6 transformer layers, 768 hidden units, and 12 attention heads, resulting in approximately 82 million parameters,[61] it is found to retain about 95% of the performance of the *BERT* model in terms of language understanding

[60] *RoBERTa* was trained on 160 GB of diverse English-language corpora, including data from BOOKCORPUS, WIKIPEDIA, CCNEWS dataset, the OPENWEBTEXT corpus, and a subset of CommonCrawl designed to mimic the narrative style of WINOGRAD schemas.

[61] In comparison, the *BERT* model has 12 transformer layers, 768 hidden units, and 12 attention heads, resulting in approximately 110 million parameters.

capabilities, while reducing the model size by 40% and calculating 60% faster (Sanh et al., 2020).[62]

We recommend accounting researchers using *DistilRoBERTa* as the foundation for domain-specific *BERT* models, as this approach allows them to fine-tune the model effectively, while maintaining the benefits of reduced computational requirements, faster processing, lower costs, and environmental sustainability.[63]

3.5.4. Step 4: Domain-adaptive Pretraining

In the next step, the domain-adaptive pretraining conducted on the pretrained general LLM takes place to tailor the model's understanding and performance to the specific domain (Gururangan et al., 2020). This approach enhances the model's ability to process text that uses specialized vocabulary, unique contexts, and intricate relationships specific to the domain (Howard & Ruder, 2018). Key components of this process include vocabulary augmentation, careful specification of training parameters, and thorough performance evaluation.[64]

Vocabulary Augmentation

Vocabulary augmentation is a technique that involves modifying or expanding the original vocabulary of the pretrained language model to include terms, phrases, and subwords that are highly relevant to the target domain. This helps the model handle specialized terminology better and improves its performance on domain-specific tasks (Nayak et al., 2020). Focusing solely on adding terms simplifies vocabulary augmentation by balancing efficiency and effectiveness, making it an ideal strategy for many domain-adaptive pretraining tasks. It improves model relevance without unnecessarily increasing complexity by including critical, high-frequency domain-specific words. The aim of this is to prevent overloading the vocabulary with subwords and phrases that may have marginal utility, while simultaneously optimizing the use of computational resources during training and inference.

[62] *RoBERTa_Large*, with 355 million parameters, could also be used, but it requires approximately five to six times more computing time. The performance gain is likely to be modest. Moreover, while *DistilRoBERTa* needs only about 6–8 GB of VRAM, *RoBERTa_Large* requires at least 24 GB, making it significantly more resource-intensive.

[63] While we acknowledge that using more advanced base models may potentially offer higher performance, our approach strikes an optimal balance between performance, efficiency, time, and cost, making it the most practical solution for the specific needs of accounting research.

[64] We train all models using the Hugging Face Transformers library on a single energy-efficient Nvidia GeForce RTX 4090 (24 GB of GPU memory), accompanied by an Intel Core i9-14900K and 64 GB of RAM.

In the process of pretraining *TaxBERT*, we follow Webersinke et al. (2022) by augmenting the vocabulary of *DistilRoBERTa* through the addition of the 235 most common tokens in our tax corpus to the tokenizer, leading to a vocabulary increase from 50,265 to 50,500.[65] We manually reviewed these added tokens to examine whether they meaningfully extend the vocabulary. All of the 235 tokens were not present in *DistilRoBERTa's* vocabulary, including the words "avoidance," "evasion," and "taxpayer."[66] Hence, this approach adds a substantial number of tax domain-specific terms not frequently used in general texts to the model's vocabulary.

The addition of new tokens to the model's vocabulary results in a linear increase in the number of parameters.[67] On the one hand, this improves the model's ability to represent domain-specific language, ensuring better performance on tasks involving specialized terminology. On the other hand, the increase in parameters can demand more memory and computational resources during training and inference and increase the risk of overfitting, especially if only limited domain-specific data is available for training.

Hyperparameters

Hyperparameters define the configuration and behavior of the model during training. Thus, they influence how effectively the model learns from the domain-specific corpus. Proper tuning of these hyperparameters is critical for achieving optimal results. In the following, we explain the most important hyperparameters in more detail. The hyperparameters used for *TaxBERT*, as presented in Table IV-3, closely resemble prior research on domain-adaptive pretraining (e.g., Alsentzer et al., 2019; Araci, 2019; Beltagy et al., 2019; Lee et al., 2019; Webersinke et al., 2022).

Table IV-3. Hyperparameter Choice

Hyperparameter	Value
Number of Epochs	12
Batch Size	512
Learning Rate	$5.00e^{-5}$
Learning Rate Scheduler	Warmup linear
Hidden Dropout Probability	0.3

[65] Theoretically, adding phrases such as "transfer pricing" ensures the model directly recognizes these important concepts as single entities. In additional model specifications for *TaxBERT*, we therefore add phrases and subwords in addition to terms. Moreover, we replace unused or less frequent tokens with domain-specific terms rather than adding them. Neither approach results in significant performance improvements.

[66] See Appendix IV-3 for an overview of all domain-specific tokens added.

[67] *TaxBERT* has 82,298,880 parameters.

Hyperparameter	Value
Attention Dropout Probability	0.3
Optimizer	AdamW
Adam Epsilon	$1.00e^{-6}$
Adam Beta Weights	(0.9, 0.999)
Weight Decay	0.001

Notes: This table presents the choice of hyperparameters for domain-adaptive pretraining of *TaxBERT*.

First, the number of epochs refers to the total number of complete passes the model makes through the training dataset during pretraining. A sufficient number of epochs allows the model to learn complex patterns and relationships in the domain-specific corpus. However, too many epochs can lead to overfitting, where the model memorizes the training data instead of generalizing well to unseen data. For domain-adaptive pretraining, the number of epochs is usually kept between 3 and 20, depending on the size and complexity of the dataset. In the process of pretraining *TaxBERT*, we use the augmented tokenizer in a 12-epoch[68] pretraining procedure. The choice of 12 epochs balances computational resources with the need to sufficiently adapt the model to the tax domain without overfitting.

Second, the effective batch size specifies the total number of training samples used collectively to update the model's parameters during training.[69] There is a trade-off between choosing a batch size that balances computational efficiency and the ability to generalize from the data. Larger batch sizes allow for smoother gradient updates but require more memory, whereas smaller batch sizes introduce more noise into the updates, which can sometimes improve generalization but slow down convergence. In the pretraining of *TaxBERT*, we employ a relatively large effective batch size of 512, which is the result of using a batch size of 64 combined with eight gradient accumulation steps (64 samples × 8 gradient accumulation steps=512 samples). This means that instead of updating the model's parameters after each batch of data is processed, the gradients are summed over eight smaller batches before performing a single update to the model's parameters, as if a single batch of 512 samples had been processed at once. This approach reduces the noise in gradient updates, making training

[68] An epoch refers to one complete pass through the entire training dataset. Training for 12 epochs allows the model to repeatedly learn from the data, refining its representations with each pass.

[69] The effective batch size can be equal to the batch size (if no gradient accumulation is used), or it can be larger if gradients from multiple smaller batches are accumulated before an update. The batch size specifies the number of samples processed in one step (forward and backward pass).

more stable and memory efficient while keeping the benefits of training with a larger effective batch size.[70]

Third, the learning rate controls the step size at which a model updates its parameters (weights) during training in response to the computed gradients. Thus, it determines how quickly the model learns from the data. Choosing the optimal learning rate is difficult. If it is too low, it causes the model to take very small steps, resulting in slow training. If it is too high, it causes the model to take large steps, potentially overshooting the optimal point in the loss function, which measures how far the model's predictions are from the true labels. For the pretraining of *TaxBERT*, we use $5e^{-5}$ as the learning rate. This threshold is often used to fine-tune pretrained *BERT* models to avoid overwriting existing knowledge. Following suggestions in Devlin et al. (2019), we use a warmup linear learning rate scheduler, where the learning rate increases linearly for the first few steps (warm-up phase) and then decreases linearly over the remaining training steps. This approach stabilizes the training process by preventing abrupt gradient updates during the early phase and improves convergence, which can indirectly help reduce the risk of overfitting.[71]

Evaluation

In the final step of the pretraining, the model's performance is assessed on a validation dataset to monitor its ability to generalize beyond the training data. This step ensures that the model learns meaningful patterns rather than overfitting to the training set. The most common validation loss metric is the Cross-Entropy Loss, a measure of how well the predicted probability distribution matches the actual distribution. Lower loss indicates better generalization to unseen text.[72] Appendix IV-2 explains the Cross-Entropy Loss in more detail.

We evaluate the performance of all *TaxBERT* models in a masked language model objective by randomly splitting the tax corpus into 80% training data and 20% validation data. Table IV-4 reports the cross-entropy loss. We find that *TaxBERT* shows the lowest validation

[70] We note that the maximum effective batch size is also highly dependent on available computational resources. For mid-range GPUs, researchers might have to reduce the batch size in order to run the code.

[71] Depending on the specific task and available computational resources, it may be helpful to perform a grid search over the ranges ϵ $\{2e^{-4}, 2e^{-5}, 3e^{-4}, 3e^{-5}, 4e^{-4}, 4e^{-5}, 5e^{-4}, 5e^{-5}\}$ for learning rate and ϵ $\{8, 16, 32, 64\}$ for batch size. This hyperparameter tuning method systematically explores a range of values for the two parameters by trying different combinations to identify the combination that produces the best performance. Alternatively, researchers could also consider using Bayesian optimization as a more efficient alternative.

[72] Importantly, the validation data is used exclusively for evaluation and not for pretraining the model to ensure unbiased performance assessment.

loss. Specifically, it decreases the cross-entropy loss by 45.2% compared to *BERT* (Devlin et al., 2019), 41.5% compared to *DistilRoBERTa* (Liu et al., 2019), and 65.1% compared to *FinBERT* (Huang et al., 2023).[73] This suggests that *TaxBERT* is better at understanding and predicting the specific language and context of tax-related disclosures.[74] It suggests that the model has learned robust representations of tax-related language that apply well across different examples within this specific domain.

Table IV-4. Loss of Language Models on a Validation Set from the Tax Corpus

Model	Validation Loss
TaxBERT	1.309
FinBERT	3.751
DistilRoBERTa	2.237
BERT	2.392

Notes: This table presents the validation loss of language models on a validation set from the tax corpus used to pretrain the *BERT* models.

Notably, the performance improvement over *FinBERT* (Huang et al., 2023) is significant, which can be attributed to several factors. *TaxBERT* is specifically pretrained on tax terminology using a rich and diverse set of tax-related documents. *TaxBERT's* vocabulary covers 72.54% of the 10,000 most frequent words in the tax corpus, indicating strong coverage of both general language and tax-specific terminology. In contrast, *FinBERT's* vocabulary covers only 32.90% of these terms. Furthermore, only 15.57% of *FinBERT's* tokens (4,806 out of 30,873 tokens) are shared with *TaxBERT*, highlighting the distinct focus of each model. This limited overlap suggests that while *FinBERT* has a comprehensive set of tokens for financial topics, it may not adequately cover the nuances and specific terms unique to tax-related documents.

3.5.5. Step 5: Fine-tuning

During the pretraining phase, the domain-specific model was trained to understand both the semantic (meaning) and syntactic (structure) aspects of language from a large corpus of unlabeled domain-related texts. In the next step, fine-tuning, the model's parameters are further refined to address specific prediction tasks tailored to downstream applications. This involves

[73] To put this performance increase into perspective, Webersinke et al. (2022) find similar improvements of 46% to 48% of *ClimateBERT* over *DistilRoBERTa* for domain-adaptive pretraining on climate-related texts.

[74] A lower validation loss generally indicates that the model is more effective at generalizing from the training data to unseen validation data, meaning it has learned relevant patterns more accurately.

training the model on labeled datasets that are specific to the task at hand, such as classification or sentiment analysis. The fine-tuning process is typically more resource-efficient than pretraining, as it requires fewer, albeit labeled, examples and focuses on aligning the model's learned representations with the requirements of the target tasks. It typically requires two steps: annotation and training on downstream tasks.

For *TaxBERT*, we first train the model to differentiate between tax-related (1) and non-tax-related (0) content within financial and non-financial disclosures. Additionally, we train *TaxBERT* to assess tax risk sentiment within disclosures, classifying sentences into the three categories: tax risk-neutral sentences (0), tax risk-related sentences (1), and tax risk mitigation-related sentences (2).

Annotation

The purpose of data annotation is to label the collected textual data with meaningful classes (Schreiner et al., 2006). The researcher has full control over choosing which paragraphs and sentences may be selected to fine-tune the model, making data annotation an instance of active learning (Cohn et al., 1994).

Sentence-level classification provides a finer granularity than paragraph-level classification that allows the model to capture relevant information within mixed-content paragraphs, where only certain sentences may pertain to relevant topics (Hemmatian & Sohrabi, 2019). This granularity is particularly useful for identifying targeted information within complex texts, such as identifying tax-related texts in financial disclosures. Moreover, it increases the number of data points available for training, which can be advantageous for model performance, especially when the annotated dataset in the domain is limited. On the other hand, not all sentences are equally helpful in providing relevant features for the model predictions (Yessenalina et al., 2010), and sentence annotation can be more subjective than paragraph-level classification.[75]

A major challenge that can arise in the annotation of domain-specific texts is the frequent appearance of specific signaling words, which can lead to biased representations and thus invalid inference (Dev et al., 2019; Papakyriakopoulos et al., 2020). For instance, when fine-tuning *TaxBERT* for risk sentiment detection, the word "risk" might frequently appear in

[75] One way to reduce the subjectivity is to hire independent annotators with domain expertise, like accountants or tax advisers, and rate each sentence by different domain experts (Hribar et al., 2022).

high-risk contexts within the training data. This may cause the model to over-rely on the presence of the word "risk," potentially biasing its ability to identify risk sentiment in sentences that do not explicitly use the term. For example, the following disclosure would not be classified as tax risk-related: "[…] the impact to us and the aviation industry would likely be adverse and could be significant, including the potential for increased fuel costs, carbon taxes or fees, or a requirement to purchase carbon credits." Consequently, the primary advantage of a *TaxBERT* model – capturing the semantic and syntactic structures of texts – would be compromised.

Three effective and resource-efficient methods to mitigate this bias are 1) word masking, 2) data augmentation, and 3) paragraph usage (Hort et al., 2024). First, in our example above, periodically masking the specific word "risk" with a neutral *BERT* token such as "[MASK]" encourages the model in the example above to focus on contextual understanding rather than specific keywords alone. Second, data augmentation can be useful, which involves expanding the training dataset by introducing variations of the original texts. Techniques include paraphrasing, synonym replacement, or injecting new examples to balance word distributions or introduce new contexts. This method reduces overfitting by increasing data diversity and minimizing the impact of any single word or phrase dominating the training set. Third, paragraph-level classification can be used instead of sentence-level classification.[76] This allows the model to consider the context provided by the surrounding sentences rather than focusing narrowly on specific words. It particularly enhances the model's ability to understand nuanced sentiment or classification signals that emerge from longer text spans.

Another significant challenge in the annotation process is determining the necessary number of manually classified examples, as this depends mostly on the complexity of the classification distinction (e.g., "tax" vs. "non-tax") and the desired accuracy. Some best practices have evolved in the literature, as outlined in Kaplan et al. (2020) and Howard and Ruder (2018). For binary classification tasks, 500 to 1,000 examples per class are generally considered to be sufficient if sentences can be clearly categorized into broader categories, such as "tax" or "non-tax." However, we recommend using more (i.e., 2,000 to 5,000) examples per class, if data availability does not hinder the annotation of such large data sets. Using more

[76] The underlying machine learning process does not differ significantly between sentence-level and paragraph-level classification. The primary difference lies in data preparation and segmentation.

examples per class significantly reduces the risk of the model overfitting to specific patterns in the smaller dataset. Moreover, a smaller dataset might not fully capture the variability of the studied domain.[77]

Training on downstream tasks

TaxBERT's first task involves labeling sentences as tax-related or not tax-related. The labels are "Yes" (tax-related) or "No" (not tax-related). Generally, mentioning the word "tax(es)" alone is not sufficient for a "Yes" label. For example, the term "pre-tax profit" would not be sufficient for a *Yes* label without proper context. A sentence is only labeled "Yes" if it includes context related to tax or fiscal topics, including compliance and litigation risks (e.g., tax evasion measures, taxable income, tax credits), levies, duties, tax authorities, financial contributions, or fiscal impacts (e.g., changes in revenue, government budget implications). A "No" label is assigned if the sentence is unrelated to income taxes, tax policies, tax obligations, or fiscal topics altogether.

For implementing the fine-tuning processes, we follow the standard practice in literature (Devlin et al., 2019; Webersinke et al., 2022; Wolf et al., 2020) and pass the final layer classification (CLS) token representation from pretraining *DistilRoBERTa* on the tax corpus to a task-specific feedforward layer for prediction. We report all hyperparameters of the implemented feedforward layers in Table IV-5.

Table IV-5. Hyperparameters for Fine-Tuning

Hyperparameter	Tax Sentence Recognition	Tax Risk Recognition
Batch Size	64	64
Learning Rate	$5.00e^{-5}$	$5.00e^{-5}$
Number of Epochs	5	3
Optimizer	AdamW	AdamW
Adam Epsilon	$1.00e^{-6}$	$1.00e^{-6}$
Adam Beta Weights	(0.9, 0.999)	(0.9, 0.999)
Weight Decay	0.01	0.01

Notes: This table presents the choice of hyperparameters for fine-tuning of *TaxBERT*.

[77] We note that also smaller, high-quality datasets can achieve high-quality results, in particular when optimized with techniques like data augmentation (Tran et al., 2017) or cumulative pretraining on related datasets, reducing the need for extensive manual annotation while still supporting strong model performance. Researchers could also use other machine learning techniques (e.g., Schreiner et al., 2006) or leverage *ChatGPT* in data augmentation (Dai et al., 2023) to reduce data annotation time.

We split our annotated datasets randomly into 80% training data and 20% validation data (Webersinke et al., 2022). Devlin et al. (2019) suggest a minimal hyperparameter tuning strategy relying on a computationally intensive grid search on the ranges ϵ {$2e^{-4}$, $2e^{-5}$, $3e^{-4}$, $3e^{-5}$, $4e^{-4}$, $4e^{-5}$, $5e^{-4}$, $5e^{-5}$} for learning rate and ϵ {8, 16, 32, 64} for batch size. We also implement an early stopping mechanism to prevent overfitting. We find that a combination of a batch size of 64, a learning rate of $5e^{-5}$, and 5 epochs is optimal for tax sentence classification (Table IV-6). Batch size, learning rate, number of epochs, and optimizer follow normal ranges in the literature (Chalkidis et al., 2020; Devlin et al., 2019; Huang et al., 2023; Webersinke et al., 2022; Wolf et al., 2020).[78] For tax risk classification, we use 3 epochs, because the stopping mechanisms that were implemented were able to detect possible overfitting with more than 3 epochs.

Table IV-6. Grid Search According to Devlin et al. (2019)

Learning rate	Batch size	Loss
$2.00e^{-4}$	8	0,1861
$2.00e^{-4}$	16	0,1857
$2.00e^{-4}$	32	0,0627
$2.00e^{-4}$	64	0,0501
$2.00e^{-5}$	8	0,0867
$2.00e^{-5}$	16	0,0515
$2.00e^{-5}$	32	0,0645
$2.00e^{-5}$	64	0,0649
$3.00e^{-4}$	8	0,6614
$3.00e^{-4}$	16	0,1850
$3.00e^{-4}$	32	0,1773
$3.00e^{-4}$	64	0,0630
$3.00e^{-5}$	8	0,0497
$3.00e^{-5}$	16	0,0723
$3.00e^{-5}$	32	0,0600
$3.00e^{-5}$	64	0,0739
$4.00e^{-4}$	8	0,6611
$4.00e^{-4}$	16	0,6610
$4.00e^{-4}$	32	0,6610
$4.00e^{-4}$	64	0,1799
$4.00e^{-5}$	8	0,0514
$4.00e^{-5}$	16	0,0570
$4.00e^{-5}$	32	0,0524
$4.00e^{-5}$	64	0,0498
$5.00e^{-4}$	8	0,6672
$5.00e^{-4}$	16	0,6655
$5.00e^{-4}$	32	0,6613
$5.00e^{-4}$	64	0,6599
$5.00e^{-5}$	8	0,0725
$5.00e^{-5}$	16	0,0458

[78] For example, for *FinBERT*, a fine-tuning strategy including a batch size of 32 and a learning rate of $2e^{-5}$ is optimal (Huang et al., 2023).

Learning rate	Batch size	Loss
5.00e^{-5}	32	0,0543
5.00e^{-5}	64	0,0444

Notes: This table presents the loss results of a grid search on the ranges ϵ {2e^{-4}, 2e^{-5}, 3e^{-4}, 3e^{-5}, 4e^{-4}, 4e^{-5}, 5e^{-4}, 5e^{-5}} for learning rate and ϵ {8, 16, 32, 64} for batch size according to Devlin et al. (2019).

Construct validity

In the next step, we evaluate the performance of *TaxBERT* compared to other LLMs by randomly splitting the annotated data corpus into 80% training data and 20% validation data. We evaluate the performance in both the text classification and the tax risk classification tasks. Tables IV-7 and IV-9 report the mean and standard deviation of the cross-entropy loss. (Mao et al., 2023; Zhang, 2004). We also follow the recommendations for construct validity in Bochkay et al. (2023) and report corresponding values for F_1-score and accuracy.

Technically speaking, the F_1-score is the harmonic mean of precision and recall with equal weight (Bochkay et al., 2023). The F_1-score builds on the concepts of true positives (TP), false positives (FP), and false negatives (FN). Following Goutte and Gaussier (2005), the F_1-score may be defined as:

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

where:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

As $Precision$, $Recall$, and F_1-score ignore true negatives (TN), we also account for the accuracy of *TaxBERT's* predictions (Hartmann et al., 2023; He & Garcia, 2009). $Accuracy$ is the share of correct predictions out of all predictions (Sokolova & Lapalme, 2009):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

In the text classification task (Table IV-7), *TaxBERT* achieves the lowest mean cross-entropy loss (0.0444) and the highest mean F_1-score (0.9904) and mean accuracy (0.9900).

Results for *FinBERT* record a loss of 0.1176, an F_1-score of 0.9686, and an accuracy of 0.9688. *DistilRoBERTa* and *GPT-4o* show higher losses and lower performance metrics, with *GPT-4o* having the highest loss (0.1805) and the lowest F_1-score (0.9372) and accuracy (0.9513). Overall, *TaxBERT* outperforms *FinBERT* by approximately 2.25%, *DistilRoBERTa* by 2.36%, and *GPT-4o* by 5.68% in terms of F_1-score performance advantage in the text classification task.

Table IV-7. Performance Analysis of the Text Classification Task

Model	Loss	F_1-score	Accuracy
TaxBERT	0.0444	0.9904	0.9900
	(0.0620)	(0.0094)	(0.0094)
FinBERT	0.1176	0.9686	0.9688
	(0.0230)	(0.0048)	(0.0048)
DistilRoBERTa	0.1435	0.9676	0.9700
	(0.0114)	(0.0039)	(0.0036)
GPT-4o	0.1805	0.9372	0.9513
	(0.0089)	(0.0117)	(0.0081)

Notes: This table presents the performance analysis of the text classification task using different LLMs. Standard deviations are presented in parentheses.

To directly compare model performance, we focus in the next step on classification disagreements between models. Table IV-8 compares the classification outcomes of *TaxBERT* with those of *DistilRoBERTa* (Panel A) and *FinBERT* (Panel B) using the McNemar test on a manually labeled dataset that was not part of the fine-tuning process, ensuring an independent evaluation of generalization performance.

In Panel A, *TaxBERT* and *DistilRoBERTa* differ in 63 cases, with *TaxBERT* correctly classifying 52 instances that *DistilRoBERTa* misclassifies, versus 11 in the opposite direction. The resulting McNemar test p-value of 0.0000 indicates a statistically significant performance advantage for *TaxBERT*. In Panel B, *TaxBERT* and *FinBERT* show 35 disagreements, but the asymmetry (23 vs. 12) yields only a p-value of 0.0895. *TaxBERT* also outperforms *FinBERT* numerically. Overall, the McNemar test highlights that *TaxBERT's* performance gains over *DistilRoBERTa* are robust and statistically validated, while the advantage over *FinBERT*, although present, is less pronounced.

Table IV-8. McNemar Contingency Test

Panel A: Cross-classification of predictions made by *TaxBERT vs. DistilRoBERTa*					
		TaxBERT			
		Misclassified	Correctly classified	Total	p-value
DistilRoBERTa	Misclassified	7	52	59	0.0000
	Correctly classified	11	3,930	3,941	
	Total	18	3,982	4,000	
Panel B: Cross-classification of predictions made by *TaxBERT vs. FinBERT*					
		TaxBERT			
		Misclassified	Correctly classified	Total	p-value
FinBERT	Misclassified	6	23	29	0.0895
	Correctly classified	12	3,959	3,971	
	Total	18	3,982	4,000	

Notes: This table presents the cross-classification of predictions made by *TaxBERT* versus *DistilRoBERTa* (Panel A) and *FinBERT* (Panel B). Correct and incorrect classifications are contrasted to assess differences in model performance using the McNemar test. Asymmetries in misclassifications indicate potential statistically significant performance differences between the models.

In the tax risk classification task (Table IV-9), *TaxBERT* again demonstrates superior performance with a mean loss of 0.3349, a mean F_1-score of 0.8859, and a mean accuracy of 0.8500. The other models exhibit higher mean losses and lower mean F_1-scores and mean accuracies. *FinBERT* tests result in a loss of 0.6288, an F_1-score of 0.7213, and an accuracy of 0.7400. *DistilRoBERTa* and *GPT-4o* also perform less effectively. Regarding the F_1-score performance advantage in the tax risk classification task, *TaxBERT* outperforms *FinBERT* by approximately 22.82%, *DistilRoBERTa* by 27.13%, and *GPT-4o* by 27.37%.

Table IV-9. Performance Analysis of the Risk Classification Task

Model	Loss	F_1-score	Accuracy
TaxBERT	0.3349 (0.1309)	0.8859 (0.0383)	0.8500 (0.0350)
FinBERT	0.6288 (0.0564)	0.7213 (0.0297)	0.7400 (0.0290)
DistilRoBERTa	0.6770 (0.0883)	0.6969 (0.0321)	0.7550 (0.0272)
GPT-4o	0.8155 (0.0177)	0.6955 (0.0096)	0.7050 (0.0177)

Notes: This table presents the performance analysis of the risk classification task using different LLMs. Standard deviations are presented in parentheses.

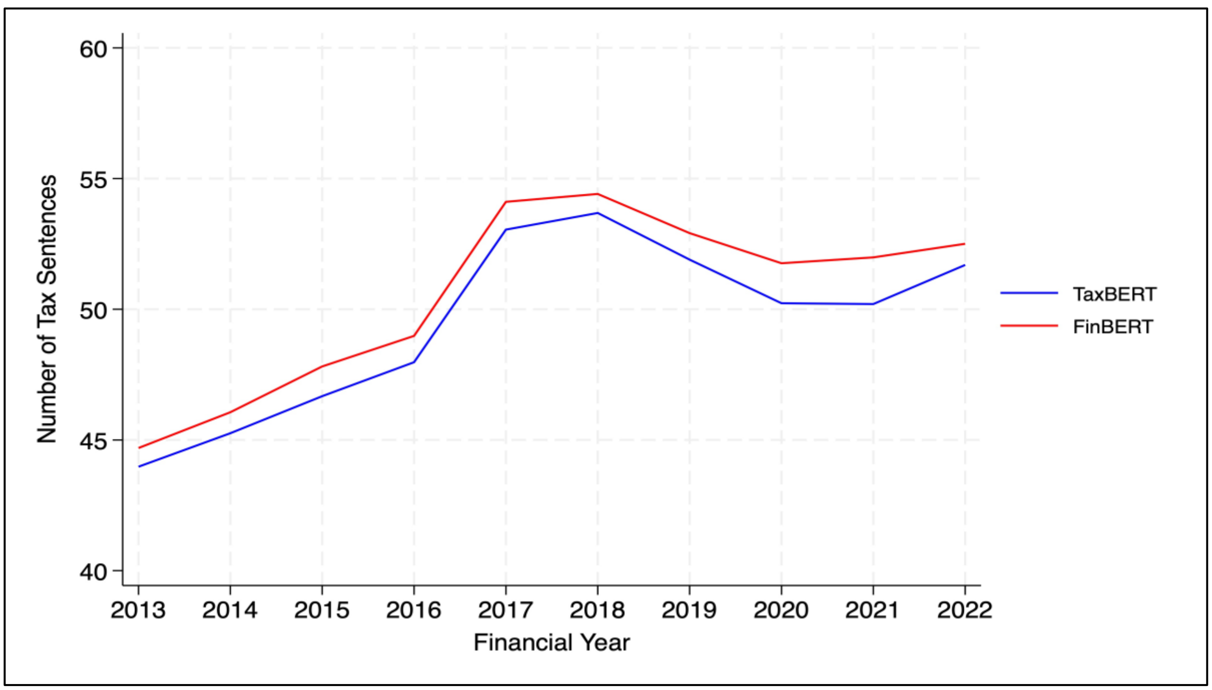
Taken together, the results suggest that *TaxBERT* consistently outperforms other LLMs in both classification tasks, achieving lower cross-entropy losses and higher F_1-score and accuracies. Overall, the initial technical results indicate that *TaxBERT* is better at capturing nuances in tax-related risk assessments.

4. Use case: Tax Avoidance and Tax Risk Sentiment in 10-K filings

This section illustrates how *TaxBERT* can be effectively applied to empirical research. To this end, we apply our model to a sample of 67,784 10-K filings of listed firms in the United States (U.S.) to evaluate its ability to measure tax risk sentiment in financial disclosures. Next, we examine the association between tax avoidance and tax risk sentiment in tax-related disclosures. This demonstration, based on a concrete use case, shows how our domain-specific model can move beyond theoretical promise and address real analytical challenges faced by researchers working with unstructured textual tax-related data. For more comprehensive use cases, see recent papers applying *TaxBERT,* including Hardeck et al. (2025) and Tan et al. (2025).

Qualitative tax-related disclosures in 10-K filings of SEC-listed firms have steadily increased over the past decade, highlighting the increasing need to leverage NLP techniques for effective analysis. As illustrated in Figure IV-1, the average number of tax-related sentences per filing rose from 44 in 2013 to 52 in 2022.

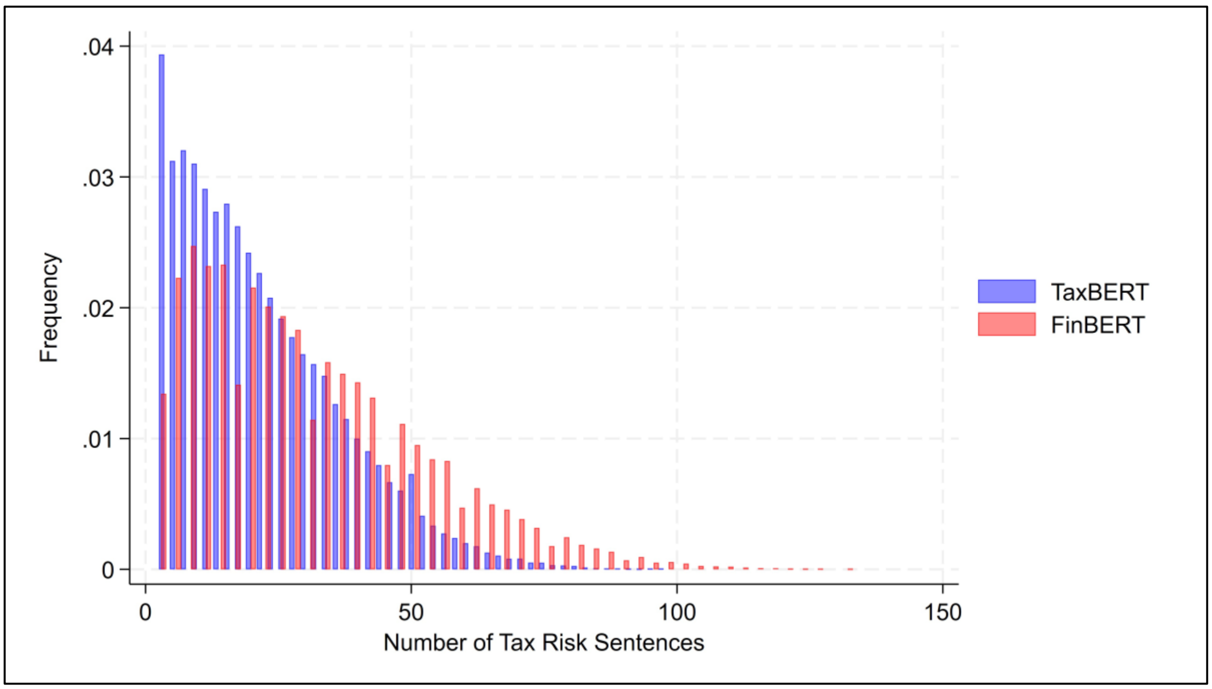
Figure IV-1. Temporal Trends in Tax-Related Sentences Identified by TaxBERT and FinBERT



Notes: This figure shows the time trends of the number of tax sentences identified by *TaxBERT* (blue) and *FinBERT* (red). This figure shows the trends in the number of tax sentences identified in the annual reports. To construct these figures, we calculate the annual average of the tax sentences identified by *TaxBERT* (*FinBERT*) using a sample of 67,784 firm-years from 2013 to 2022.

When classifying disclosures in 10-K filings into tax-related and non-tax-related information, it is crucial to use domain-specific models like *TaxBERT* that can accurately distinguish technical tax language from broader financial or operational content. Figure IV-2 graphically shows how *FinBERT* systematically overestimates the number of tax risk-related sentences, compared to *TaxBERT*. For instance, *FinBERT* incorrectly classifies the following sentence as tax risk-related: "However, the definition of personally identifiable information, or personal data varies by country, and continues to evolve in ways that may require us to adapt our practices to avoid violating laws or regulations related to the collection, storage, and use of consumer data."

Figure IV-2. Distribution of Tax Risk Sentence Counts in 10-K Filings Identified by TaxBERT and FinBERT



Notes: This figure shows the frequency distribution of the number of tax risk sentences in the annual reports. Frequency is calculated as the percentage of 10-K filings with a given count of tax risk sentences identified by *TaxBERT* (blue) and *FinBERT* (red). We use a sample of 67,784 firm-years from 2013 to 2022.

Vice versa, *FinBERT* identifies fewer tax risk mitigation-related sentences than *TaxBERT*. For instance, *FinBERT* classifies the following sentence as tax risk-related (i.e., indicating higher tax risk), whereas *TaxBERT* correctly identifies it as related to tax risk mitigation: "The Company mitigates this risk in many ways including country risk analysis, government relations and tax compliance activities […]." In line with the superior F_1-score and accuracy of *TaxBERT* presented in Tables IV-7 and IV-9, our empirical results suggest that *TaxBERT* outperforms the other models in extracting meaningful insights from tax disclosures.

We next differentiate between gross and net tax risk. The correlation coefficients in Table IV-10 reveal that the tax risk sentiment scores generated by *TaxBERT*, *FinBERT*, and *DistilRoBERTa* are positively correlated. However, the correlation between net tax risk sentiment (*RISK_NET*) as measured by *TaxBERT* and the respective sentiment scores from *FinBERT* and *DistilRoBERTa* is moderate, with coefficients of 0.376 and 0.216, respectively. These relatively low correlations suggest that the models interpret and measure tax risk sentiment differently, mainly due to the domain-specific pretraining of *TaxBERT*.

Table IV-10. Pearson and Spearman Correlations

	(1)	(2)	(3)	(4)	(5)	(6)
(1) *RISK_GROSS* (*TaxBERT*)	1.000	0.626 ***	0.885 ***	0.519 ***	0.866 ***	0.851 ***
(2) *RISK_NET* (*TaxBERT*)	0.472 ***	1.000	0.400 ***	0.673 ***	0.324 ***	0.354 ***
(3) *RISK_GROSS* (*FinBERT*)	0.670 ***	0.246 ***	1.000	0.622 ***	0.949 ***	0.923 ***
(4) *RISK_NET* (*FinBERT*)	0.331 ***	0.376 ***	0.624 ***	1.000	0.484 ***	0.503 ***
(5) *RISK_GROSS* (*DistilRoBERTa*)	0.649 ***	0.176 ***	0.951 ***	0.511 ***	1.000	0.978 ***
(6) *RISK_NET* (*DistilRoBERTa*)	0.644 ***	0.216 ***	0.921 ***	0.509 ***	0.975 ***	1.000

Notes: This table presents the correlations for the different tax risk sentiment variables prior to percentile rank transformation. We use a sample of 67,784 firm-years from 2013 to 2022. Pearson (Spearman) correlations are shown below (above) the diagonal. Statistical significance is denoted by *, **, and *** for the 10%, 5%, and 1% levels using a two-tailed t-test. All variables are defined in Appendix IV-6.

We further analyze the relationship between tax risk sentiment and corporate behavior. Appendix IV-5 provides an overview of the sample selection process for the multivariate regression analysis. Appendix IV-6 shows variable definitions. The results presented in Table IV-11 show that tax avoidance (*TAX_AVOID*) is significantly negatively related to different measures of tax risk sentiment (*RISK_GROSS* and *RISK_NET*). This suggests that companies with higher levels of tax avoidance tend to disclose less tax risk sentiment in their filings, potentially indicating strategic reporting behavior. Firms engaged in tax avoidance may downplay or obfuscate risks in their narratives to avoid drawing regulatory or public scrutiny. Notably, while *DistilRoBERTa* produces weaker results, *FinBERT* and the dictionary-based approach fail to capture this relationship, likely due to their limitations in accurately measuring tax risk sentiment.

Table IV-11. Effect of Tax Avoidance on Tax Risk Disclosure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	TaxBERT		*FinBERT*		*DistilRoBERTa*			*LM*	
Dependent Variable =	*RISK_ GROSS*	*RISK_ NET*	*RISK_ GROSS*	*RISK_ NET*	*RISK_ GROSS*	*RISK_ NET*	*FIN_ UNC*	*FIN_ UNC*	*FIN_ UNC*
TAX_AVOID	-2.653**	-2.960***	-1.296	-1.435	-1.838*	-2.166**	-0.112	-0.102	-0.102
	(-2.533)	(-2.647)	(-1.295)	(-1.413)	(-1.807)	(-2.080)	(-0.0970)	(-0.0886)	(-0.0886)
LENGTH_ TaxBERT	24.41***	19.97***					6.085***		
	(44.19)	(35.79)					(9.003)		
LENGTH_ FinBERT			25.77***	24.51***				6.171***	
			(48.10)	(45.31)				(9.128)	
LENGTH_ DistilRoBERTa					27.01***	24.70***			6.148***
					(47.37)	(42.85)			(9.099)
FOG	0.414***	0.206**	0.294***	0.248***	0.996***	0.891***	-0.666***	-0.659***	-0.662***
	(4.945)	(2.400)	(3.555)	(2.981)	(12.15)	(10.76)	(-7.369)	(-7.293)	(-7.326)
AGE	2.463***	3.640***	2.979***	3.252***	3.424***	4.042***	-2.904***	-2.901***	-2.898***
	(7.114)	(10.16)	(8.863)	(9.591)	(10.07)	(11.66)	(-7.634)	(-7.628)	(-7.620)
CAPX	-38.32***	-28.95***	-46.07***	-42.81***	-41.37***	-34.05***	-35.18***	-35.19***	-35.17***
	(-6.555)	(-4.774)	(-8.073)	(-7.396)	(-7.084)	(-5.691)	(-5.178)	(-5.181)	(-5.176)
INTANG	-6.482***	-5.118***	-5.411***	-5.146***	-2.976***	-2.287*	4.606***	4.584***	4.601***
	(-5.488)	(-4.079)	(-4.812)	(-4.471)	(-2.593)	(-1.923)	(3.511)	(3.495)	(3.508)
LEV	-4.089***	-3.752***	-4.885***	-4.519***	-5.017***	-3.887***	-0.122	-0.123	-0.129
	(-3.599)	(-3.130)	(-4.535)	(-4.131)	(-4.584)	(-3.456)	(-0.0949)	(-0.0959)	(-0.101)
ROA	2.657	1.323	-0.742	-1.725	4.278	3.810	-10.93***	-10.88***	-10.89***
	(0.933)	(0.440)	(-0.268)	(-0.615)	(1.528)	(1.322)	(-3.391)	(-3.376)	(-3.380)
SALEGR	-0.928	0.268	-2.141*	-1.976	-3.242**	-2.606*	1.280	1.281	1.282
	(-0.681)	(0.189)	(-1.653)	(-1.500)	(-2.448)	(-1.914)	(0.844)	(0.845)	(0.845)
SGA	8.789***	6.834***	8.898***	8.592***	7.276***	6.697***	12.09***	12.09***	12.09***
	(6.333)	(4.596)	(6.664)	(6.336)	(5.389)	(4.836)	(7.987)	(7.987)	(7.989)
SIZE	-0.597***	-0.670***	-0.431***	-0.397**	-0.515***	-0.557***	1.143***	1.132***	1.135***
	(-3.713)	(-4.035)	(-2.727)	(-2.475)	(-3.200)	(-3.375)	(6.449)	(6.392)	(6.405)
Constant	-60.09***	-41.40***	-65.58***	-60.83***	-85.26***	-76.04***	35.76***	35.39***	35.49***
	(-18.98)	(-12.74)	(-21.21)	(-19.48)	(-26.65)	(-23.48)	(9.327)	(9.229)	(9.258)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	TaxBERT		*FinBERT*		*DistilRoBERTa*			*LM*	
Dependent Variable =	*RISK_GROSS*	*RISK_NET*	*RISK_GROSS*	*RISK_NET*	*RISK_GROSS*	*RISK_NET*	*FIN_UNC*	*FIN_UNC*	*FIN_UNC*
Adjusted R^2	0.278	0.211	0.328	0.309	0.319	0.282	0.098	0.098	0.098
Observations	9,986	9,986	9,986	9,986	9,986	9,986	9,986	9,986	9,986

Notes: This table presents estimation results analyzing the impact of tax avoidance on tax risk disclosure. The dependent variables are *RISK_GROSS*, *RISK_NET*, and *FIN_UNC*. Columns 1 and 2 use *TaxBERT*, columns 3 and 4 use *FinBERT*, columns 5 and 6 use *DistilRoBERTa*, and columns 7, 8, and 9 use Loughran/McDonald (LM). T-statistics based on robust standard errors are reported in parentheses below the coefficient estimates. Statistical significance is denoted by *, **, and *** for the 10%, 5%, and 1% levels, respectively. Variable definitions can be found in Appendix IV-6. The following regression equation is estimated:

$$\boldsymbol{Dependent\ Variable_{i,t} = \alpha_0 + \alpha_1 TAX\_AVOID_{i,t} + \textstyle\sum_{j=2}^{10} \alpha_j Controls_{i,t} + Industry\ \&\ Year\ Fixed\ Effects + \varepsilon.}$$

In supplementary analyses, we use a median split to divide firms into high and low tax avoiders. In Table IV-12, we observe no significant difference in the level of disclosed tax risk sentiment between the two groups. This indicates that, on average, both high and low tax avoiders may disclose similar levels of tax risk sentiment in their narratives. However, when high tax avoiders choose to disclose tax risks, they may emphasize them more significantly, potentially as a way to pre-emptively justify their tax strategies or highlight efforts to mitigate associated risks. The positive interaction term suggests that for high tax avoiders, disclosures on tax risks may be a deliberate strategy to frame their behavior as transparent and compliant.

Table IV-12. Effect of Tax Avoidance on Tax Risk Disclosure with High/Low Tax Avoidance Interactions

	(1)	(2)
	TaxBERT	
Dependent Variable =	*RISK_GROSS*	*RISK_NET*
TAX_AVOID	-7.692***	-7.476***
	(-6.181)	(-5.599)
HIGH_TAX_AVOID	0.987	1.077
	(1.457)	(1.517)
TAX_AVOID × HIGH_TAX_AVOID	26.36***	22.29***
	(5.159)	(4.144)
LENGTH_TaxBERT	24.25***	19.83***
	(44.01)	(35.55)
FOG	0.421***	0.213**
	(5.051)	(2.481)
AGE	2.752***	3.892***
	(7.881)	(10.78)
CAPX	-37.45***	-28.21***
	(-6.418)	(-4.655)
INTANG	-6.292***	-4.971***
	(-5.342)	(-3.970)

	(1)	(2)
	TaxBERT	
Dependent Variable =	*RISK_GROSS*	*RISK_NET*
LEV	-4.567***	-4.174***
	(-4.024)	(-3.481)
ROA	7.046**	5.244*
	(2.426)	(1.712)
SALEGR	-1.512	-0.251
	(-1.111)	(-0.176)
SGA	8.823***	6.875***
	(6.383)	(4.638)
SIZE	-0.547***	-0.628***
	(-3.412)	(-3.794)
Constant	-62.99***	-43.97***
	(-19.74)	(-13.41)
Industry FE	Yes	Yes
Year FE	Yes	Yes
Adjusted R^2	0.282	0.214
Observations	9,986	9,986

Notes: This table presents estimation results analyzing the impact of tax avoidance on tax risk disclosure with the interaction of high and low tax avoidance. The dependent variables are *RISK_GROSS* and *RISK_NET*. All columns use *TaxBERT*. T-statistics based on robust standard errors are reported in parentheses below the coefficient estimates. Statistical significance is denoted by *, **, and *** for the 10%, 5%, and 1% levels, respectively. Variable definitions can be found in Appendix IV-6. The following regression equation is estimated:

$$\boldsymbol{Dependent\ Variable_{i,t} = \alpha_0 + \alpha_1 TAX\_AVOID_{i,t} + \alpha_2 HIGH\_TAX\_AVOID_{i,t} + \alpha_3 TAX\_AVOID_{i,t}} \times \boldsymbol{HIGH\_TAX\_AVOID_{i,t} + \sum_{j=4}^{10} \alpha_j Controls_{i,t} + Industry} \mathbin{\&} \boldsymbol{Year\ Fixed\ Effects + \varepsilon.}$$

5. Conclusion

In this study, we provide a comprehensive guide on designing and applying specialized large language models that are environmentally sustainable and practically feasible for accounting and tax researchers. Using a case-based approach, we introduce *TaxBERT*, a domain-specific model developed to enhance the analysis of corporate tax disclosures, particularly focusing on tax strategy, risk management, and compliance. By leveraging domain-adaptive pretraining on a large corpus of qualitative tax disclosure, we show that *TaxBERT* effectively captures the nuances of tax-related language, providing more accurate and context-sensitive interpretations. Specifically, our results demonstrate that *TaxBERT* outperforms general-purpose language models like *DistilRoBERTa* and other domain-specific models like *FinBERT* in terms of validation loss, F_1-score, and accuracy, as well as in multivariate regression analyses, which highlights its enhanced capability to understand the unique linguistic patterns and context of tax discourse. Thereby, we contribute to the tax literature by addressing a critical gap in the application of NLP to tax disclosures and providing an open-source framework for future research at this intersection.

We acknowledge that despite our efforts to develop an environmentally sustainable domain-specific *BERT* model, the development and application of *TaxBERT* remains computationally demanding. We limited the carbon footprint of this study by utilizing energy-efficient infrastructure and using renewable energy sources.

Our publicly available training code and model weights offer a foundation for further advancements in developing sophisticated NLP tools tailored for the accounting domain. Moreover, *TaxBERT* serves as a tax domain-specific base model that can be fine-tuned for a wide range of NLP tasks, such as text classification, named entity recognition, or question answering, depending on the specific needs of tax research. We note that some recent tax studies have already successfully applied *TaxBERT*, reporting notably higher performance on classification tasks compared to non-specialized language models (Hardeck et al., 2025; Tan et al., 2025). These findings suggest that our model can substantially enhance the accuracy and interpretability of NLP applications in tax research.

6. References

Acito, A. A., & Nessa, M. (2022). Law Firms as Tax Planning Service Providers. *Accounting Review*, *97*(4), 1–26. https://doi.org/10.2308/TAR-2018-0712

Alsentzer, E., Murphy, J. R., Boag, W., Weng, W.-H., Jin, D., Naumann, T., & McDermott, M. B. A. (2019). Publicly Available Clinical BERT Embeddings. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1904.03323

Araci, D. (2019). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1908.10063

Azimi, M., & Agrawal, A. (2021). Is Positive Sentiment in Corporate Annual Reports Informative? Evidence from Deep Learning. *The Review of Asset Pricing Studies*, *11*(4), 762–805. https://doi.org/10.1093/rapstu/raab005

Baker, A. C., Larcker, D. F., McClure, C. G., Saraph, D., & Watts, E. M. (2024). Diversity Washing. *Journal of Accounting Research*, *62*(5), 1661–1709. https://doi.org/10.1111/1475-679X.12542

Balakrishnan, K., Copat, R., de La Parra, D., & Ramesh, K. (2023). Racial Diversity Exposure and Firm Responses Following the Murder of George Floyd. *Journal of Accounting Research*, *61*(3), 737–804. https://doi.org/10.1111/1475-679X.12484

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A Pretrained Language Model for Scientific Text. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1903.10676

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bilicka, K. A., Casi-Eberhard, E., Seregni, C., & Stage, B. (2024). Tax Strategy Disclosure: A Greenwashing Mandate? CESifo Working Paper No. 9030. http://dx.doi.org/10.2139/ssrn.3832485

Bingler, J. A., Kraus, M., Leippold, M., & Webersinke, N. (2022). Cheap talk and cherry-picking: What ClimateBert has to say on corporate climate risk disclosures. *Finance Research Letters*, *47*(B), 102776. https://doi.org/10.1016/j.frl.2022.102776

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning* (1st ed.). Springer.

Blankespoor, E. (2019). The Impact of Information Processing Costs on Firm Disclosure Choice: Evidence from the XBRL Mandate. *Journal of Accounting Research*, *57*(4), 919–967. https://doi.org/10.1111/1475-679X.12268

Blaufus, K., Jarzembski, J. K., Reineke, J., & Trenn, I. (2025). Public Disclosure of Tax Strategies and Firm's Actual Tax Policy. *Journal of Accounting, Auditing & Finance*, *40*(3), 863-887. https://doi.org/10.1177/0148558X231200913

Bochkay, K., Brown, S. v., Leone, A. J., & Tucker, J. W. (2023). Textual Analysis in Accounting: What's Next?. *Contemporary Accounting Research*, *40*(2), 765–805. https://doi.org/10.1111/1911-3846.12825

Bozanic, Z., & Thevenot, M. (2015). Qualitative Disclosure and Changes in Sell-Side Financial Analysts' Information Environment. *Contemporary Accounting Research*, *32*(4), 1595–1616. https://doi.org/10.1111/1911-3846.12123

Cao, S., Jiang, W., Yang, B., & Zhang, A. L. (2023). How to Talk When a Machine Is Listening: Corporate Disclosure in the Age of AI. *The Review of Financial Studies*, *36*(9), 3603–3642. https://doi.org/10.1093/rfs/hhad021

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, *538*(7623), 20–23. https://doi.org/10.1038/538020a

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The Muppets straight out of Law School. ArXiv Working Paper. https://doi.org/10.48550/arXiv.2010.02559

Cho, H., & Muslu, V. (2021). How Do Firms Change Investments Based on MD&A Disclosures of Peer Firms? *Accounting Review*, *96*(2), 177–204. https://doi.org/10.2308/TAR-2017-0646

Cohn, D., Atlas, L., & Ladner, R. (1994). Improving generalization with active learning. *Machine Learning*, *15*(2), 201–221. https://doi.org/10.1007/BF00993277

Dai, H., Liu, Z., Liao, W., Huang, X., Cao, Y., Wu, Z., Zhao, L., Xu, S., Liu, W., Liu, N., Li, S., Zhu, D., Cai, H., Sun, L., Li, Q., Shen, D., Liu, T., & Li, X. (2023). AugGPT: Leveraging ChatGPT for Text Data Augmentation. ArXiv Working Paper. https://doi.org/10.48550/arXiv.2302.13007

Dev, S., Li, T., Phillips, J., & Srikumar, V. (2019). On Measuring and Mitigating Biased Inferences of Word Embeddings. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1908.09369

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1810.04805

Dinh, T., Seebeck, A., & Wolter, R. (2025). The More, the Better? Multiple Sustainability Reporting Standards and the Usefulness of Disclosures for Passive ESG Raters. SSRN Working Paper. https://doi.org/10.2139/ssrn.5193263

Durnev, A., & Mangen, C. (2020). The spillover effects of MD&A disclosures for real investment: The role of industry competition. *Journal of Accounting and Economics*, *70*(1), 101299. https://doi.org/10.1016/j.jacceco.2020.101299

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics*, *64*(2–3), 221–245. https://doi.org/10.1016/j.jacceco.2017.07.002

Dyreng, S. D., & Lindsey, B. P. (2009). Using Financial Accounting Data to Examine the Effect of Foreign Operations Located in Tax Havens and Other Countries on U.S. Multinational Firms' Tax Rates. *Journal of Accounting Research*, *47*(5), 1283–1316. https://doi.org/10.1111/j.1475-679X.2009.00346.x

Fiechter, P., Hitz, J.-M., & Lehmann, N. (2022). Real Effects of a Widespread CSR Reporting Mandate: Evidence from the European Union's CSR Directive. *Journal of Accounting Research*, *60*(4), 1499–1549. https://doi.org/10.1111/1475-679X.12424

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5). https://doi.org/10.1214/aos/1013203451

Geiger, R. S., Cope, D., Ip, J., Lotosh, M., Shah, A., Weng, J., & Tang, R. (2021). "Garbage in, garbage out" revisited: What do machine learning application papers report about human-labeled training data? *Quantitative Science Studies*, *2*(3), 795–827. https://doi.org/10.1162/qss_a_00144

Global Sustainability Standards Board (GSSB). (2019). *GRI 207: Tax 2019*. https://www.globalreporting.org/search/?query=GRI+207

Goutte, C., & Gaussier, E. (2005). A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In D. E. Losada & J. M. Fernández-Luna (Eds.), *Advances in Information Retrieval. ECIR 2005. Lecture Notes in Computer Science* (Vol. 3408, pp. 345–359). Springer. https://doi.org/10.1007/978-3-540-31865-1_25

Gupta, M., Parra, C. M., & Dennehy, D. (2022). Questioning racial and gender bias in AI-based recommendations: Do espoused national cultural values matter?. *Information Systems Frontiers*, 24(5), 1465–1481

Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. ArXiv Working Paper. https://doi.org/10.48550/arXiv.2004.10964

Hail, L., Lang, M., & Leuz, C. (2020). Reproducibility in Accounting Research: Views of the Research Community. *Journal of Accounting Research*, *58*(2), 519–543. https://doi.org/10.1111/1475-679X.12305

Hardeck, I., Inger, K. K., Moore, R. D., & Schneider, J. (2024). The Impact of Tax Avoidance and Environmental Performance on Tax Disclosure in CSR Reports. *Journal of the American Taxation Association*, *46*(1), 83–111. https://doi.org/10.2308/JATA-2021-030

Hardeck, I., Hechtner, F., Seebeck, A., & Weiß, M. (2025). Does Public Tax Sustainability Disclosure Deter Corporate Tax Avoidance? Evidence from GRI 207 Reporting. https://ssrn.com/abstract=4910591

Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2023). More than a Feeling: Accuracy and Application of Sentiment Analysis. *International Journal of Research in Marketing*, *40*(1), 75–87. https://doi.org/10.1016/j.ijresmar.2022.05.005

He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Hemmatian, F., & Sohrabi, M. K. (2019). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*, *52*(3), 1495–1545. https://doi.org/10.1007/s10462-017-9599-6

Henry, E., & Leone, A. J. (2016). Measuring Qualitative Information in Capital Markets Research: Comparison of Alternative Methodologies to Measure Disclosure Tone. *Accounting Review*, *91*(1), 153–178. https://doi.org/10.2308/accr-51161

Hoitash, R., & Hoitash, U. (2018). Measuring Accounting Reporting Complexity with XBRL. *Accounting Review*, *93*(1), 259–287. https://doi.org/10.2308/accr-51762

Hoopes, J. L., Robinson, L., & Slemrod, J. (2018). Public tax-return disclosure. *Journal of Accounting and Economics*, *66*(1), 142–162. https://doi.org/10.1016/j.jacceco.2018.04.001

Hoopes, J. L., Robinson, L., & Slemrod, J. (2024). Corporate Tax Disclosure. *Journal of the American Taxation Association*, *46*(2), 31–61. https://doi.org/10.2308/JATA-2022-037

Hort, M., Chen, Z., Zhang, J. M., Harman, M., & Sarro, F. (2024). Bias Mitigation for Machine Learning Classifiers: A Comprehensive Survey. *ACM Journal on Responsible Computing*, *1*(2), 1–52. https://doi.org/10.1145/3631326

Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1801.06146

Hribar, P., Mergenthaler, R., Roeschley, A., Young, S., & Zhao, C. X. (2022). Do Managers Issue More Voluntary Disclosure When GAAP Limits Their Reporting Discretion in Financial Statements? *Journal of Accounting Research*, *60*(1), 299–351. https://doi.org/10.1111/1475-679X.12401

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach. *Management Science*, *64*(6), 2833–2855. https://doi.org/10.1287/mnsc.2017.2751

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A Large Language Model for Extracting Information from Financial Text. *Contemporary Accounting Research*, *40*(2), 806–841. https://doi.org/10.1111/1911-3846.12832

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling Laws for Neural Language Models. ArXiv Working Paper. https://doi.org/10.48550/arXiv.2001.08361

Kravet, T., & Muslu, V. (2013). Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies*, 18, 1088–1122

Kubick, T. R., Lynch, D. P., Mayberry, M. A., & Omer, T. C. (2016). The Effects of Regulatory Scrutiny on Tax Avoidance: An Examination of SEC Comment Letters. *Accounting Review*, *91*(6), 1751–1780. https://doi.org/10.2308/accr-51433

Lai, G., Xie, Q., Liu, H., Yang, Y., & Hovy, E. (2017). RACE: Large-scale ReAding Comprehension Dataset From Examinations. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1704.04683

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. arXiv Preprint. https://doi.org/10.48550/arXiv.1909.11942

Lang, M., & Stice-Lawrence, L. (2015). Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics*, *60*(2–3), 110–135. https://doi.org/10.1016/j.jacceco.2015.09.002

Lauriola, I., Lavelli, A., & Aiolli, F. (2022). An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. *Neurocomputing*, *470*, 443–456. https://doi.org/10.1016/j.neucom.2021.05.103

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, *36*(4), 1234–1240. https://doi.org/10.1093/bioinformatics/btz682

Li, F. (2010). The Information Content of Forward-Looking Statements in Corporate Filings – A Naïve Bayesian Machine Learning Approach. *Journal of Accounting Research*, *48*(5), 1049–1102. https://doi.org/10.1111/j.1475-679X.2010.00382.x

Lin, Y., Shen, R., Wang, J., & Julia Yu, Y. (2024). Global Evolution of Environmental and Social Disclosure in Annual Reports. *Journal of Accounting Research*, *62*(5), 1941–1988. https://doi.org/10.1111/1475-679X.12575

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1907.11692

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, *54*(4), 1187–1230. https://doi.org/10.1111/1475-679X.12123

Luo, L., Ma, M. S., Omer, T. C., & Xie, H. (2024). Tax avoidance and firm value: does qualitative disclosure in the tax footnote matter? *Review of Accounting Studies*, *29*(3), 2927–2970. https://doi.org/10.1007/s11142-023-09773-w

Mao, A., Mohri, M., & Zhong, Y. (2023). Cross-Entropy Loss Functions: Theoretical Analysis and Applications. ArXiv Working Paper. https://doi.org/10.48550/arXiv.2304.07288

Mauler, L. M. (2019). The Effect of Analysts' Disaggregated Forecasts on Investors and Managers: Evidence Using Pre-Tax Forecasts. *Accounting Review*, *94*(3), 279–302. https://doi.org/10.2308/accr-52268

Müller, R., Spengel, C., & Vay, H. (2020). On the Determinants and Effects of Corporate Tax Transparency: Review of an Emerging Literature. ZEW – Centre for European Economic Research Discussion Paper No. 20-063. https://doi.org/http://dx.doi.org/10.2139/ssrn.3736747

Nayak, A., Timmapathini, H., Ponnalagu, K., & Gopalan Venkoparao, V. (2020). Domain adaptation challenges of BERT in tokenization and sub-word representations of Out-of-Vocabulary words. *Proceedings of the First Workshop on Insights from Negative Results in NLP*, 1–5. https://doi.org/10.18653/v1/2020.insights-1.1

Nesbitt, W. L., Outslay, E., & Persson, A. v. (2023). A reexamination of investors' reaction to tax shelter news: Evidence from the Luxembourg tax leaks. *Journal of Accounting and Economics*, *75*(2–3), 101537. https://doi.org/10.1016/j.jacceco.2022.101537

Nguyen, T., Raghunandan, A., & Scherf, A. (2024). How Does Judges' Personal Exposure to Financial Fraud Affect White-Collar Sentencing? *Journal of Accounting Research*. *63*(2). https://doi.org/10.1111/1475-679X.12584

Papakyriakopoulos, O., Hegelich, S., Serrano, J. C. M., & Marco, F. (2020). Bias in word embeddings. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 446–457. https://doi.org/10.1145/3351095.3372843

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1802.05365

Radford, A., & Narasimhan, K., Salimans, T. & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. OpenAI Working Paper. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). SQuAD: 100,000+ Questions for Machine Comprehension of Text. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1606.05250

Rasmy, L., Xiang, Y., Xie, Z., Tao, C., & Zhi, D. (2021). Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *Npj Digital Medicine*, *4*(1), 86. https://doi.org/10.1038/s41746-021-00455-y

Rennekamp, K. (2012). Processing Fluency and Investors' Reactions to Disclosure Readability. *Journal of Accounting Research*, *50*(5), 1319–1354. https://doi.org/10.1111/j.1475-679X.2012.00460.x

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1910.01108

Schreiner, C., Torkkola, K., Gardner, M., & Zhang, K. (2006). Using Machine Learning Techniques to Reduce Data Annotation Time. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(22), 2438–2442. https://doi.org/10.1177/154193120605002219

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2019). Green AI. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1907.10597

Seebeck, A., & Kaya, D. (2023). The Power of Words: An Empirical Analysis of the Communicative Value of Extended Auditor Reports. *European Accounting Review*, *32*(5), 1185–1215. https://doi.org/10.1080/09638180.2021.2021097

Sheng, V. S., Provost, F., & Ipeirotis, P. G. (2008). Get another label? improving data quality and data mining using multiple, noisy labelers. *Proceedings of the 14th ACM*

SIGKDD International Conference on Knowledge Discovery and Data Mining, 614–622. https://doi.org/10.1145/1401890.1401965

Si, Y., Wang, J., Xu, H., & Roberts, K. (2019). Enhancing Clinical Concept Extraction with Contextual Embeddings. *Journal of the American Medical Informatics Association*, *26*(11), 1297–1304. https://doi.org/10.1093/jamia/ocz096

Siano, F., & Wysocki, P. (2021). Transfer Learning and Textual Analysis of Accounting Disclosures: Applying Big Data Methods to Small(er) Datasets. *Accounting Horizons*, *35*(3), 217–244. https://doi.org/10.2308/HORIZONS-19-161

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Tan, Y., Wu, B., Cao, J., & Jiang, B. (2025). LLaMA-UTP: Knowledge-Guided Expert Mixture for Analyzing Uncertain Tax Positions. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3571502

Tran, T., Pham, T., Carneiro, G., Palmer, L., & Reid, I. (2017). A Bayesian Data Augmentation Approach for Learning Deep Models. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1710.10564

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62(3), 1139–1168

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). Attention Is All You Need. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1706.03762

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2019). GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1804.07461

Webersinke, N., Kraus, M., Bingler, J. A., & Leippold, M. (2022). ClimateBert: A Pretrained Language Model for Climate-Related Text. ArXiv Working Paper. https://doi.org/10.48550/arXiv.2110.12010

Wilde, J. H. (2017). The Deterrent Effect of Employee Whistleblowing on Firms' Financial Misreporting and Tax Aggressiveness. *Accounting Review*, *92*(5), 247–280. https://doi.org/10.2308/accr-51661

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. le, Gugger, S., … Rush, A. M. (2020). HuggingFace's Transformers: State-of-the-art Natural Language Processing. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1910.03771

Yessenalina, A., Yue, Y., & Cardie, C. (2010). Multi-Level Structured Models for Document-Level Sentiment Classification. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 1046–1056. https://aclanthology.org/D10-1102/

Zhang, T. (2004). Statistical Analysis of Some Multi-Category Large Margin Classification Methods. *Journal of Machine Learning Research*, *5*(Oct), 1225–1251.

Zhu, H., Paschalidis, I. Ch., & Tahmasebi, A. (2018). Clinical Concept Extraction with Contextual Word Embedding. ArXiv Working Paper. https://doi.org/10.48550/arXiv.1810.10566

7. Appendices

7.1. Appendix IV-1. Description of the Tax Corpus

In this appendix, we describe the data used to construct the tax corpus. The corpus consists of diverse disclosure types: 10-K disclosures, GRI 207 reports, UK Tax Strategy reports, Australian TTC reports, Tax Key Audit Matters, and Corporate Tax News.

Specifically, we include the tax-related paragraphs of 21,643 XBRL 10-K disclosures. These paragraphs encompass a wide range of tax-related content, including routine compliance details, statutory tax rates, deferred tax assets and liabilities, and tax contingencies. They often contain diverse and highly technical accounting language, which may not be directly relevant to identifying tax risks.

Moreover, we include 657 Australian TTC reports published on the ATO's website containing narrative information on corporate tax transparency, such as tax contributions, compliance with tax regulations, effective tax rates, and explanations of discrepancies between accounting and taxable income. In addition, we gathered 643 GRI 207 reports of listed firms in the European Union between 2019 and 2023 and 476 UK Tax Strategy reports of FTSE 250 firms between 2020 and 2023. In GRI 207 reports, firms extensively report on their tax strategy, tax risk management, and stakeholder engagement. GRI 207 directly links tax information to sustainability and sustainable development (Global Sustainability Standards Board (GSSB), 2019). Some of the GRI 207 reports contain a public Country-by-Country Reporting (CbCR). The UK Tax Strategy requires firms to describe the business's approach to tax risk management and governance (Bilicka et al., 2024; Blaufus et al., 2025).

Finally, we include 3,585 Tax KAMs derived from Audit Analytics for listed firms in the European Union and the UK for the period 2010 to 2023. These tax risk descriptions also provide valuable insights into a firm's tax environment. For instance, ISA 700, as implemented in the UK and Ireland in 2013, requires auditors to disclose and describe the risks of material misstatement that had the greatest effect on the audit (Seebeck & Kaya, 2023). As most KAMs are differentiated into sections for a) description, b) response, and c) conclusion, we expect Tax KAMs to be especially useful in identifying material tax risks and extracting the auditor's narrative response to them.

We additionally searched 1,176 news articles from the LSEG News Monitor published between July 1, 2023, and July 1, 2024, that are tagged with tax topics including "tax avoidance," "tax strategy," "tax compliance," and "tax risk management." All researchers manually validated the tax relevance of each article. By incorporating news articles, *TaxBERT* captures a wider range of perspectives and interpretations on tax-related topics. These articles often discuss tax issues from various viewpoints, such as those of policymakers, businesses, analysts, and the public. This diversity enables the model to develop a more nuanced understanding of how tax topics are perceived and discussed across different contexts, making it more versatile in handling various types of tax-related content. Moreover, the inclusion of news articles published within a recent timeframe ensures that *TaxBERT* is exposed to the most current language, trends, and developments in tax-related discourse. Finally, news articles often cover real-world implications of tax strategies, compliance issues, and policy changes, providing a broader context of how these topics affect businesses, governments, and the public. They include sentiment-rich language, capturing the tone and public perception of various tax topics. Training on this content enables *TaxBERT* to better recognize and interpret sentiment and tone.

We exclude tax policy and regulatory documents from the corpus, as they primarily focus on setting guidelines, legal frameworks, and compliance requirements from a regulatory perspective, rather than on the strategic or risk-management perspectives found in corporate disclosures. These documents are also less relevant for understanding the stakeholder perspective, which often centers on how firms communicate their tax strategies, risks, and compliance in ways that build trust and transparency. Consequently, including such documents may dilute the model's ability to understand the specific language, tone, and context used by firms when discussing their tax strategies and risk management practices. We also do not include social media data, as we assume that these texts are too noisy.

7.2. Appendix IV-2. Cross-Entropy Validation Loss

Following Zhang (2004) and Mao et al. (2023), the Cross-Entropy Validation Loss Function may be defined as:

$$Loss_V = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\log(p_{ij})$$

where:

$Loss_V$: The average cross-entropy validation loss.

N: The number of examples in the validation dataset.

C: The number of classes.

y_{ij}: The true label indicator, for example, *i* and class *j*. It is a binary indicator that is 1 for the correct class and 0 for all other classes for each example.

p_{ij}: The predicted probability that example *i* belongs to class *j*, obtained from the model's output after applying the softmax function: $p_{ij} = \frac{e^{z_{ij}}}{\sum_{k=1}^{C} e^{z_{ik}}}$, where z_{ij} is the logit (= raw output) from the model for class *j* of example *i*. (Bishop, 2006; Vaswani et al., 2023).

The following example serves as a foundational illustration for understanding the principles of loss computation in machine learning, with implications for developing more accurate predictive models in various domains. Suppose we have:

N=two examples

C=three classes (Risk, Neutral, Mitigation)

True Labels (y_{ij}):

Example 1 belongs to Class 2: $y_1 = [0, 1, 0]$

Example 2 belongs to Class 3: $y_2 = [0, 0, 1]$

Predicted probabilities (p_{ij}):

Example 1: $p_1 = [0.2, 0.7, 0.1]$

Example 2: $p_1 = [0.1, 0.2, 0.7]$

Then the loss for each example is calculated as:

$Loss_1 = -\sum_{j=1}^{3} y_{1j}\log(p_{1j}) = -[0 * \log(0.2) + 1 * \log(0.7) + 0 * \log(0.1)]$

$Loss_2 = -\sum_{j=1}^{3} y_{2j}\log(p_{2j}) = -[0 * \log(0.1) + 0 * \log(0.2) + 1 * \log(0.7)]$

So that the total loss is:

$$Loss_v = -\frac{1}{2}[\log(0.7) + \log(0.7)] = -\log(0.7) = 0.3567$$

7.3. Appendix IV-3. Added Tokens

The following provides an overview of all added tokens during vocabulary augmentation:

'proposed', 'ruling', 'exchange', 'commissioner', 'corporation', 'determine', 'authority', 'submitted', 'ensure', 'previous', 'refund', 'country', 'investigation', 'governments', 'individuals', 'finance', 'records', 'appeal', 'already', 'employees', 'reduce', 'returns', 'multinational', 'penalties', 'possible', 'pricing', 'reform', 'compared', 'countries', 'refunds', 'accounts', 'estimates', 'shifting', 'taken', 'investors', 'domestic', 'filed', 'firm', 'accounting', 'federal', 'gdp', 'become', 'savings', 'decision', 'reforms', 'administration', 'employee', 'minister', 'advice', 'indirect', 'opportunity', 'requirements', 'customers', 'commission', 'payments', 'considered', 'audits', 'client', 'havens', 'preparer', 'purposes', 'several', 'processes', 'particular', 'fraudulent', 'various', 'alleged', 'reduction', 'deduction', 'strategies', 'further', 'avoidance', 'insurance', 'additional', 'efforts', 'transaction', 'governance', 'industry', 'disclosure', 'plans', 'fraud', 'inflation', 'clients', 'strategy', 'credits', 'regulations', 'introduced', 'approach', 'claims', 'annual', 'basis', 'attorney', 'previously', 'evasion', 'european', 'provisions', 'taxed', 'scheme', 'restitution', 'prevent', 'befit', 'overall', 'filing', 'supreme', 'purchase', 'taxable', 'therefore', 'estimated', 'fees', 'u.s.', 'provides', 'burden', 'transactions', 'following', 'taxes', 'increased', 'professionals', 'spending', 'oecd', 'says', 'procedures', 'combined', 'sales', 'operations', 'solutions', 'penalty', 'jurisdictions', 'corporations', 'crypto', 'authorities', 'firms', 'pleaded', 'activities', 'practices', 'obligations', 'potential', 'district', 'costs', 'prior', 'persons', 'vat', 'parties', 'wages', 'reduced', 'provide', 'auditor', 'wealthy', 'taxpayer', 'implementation', 'receive', 'corporate', 'addition', 'gains', 'revenue', 'areas', 'shares', 'receipts', 'guilty', 'sentenced', 'department', 'arrangements', 'risks', 'companies', 'transparency', 'standards', 'advantage', 'dollars', 'legislation', 'investment', 'planning', 'directly', 'nearly', 'losses', 'however', 'benefits', 'expenses', 'regime', 'notices', 'payroll', 'debt', 'collected', 'supervised', 'revenues', 'congress', 'treasury', 'goods', 'increase', 'means', 'documents', 'strategic', 'asked', 'offshore', 'caused', 'operating', 'cannot', 'investments', 'employer', 'operative', 'taxpayers', 'prepared', 'challenges', 'partner', 'relief', 'assessment', 'collections', 'others', 'liability', 'economy', 'proposal', 'least', 'worldwide', 'regarding', 'policies', 'funds', 'towards', 'deductions', 'asset', 'fiscal', 'audit', 'regulatory', 'across', 'included', 'providing', 'entities', 'schemes', 'reducing', 'comprehensive', 'exemption', 'voluntary', 'businesses', 'ministry', 'withholding', 'taxation', 'agreement', 'systems'

7.4. Appendix IV-4. Overview of BERT-based models

In the following section, we provide an overview of five key *BERT*-based models, i.e., *BERT-Base*, *BERT-Large*, *DistilBERT*, *ALBERT*, and *RoBERTa*, highlighting their unique characteristics, advantages, and typical applications. By understanding the differences among these models, researchers and practitioners can make informed decisions about which model best suits their computational resources, task complexity, and performance needs.

7.4.1. *BERT*-Base

BERT-Base is the original version of the *BERT* model, consisting of 12 layers (transformer blocks), 768 hidden units, and 12 attention heads. It has approximately 110 million parameters. This model is often chosen for tasks where computational resources are limited or the domain-specific dataset is relatively small. *BERT-Base* strikes a balance between performance and computational efficiency, making it suitable for most NLP tasks, including domain-specific pretraining.

7.4.2. *BERT*-Large

BERT-Large is a more complex version of the *BERT* model with 24 layers, 1024 hidden units, and 16 attention heads, resulting in about 340 million parameters. It is typically selected when more computational resources are available and when the task demands a deeper understanding of the text, such as with very complex domain-specific corpora or when working with larger datasets. It tends to perform better than *BERT-Base* on many tasks due to its larger size and capacity, although it requires significantly more training time and resources.

7.4.3. *DistilBERT*

DistilBERT is a smaller, distilled version of *BERT*, with roughly 60% of *BERT's* parameters while retaining about 97% of its performance. It uses only six transformer layers, 768 hidden units, and 12 attention heads. *DistilBERT* is used when computational resources are constrained, or when faster inference is needed without a substantial loss in performance. It is particularly useful for deploying models in production environments where latency and resource consumption are critical concerns.

7.4.4. *ALBERT* (A Lite *BERT*)

ALBERT reduces the number of parameters in *BERT* by sharing parameters across layers and employing factorized embedding parameterization, while also increasing training speed and efficiency. It achieves comparable or even superior performance to *BERT-Large* with far fewer parameters. It is selected when the goal is to maximize model efficiency and scalability without sacrificing too much performance. It is suitable for scenarios where memory constraints are an issue or when scaling to very large datasets.

7.4.5. *RoBERTa* (Robustly Optimized *BERT* Approach)

RoBERTa is an optimized version of *BERT* that uses a larger training corpus and removes some of *BERT's* original training constraints, such as the next sentence prediction task. Also, it increases the batch size, training data, and the length of training, which enhances the model's robustness. Furthermore, *RoBERTa* trains on longer sequences and dynamically changes the masking pattern applied to the training data (Liu et al., 2019). It generally performs better than *BERT* on many NLP benchmarks. *RoBERTa* is chosen when there is a need for a more robust model that can handle a wider range of NLP tasks. Its superior performance over *BERT* makes it a popular choice for tasks where achieving the highest possible accuracy is critical, though it comes at the cost of increased computational resources.

7.5. Appendix IV-5. Sample Selection Process

	Firm-Years	
	Obs. Dropped	Obs. Remaining
Start: Firm-years with annual reports eligible for textual analysis from 2013 to 2022		67,784
Less: Firm-years with missing Compustat financial statement data	(32,778)	35,006
Less: Firm-years with Standard Industrial Classifications (SIC) codes 6000–6799 (Financial firms)	(7,732)	27,274
Less: Firm-years with missing values of regression variables	(17,288)	9,986

Notes: This table describes the sample selection process.

7.6. Appendix IV-6. Variable Definitions

Variable	Definition (Source is in parentheses and is Compustat unless otherwise noted)
RISK_GROSS	Gross tax risk sentiment in the annual report, measured as the ratio of tax risk-related sentences to the total number of sentences (*TaxBERT*, *FinBERT* and *DistilRoBERTa* respectively). We transform this measure into percentile rank form.
RISK_NET	Net tax risk sentiment in the annual report, measured as the ratio of tax risk-related sentences minus tax risk mitigation-related sentences to the total number of sentences (*TaxBERT*, *FinBERT* and *DistilRoBERTa* respectively). We transform this measure into percentile rank form.
FIN_UNC	Financial uncertainty in the tax sentences of the annual reports, measured as the number of words in the corresponding list to the total number of words in the tax sentences of the annual reports (Loughran & McDonald, 2011). We transform this measure into percentile rank form.
TAX_AVOID	Tax avoidance, measured as the industry-adjusted three-year sum of current tax expense (txt) starting from fiscal year *t* divided by the three-year sum of pretax income (pi) less special items (spi) starting from year *t*.
HIGH_TAX_AVOID	Indicator equal to 1 if *TAX_AVOID* is greater than the industry median by industry and year.
LENGTH_TaxBERT	Length of tax sentences, measured as the logarithm of the number of tax sentences in the annual report. Source: *TaxBERT*.
LENGTH_FinBERT	Length of tax sentences, measured as the logarithm of the number of tax sentences in the annual report. Source: *FinBERT*.
LENGTH_DistilRoBERTa	Length of tax sentences, measured as the logarithm of the number of tax sentences in the annual report. Source: *DistilRoBERTa*.
FOG	Readability of the tax sentences in the annual report, measured with the Gunning Fog Index.
AGE	Firm age, measured as the logarithm of the sum of 1 and the number of years listed in Compustat.
CAPX	Capital expenditures, measured as the ratio of capital expenditures (capx) to lagged total assets (at).
INTANG	Intangibility, measured as the ratio of intangible assets (intan) to lagged total assets (at).
LEV	Leverage, measured as the ratio of the sum of long-term debt (dltt) and deferred revenue (drlt) to lagged total assets (at). Long-term debt (dltt) set to 0 if the data are missing following Dyreng and Lindsey (2009).
ROA	Return on assets, measured as the ratio of pre-tax income (pi) to lagged total assets (at).
SALEGR	Sales growth, measured as subtraction of 1 from the ratio of sales (sale) in year *t* to sales (sale) in year *t-1*.
SGA	Selling, general and administrative expenses, measured as the ratio of selling, general and administrative expenses (xsga) to lagged total assets (at).
SIZE	Size, measured as the logarithm of the market value of equity (csho×prcc_f).

Notes: This table reports the variable definitions.

8. Statement on Funding

This paper received the Dr.-Michael-Munkert Research Award 2024. No additional financial resources, grants, or assistance were received.

9. Statement on Declaration of Interest

No potential conflict of interest was reported by the author(s).

10. Statement on the Use of Artificial Intelligence (AI)

During the preparation of this work, the authors used *ChatGPT*, *DeepL Translate*, *DeepL Write*, and *Grammarly* to improve the language and readability, exercising caution. After using these tools, the author reviewed and edited the content as needed, taking full responsibility for the work's content.

11. Statement on Data Availability

The data presented in this study are available on request from the authors. The training code and model weights for all LLMs used in this study are available at the following GitHub repository: https://github.com/TaxBERT/TaxBERT and Hugging Face: https://huggingface.co/mariusweiss/TaxBERT