Unstandardized Accounting Terminology*

Holger Daske
University of Mannheim
Carol Seregni
The Wharton School
University of Pennsylvania
Matthias Uckert
University of Amsterdam

This version: October 2025

ABSTRACT. The communication of accounting information requires a domain-specific vocabulary, and, in specialized languages, standardization is considered a key to clear communication, i.e., one term should only be assigned to one concept and vice versa. In practice, accounting terminology is unstandardized and produces undue complexity. We provide the first large-sample evidence on the level and the implications of unstandardized accounting terminology for a global corpus of annual reports. Our study shows that unstandardized accounting terminology is widespread and has economic consequences, increasing human and computerized information processing costs.

* We thank Joachim Gassen, Christian Leuz, Bradford Levy, Volkan Muslu (discussant), Beatriz García Osma, Cathy Schrand, Emily Shafron (discussant), Steven Young, Rodrigo Verdi, Joe Weber, Eric Weisbrod, Christina Zhu, and workshop participants at Free University of Bozen-Bolzano, Harvard Business School, IESE Business School, the Junior Accounting Faculty Conference at Columbia Business School, MIT, the Yale Accounting Conference, and the University of Amsterdam (UvA) for helpful comments and suggestions. We are grateful to the Wharton Behavioral Lab for helping with data collection for the experimental evidence. The survey in this study was approved by the IRB at the University of Pennsylvania. We gratefully acknowledge the financial support from the German Research Foundation Project-ID 403041268 – TRR 266 Accounting for Transparency, the Graduate School of Economic and Social Sciences of the University of Mannheim, and the Wharton School.

It is true that a good many of us hope to live long enough to derive lasting benefits from a standardization of accounting terminology, but the ultimate success of our undertaking depends almost entirely upon the attitude of the next generation. — Preinreich, 1933

1. Introduction

Regulators and standard setters have long emphasized improving the clarity of financial disclosures by reducing undue complexity (i.e., the SEC Plain English Mandate, the IASB's Disclosure Initiative). We study firms' use of accounting terminology, a fundamental but often overlooked ingredient of accounting communication (Mills, 1989; Loughran and McDonald, 2014), and its consequences for users' information processing costs.

Accounting is the language of business. As a language, it includes grammatical, i.e., accounting rules and methods, and lexical characteristics, i.e., accounting terms and concepts (Belkaoui, 1978; Riahi-Belkaoui, 1995). In a specialized vocabulary, standardization is considered a key to clear communication (Cabré Castellví, 1999) and occurs when *one concept* is referred to by only *one term*, and *one term* only refers to that concept (Temmerman, 2000).[1] However, this principle is often violated, as accounting terms emerge from business practice, regulation, and translations (Fuertes-Olivera and Nielsen, 2011). For example, firms can talk about "property, plant, and equipment" or "tangible fixed assets" as well as "deferred revenue" or "unearned income." They use different terms to refer to the same accounting concept.

Divergent accounting terminology has long been a problem in the accounting profession (AICPA 1953, Silhan, 1978). Academics (Koonce et al., 2005; Blankespoor et al., 2014; Hoitash et al., 2021), standard setters (IASB, 2015), and regulators (ESMA, 2013) have suggested that the

[1] According to the univocity principle both synonyms (multiple terms with the same meaning) and polysemy (terms with multiple distinct meanings) should not exist. While polysemy is also present within the accounting language, its presence is very limited (e.g., the term "stock" can both refer to both the concepts inventory and shares). We focus on synonyms as a more apparent and longstanding phenomenon.

use of unstandardized terminology can impair the comparison of firms' financial statements and, in economic terms, increase users' information processing costs (Blankespoor et al., 2020). This is also the concern underlying recent initiatives on digital reporting: establishing an XBRL taxonomy requires the creation of a one-to-one mapping where *one tag* describes *one element* (i.e., an accounting concept) (IASB, 2019; Hoitash et al., 2021). The rationale for these arguments, found in the early work on accounting terminology, is that a lack of standardized terminology gives rise to unnecessary complexity and erodes understanding (Preinreich, 1933; Singer, 1960; Haried, 1972). As accounting instructors, we can relate to these assertions: we frequently face questions from students inquiring about whether two terms have the same meaning. When we take this conjecture to the lab, students who receive a statement with a more standardized terminology make fewer mistakes and process information faster. Yet terminology standardization has received little attention in empirical research.

We provide the first large-sample evidence on accounting terminology. Our analysis proceeds in two steps. First, we describe firms' use of accounting terms and concepts and shed light the level of terminology standardization in reporting practice. Second, we evaluate the implications of unstandardized terminology on users' information processing costs.

We focus on firms' annual reports as the primary output of the financial accounting system and collect them for firms worldwide over the past two decades. We rely on SEC EDGAR for US firms and Perfect Information Filings Expert for firms incorporated in other countries.

As no comprehensive accounting thesaurus exists, we start by creating two complementary term and concept lists of accounting terminology, including multi-word expressions, n-grams. For our "top-down" lists, we gather accounting terms from a range of authoritative sources focused on accounting terminology (accounting dictionaries, standards, and XBRL taxonomies). Our top-down term and concept lists consist of 20,492 unique accounting terms and 964 concepts with an

average of 5.4 terms per concept (total 5,166 synonyms). For our "bottom-up" lists, we exploit the structure of firms' XBRL filings. Their tagging allows us to parse the terms firms use to describe the same line item and therefore map the different terms firms use in practice for the same accounting concept. Our bottom-up lists consist of 63,099 unique terms and 6,155 concepts with an average of 11.6 terms per concept (total 71,485 synonyms). These lists have unique advantages and disadvantages: while the top-down approach builds on specialized dictionaries, it relies on the sources' classification of synonyms, which may not reflect subtle differences in meaning and the extent of variation in practice. Instead, the bottom-up approach controls for the underlying economics of transactions, but the labels also capture writing conventions (stylistic differences in word choice, word order, or additional details) rather than real synonyms, and the same term can describe different but semantically related tags (due to the hierarchy of the XBRL taxonomy).[2]

We start by describing firms' use of accounting terminology. We find that, on average, words comprising accounting terms represent around 15 to 20 percent of the vocabulary a firm uses in its annual report (excluding stop words). For an average concept in our lists, the likelihood that two firms within an industry-year use the same term for a concept is only between 50 percent and 60 percent, even when firms apply the same accounting standards, such as in the United States. We report the most frequent terms and concepts that appear in firms' reports and provide examples of the variation we observe and concepts' evolution. Collectively, our statistics uncover novel macro-level insights into firms' use of accounting terminology and document that accounting terminology is often unstandardized in practice.

Next, we identify real-world settings where we expect unstandardized terminology to represent a key friction. We examine three sets of users of accounting information. First, we study analysts

[2] For example, firms may use the term "accounts receivable" when describing the XBRL tags <us-gaap:receivablesnetcurrent> and <us-gaap:accountsreceivablenetcurrent>. Both tags appear in the US GAAP XBRL Taxonomy, but the former should include accounts receivable and other types of receivables.

who collect accounting data for commercial databases (i.e., Compustat and Worldscope). Second, we consider recent advancements in technology and examine the ability of artificial intelligence (AI) in the form of a generative pre-trained transformer (GPT) to identify accounting items, mimicking the job of a data analyst. Third, we study financial analysts, a traditional group of users of accounting information. Each setting comes with unique features that help provide specific insights with respect to unstandardized terminology. The data analysts' test allows a direct mapping between terminology standardization and a measure of information processing cost at the concept level. With the GPT test, we account for the fact that disclosures are subject to both manual (human) and automated (computerized) processing (or a combination) (Allee et al., 2018). We showcase the importance of the underlying training corpus and prompting (task complexity) in assessing whether AI can process accounting data and whether unstandardized language represents a friction for machines. In our financial analysts' tests, we exploit information about the portfolio of companies an analyst follows. Our presumption is that, if we can show that unstandardized terminology affects these sophisticated parties, it will matter even more to other users, who have less attention, expertise, or time.

Empirically, we need to capture the level of terminology standardization. As our main proxy, we introduce a measure from linguistics, concept-based standardization (CBS). The CBS (based on the Manhattan distance) captures whether two firms use the same term when referring to a concept. For the GPT test, we develop three sets of measures based on core characteristics of the technology—masking predictions, term frequency, and cosine similarity of a term embedding—at identifying the term GPT would recognize as the standard one within a concept. We use these in the analyses described below. The granularity of our measurement, at the concept-firm-year level, allows for a tight fixed effects structure (i.e., including firm-year fixed effects), which helps rule out alternative explanations.

We begin our investigation with commercial data providers (i.e., Compustat and Worldscope), whose analysts extract large quantities and many different types of accounting data from firms' annual reports. As the financial industry and investors outsource the collection of firms' financials to data providers (CFA Institute, 2016; Tarca, 2020) that are subject to processing constraints (e.g., Akbas et al., 2018; Schaub, 2018), missing or wrong values of data items spill over to many users' decision-making. We posit that issues across commercial databases shown in the literature may also stem from ambiguities in terminology. (For a review, see Chychyla and Kogan, 2015.) We focus on missing items (e.g., Chen et al., 2015). We hypothesize that missing items are negatively associated with the level of terminology standardization. Our results support these predictions. Across different term lists, samples, providers, and fixed effects structures, we find that greater terminology standardization at the concept level is associated with a lower likelihood of having a missing item for the corresponding data item. In the cross-section, the effect is driven by those line items that exhibit more variation to start with—that have more synonyms, display a lower textual similarity, and for which no dominant term emerges—where information processing costs are higher. These findings back the idea that terminology standardization influences data analysts' information acquisition.

Second, recent advancements in the field of AI and emerging evidence on the capabilities of these new technologies may cast doubt on whether unstandardized terminology affects information acquisition by these new users (e.g., Bernard et al., 2024; Shaffer and Wang, 2025). We test this conjecture by asking a large language model (i.e., GPT) to match accounting items from firms' annual reports to variables in Compustat and Worldscope, essentially mimicking the job of data analysts. When we provide the GPT with real-world financial statements, we find that it makes fewer mistakes in identifying accounting items when terminology standardization is higher, across several standardization proxies and prompts, even when provided with the simplest one-to-one

matching task. This finding is consistent with unstandardized terminology also increasing complexity for machines.

Third, we focus on financial analysts and examine the association between terminology standardization and their output (coverage, accuracy, and dispersion). We find that terminology standardization, as proxied by the CBS at the firm-year, is positively associated with the properties of analysts' forecasts. Exploiting the granular information about analysts' portfolios, we further show that standardization especially benefits less experienced analysts and that standardization among the firms within an analyst's portfolio is what matters for these users.

Overall, our analyses provide consistent evidence from different empirical angles that terminology standardization impacts information processing costs.

Our research makes several contributions to the literature. First, we contribute to a longstanding literature (e.g., Preinreich, 1933; Fjeld, 1936; Bowers, 1944; Singer, 1960; Haried, 1972) by conducting the first large-scale analysis of accounting terminology, its level of standardization, and the consequences for users' information processing costs. Second, we add to the textual analysis literature in accounting and finance (for a review, see Loughran and McDonald, 2016) by constructing term and concept lists that consider the specificity of accounting language (i.e., thesauri of accounting terminology, considering n-grams) and by introducing measures to capture unstandardized terminology empirically. Citing Egozi et al. (2011), Loughran and McDonald (2016) note that it would be valuable to account for semantic relationships (i.e., meaning) when measuring document similarity. Our approach addresses this by grouping terms according to their meaning. Moreover, our descriptive evidence on firms' use of accounting terminology and the degree of terminology standardization complements the analyses of textual characteristics and disclosure topics of Lang and Stice-Lawrence (2015) and Dyer et al. (2017) by examining a distinct yet central and previously overlooked dimension of financial reporting communication. Third, we

respond to Blankespoor et al.'s (2020) suggestions for future research by illuminating that targeted linguistic properties of financial reports (i.e., the use of accounting terminology) matter for the information provided in widely used financial databases. Fourth, our evidence adds to the emerging literature on the capabilities of LLMs in financial applications (Bernard et al., 2024; Shaffer and Wang, 2025) and shows that they face information processing frictions as well.

Overall, our findings can inform standards setters' and regulators' discussions of the standardization of financial reporting (e.g., IFRS Taxonomy Project, European Single Electronic Format). Our results suggest they should care about the choice of words, as unstandardized terminology produces undue complexity. To this point, the IASB has started to recognize the importance of terminology, acknowledging that "the use of multiple English terms for very similar concepts in IFRS Standards can be maddeningly confusing to non-English speakers," and so the organization has "tried to be more disciplined in our use of English terms" (Hoogervorst, 2019). Our paper shows that word choices in disclosure regulation have economic consequences in themselves.

2. Conceptual Underpinnings

From a conceptual perspective, our paper is rooted in the early days of the accounting discipline. Accounting theorists have long suggested that firms' use of unstandardized terminology can affect the clarity of financial reporting information and thus threaten mutual understanding (Preinreich, 1933; Kohler, 1935; Tenner, 1941; Bowers, 1944; Singer, 1960). Theories from linguistics support this argument. Terminologists argue that standardization, defined by the univocity principle, is key to effective and unambiguous communication in specialized languages (Cabré Castellví, 1999; Temmerman, 2000). That is, to avoid ambiguity, "a *single term* should be assigned to *one concept* only and vice versa." (Felber, 1980). However, in practice, this is often not the case. For example, according to IFRS 10, the concept defined as "the equity in a subsidiary not attributable, directly or indirectly, to a parent" should be referred to as "non-controlling interests," but many

firms use the term "minority interests" instead. While other specialized disciplines (medicine or law) have addressed the issue of unstandardized terminology (e.g., in the medical field, Wermuth and Verplaetse, 2019), no study has investigated the true extent to which *accounting* terminology is unstandardized and the potential implications for users, i.e., whether this variation has economic consequences.

Based on these arguments, we expect unstandardized terminology to affect the understandability of firms' financial reporting information and, in economic terms, the information processing costs of users (Blankespoor et al., 2019). The idea that language standardization could affect information processing is also underscored by Blankespoor et al. (2020): "Boilerplate language could also provide a standardized structure that aids in finding relevant information …. Similarly, using consistent language over time could make it easier to identify new information in period-over-period comparisons." (p. 26). In line with this, we predict that the level of standardization of accounting terminology in firms' annual reports will affect the users' ease in identifying relevant data.

Anecdotally, we witness this conjecture as accounting instructors. Students often ask whether different terms firms use have the same meaning.[3] To illustrate the *intuition* of our line of argument, we start by using a laboratory setting where we can hold all other factors constant while examining the effects of unstandardized terminology on information processing costs.[4] We conduct a simple one-by-two between-subjects experiment and task undergraduate business students with matching accounting items from a firm's financial statements to a set of data items in Compustat. A prerequisite to participation is having taken the financial accounting class. We provide participants with (1) the income statement, (2) the asset side, and (3) the liabilities and equity side

[3] This issue is real to the point that colleagues at different institutions have told us that they provide students with lists of accounting terms and their synonyms.
[4] Our approach resembles the experiment used to validate the Bog Index in Bonsall et al. (2017).

of a hypothetical company. Each student is randomly assigned to a statement with more (less) standardized terminology. For example, a student can receive a more standardized income statement but be presented with a less standardized asset side. To operationalize, we create the statements such that (a) the more (less) standardized disclosure consists of terms that are more (less) frequently used in practice (based on our samples of US financial statements) and (b) the less standardized terms exhibit the higher textual distance from the data item names in Compustat. See Appendix A Panel A for an example set of high (low) standardized statements and Panel B for the students' task description. Other than the terms in the financial statements, everything else is held constant across conditions. Each student must match 40 different line items with Compustat data items. We recruited 73 students, resulting in a total of 2,920 observations.

The results in Table 1 show that participants receiving the more standardized disclosure take less time and make fewer errors. Panel A displays that participants, on average, take four seconds less and make 16% fewer errors when provided with a more standardized term. Results are consistent across financial statements (Panels B, C, and D), and the inferences hold when we exclude line items where the Compustat data item name corresponds to the term displayed in the financial statement (Panel E) or when we conduct a within-participant analysis (Panel F).[5] In sum, this experimental evidence from the lab illustrates that unstandardized terminology impairs human information processing (time of extraction and error).

We take this intuition into the field by identifying real-world settings where the standardization of terminology should directly impact information acquisition. We examine (1) the effects of unstandardized terminology on commercial data providers whose data analysts' specific task is to extract large quantities and many different accounting data from firms' annual reports; (2) whether

[5] In Panel F, we have to limit the sample to those participants who were assigned statements from both conditions, leaving us with 57 participants. We compute the average error rate and processing time for the items displayed in the (more) less standardized statement and run a t-test of the difference.

unstandardized terminology represents a friction for machines; (3) whether unstandardized terminology affects financial analysts, the group of users of accounting information studied most.

First, in our data providers' test, we study the association between terminology standardization and missing data as a proxy for information acquisition costs. Traditional data providers rely on a combination of computerized and manual collection to extract accounting data from firms' filings. After a first pass based on automated searches of companies' financials, data analysts trained on the collection methodologies start working on the document (see Becker et al., 2021, Section 4.2.2 for more details).[6] Studies have pinpointed data quality issues in accounting databases (e.g., Chychyla and Kogan, 2015; Boritz and No, 2020). And the Worldscope guide explicitly mentions differences in terminology as one of the main challenges analysts face when collecting raw data (Refinitiv 2020, p. 30). Nobes and Stadler (2018) document that the less common translations of the term "impairment" are associated with missing data in Worldscope. In semi-structured interviews we conducted with analysts at Compustat and Worldscope, representatives from both providers highlighted that differences in terminology affect their information acquisition.[7] Collectively, we predict a negative association between the level of terminology standardization for an accounting concept and the likelihood of having a missing item for that concept.

Second, given the recent developments in AI and the possibility of relying on these new technologies to process accounting information (Allee et al., 2018), we ask GPT to identify accounting items in real-world firms' financial statements that exhibit different levels of standardization, i.e., effectively mimicking the job of data analysts. Recent evidence from both research and practice

[6] According to our conversations with them and as observed by Du et al. (2022), traditional data providers like Worldscope and Compustat do not rely on XBRL financial statements, even when available.

[7] For example, representatives stated: "Firms do use different terms to refer to similar things; we see this across North American and Global companies. Our internal definitions and methodologies do document these differences to best categorize like items together" (correspondence with the authors). Others said: "Yes, companies may use different terminology for the same item under different accounting methods, and we do a thorough research of the document and make the call accordingly" (correspondence with the authors).

regarding real-world applications is conflicting: while OpenAI's benchmarks demonstrate that frontier AI models can handle complex tasks in economically valuable settings (Patwardhan et al., 2025), industry research suggests that only 5% of task-specific AI tools reach production deployment (Challapally et al., 2025). To what extent unstandardized terminology affects GPT's ability to retrieve accounting information is unclear. Recent studies suggest that a GPT can process financial disclosures (Shaffer and Wang, 2025), but whether it can deal with variation in technical language depends on the prevalence of that kind of variation in a GPT's training corpus. A GPT is trained on vast amounts of publicly available internet data (Achiam et al., 2023). Insofar as the training corpus includes technical language and maps synonyms of accounting terms to the same concept, the GPT should be able to identify items correctly. However, Levy and Wang (2024) document that only a small percentage of the business corpus is included in commonly crawled datasets, and a GPT therefore may not be sufficiently familiar with accounting synonyms. In addition, Bernard et al. (2024) show how complexity influences a GPT's confidence in predicting the relevant XBRL tags. Thus, unstandardized terminology could still increase complexity, affecting GPT's ability to identify the relevant accounting item.

Third, we investigate whether unstandardized terminology imposes costs on traditional capital market participants. We focus on financial analysts who rely on accounting information (Brown et al., 2015) and study the association between terminology standardization and the properties of their forecasts as a proxy for information processing costs (De Franco et al., 2011; Peterson et al., 2015). The literature on whether analysts are affected by financial reports' linguistic properties is inconclusive (e.g., Lehavy et al., 2011; Bonsall et al., 2017). While financial analysts are sophisticated users, the pure magnitude of technical terms and synonyms in reporting practice likely challenges their skills, cognitive abilities, and time. The extent to which unstandardized accounting

terminology spills over into their research output is less direct than for data analysts and therefore ultimately an empirical question.

3. Data

3.1. Capturing Accounting Terminology: Term and Concept Lists

To build a comprehensive thesaurus of accounting terminology, we construct two alternative accounting term and concept lists following a top-down and bottom-up approach (see online appendix for details).

We build the *top-down* lists as follows. We collect accounting terms and concepts from multiple sources to proxy for the universe of accounting terminology.[8] Our term list includes terms from IFRS, US GAAP, and UK GAAP authoritative sources and specialized accounting dictionaries. Terms that are explicitly classified as synonyms by these sources are grouped by the accounting concept they refer to (if any) and constitute our concept list. After some initial cleaning and standardization, these lists are refined and augmented using a GPT-based procedure.[9] After careful manual validation, the resulting thesaurus is restricted to the language found in our global corpus, which ensures that our lists reflect terminology actually used in reporting practice. The summary statistics of the top-down term and concept lists are reported in Tables 2 and 3.

[8] We acknowledge that what constitutes accounting terminology as well as its boundary is not well defined. We therefore rely on the selection of independent sources in English that specifically target "accounting terminology."

[9] Research indicates that LLMs can identify synonyms in scientific jargon (Thießen et al., 2023). We verify that this is true for accounting terminology by examining how these models represent term similarities. LLMs like GPT convert text into their vector space representations, where semantic similarities are captured by vector proximity (i.e., closely related terms or synonyms should be located closer in the n-dimensional embedding space than unrelated terms). Using a publicly available embedding model from OpenAI (TextEmbedding3Large), as well as the domain-specific ProsusAi FinBERT model for comparison, we can analyze how accounting terms are represented in vector space. We start by plotting the vector space representation of all terms part of a concept in our lists. Across all models, we consistently observe that terms within the same concept are closer together than terms from another concept. We validate the graphical evidence by computing the cosine similarity of the vector space representation of synonyms versus terms belonging to different concepts. Terms belonging to the same accounting concept exhibit a statistically significantly higher similarity than terms belonging to different concepts. See online appendix for supporting figures and tables. Note that the task of providing us with synonyms for accounting concepts differs from the task of mimicking the job of a data analyst that we study in Section 4.2.2., where the latter entails different layers of complexity that go beyond individual synonym recognition.

Our top-down approach builds on specialized dictionaries, capturing synonyms across borders and reporting regimes. However, it may have two main limitations. First, our concept list may be incomplete: we identify synonyms only for a subset of terms (i.e., the majority of terms do not have a synonym), and, for those for which we have a synonym, we may not capture the complete variation (i.e., all existing synonyms per concept). We may therefore be neglecting variation in terminology, biasing our standardization measures upward. Second, we assume the synonyms defined by our sources are real (i.e., they are used interchangeably). To the extent that our list is imprecise (i.e., identifies terms as synonyms when, in specific contexts, they may have subtle differences in meaning), we are making dissimilar things look similar, biasing our standardization measures downward.

To overcome these limitations, we develop an alternative dictionary following a *bottom-up* approach. To this end, we rely on XBRL filings on EDGAR. Specifically, we parse financial statements and extract from the respective Exhibit 101.LAB files the different terms firms use to describe the same line item, mapping them to the accounting concepts via taxonomy tags. As with the top-down lists, raw labels are standardized. A careful inspection suggests that many terms reflect different writing conventions, such as stylistic differences in word choices, word order, or added detail rather than real synonyms. Tagging errors are also frequent (in line with previous research on XBRL; see Hoitash et al., 2021). To reduce noise, we impose a frequency threshold: a term must appear in at least 20 distinct filings. All remaining terms are grouped by their underlying taxonomy tag to form bottom-up concepts. We restrict this list to the terminology actually observed in our US 10-K corpus. Finally, since some terms are linked to multiple tags (sometimes erroneously), we apply a majority disambiguation rule. That is, we remove terms within our

concepts that appear in less than 5% of all filings for a given concept (tag).[10] We create an international version of the bottom-up list by repeating these steps for non-US firms cross-listed in the United States that file 20-F reports using the IFRS Taxonomy. The summary statistics of the bottom-up term and concept lists are reported in Tables 2 and 3.

Due to their construction, the bottom-up lists ensure that we capture terms that describe the same economics (i.e., they represent different human-readable labels for the *same* XBRL tag) and that we are identifying the variation in reporting practice for a large sample of firms and over time.

When comparing the top-down and bottom-up *term* lists in Table 2, we observe that the latter consists of more terms and includes higher n-gram terms; that is, the number of terms with 3+ grams is higher than in the top-down list. This reflects the prescriptive and more technical nature of the XBRL taxonomies and the fact that preparers borrow the terminology from the XBRL tags to label accounting items in their reports.

When examining the *concept* lists in Table 3, we note, on average, more synonyms per concept in the bottom-up (10-Ks) relative to the top-down list (12.0 versus 5.4 terms per concept). We attribute this difference to the fact that the bottom-up approach captures variation in reporting practice that not only relates to what we would generally consider real synonyms but also due to different writing conventions. Concepts in our bottom-up lists exhibit a higher textual similarity (0.86 versus 0.78) and a lower concentration (0.54 versus 0.58) relative to the concepts in our top-down dictionary. While this latter type of variation at the individual concept level is most likely understandable to a careful user, it still adds complexity, as these differences, even if small, add up for users who need to process much data or compare firms' financial statements.

[10] For example, a few companies erroneously use the term "net income" for the income tax expense benefit tag, and thus "net income" appears as a synonym for "income tax expense" in our original bottom-up (10-K) list. The 5% frequency filter removes "net income" from the group of terms associated with the income tax expense benefit tag.

3.2. Usage of Accounting Terminology: Sample

Annual reports constitute the second key element of our investigation.[11] In the absence of a centralized repository for non-US filings, we employ distinct procedures to collect this data (see online appendix for details).

For non-US firms, we collect annual reports from the Perfect Information database (https://www.filingsexpert.com/) and match them to LSEG Datastream by firm name to retrieve firm identifiers and company information.[12] We remove scanned, non-English, and duplicated documents. We restrict the sample period to 2000–2019 and exclude countries with fewer than 100 available reports. To ensure sufficient accounting terminology for our subsequent tests, we require documents to contain at least the first percentile (P1) of unique accounting terms within the non-US sample. We convert all PDFs into plain text and apply the same standardization procedure as for our term and concept lists. Subsequently, we tokenize the reports at the word level while preserving the sentence structure. Tokenization enables us to assign a word to a single accounting term if that word is part of a higher n-gram (i.e., avoid double-counting). Table 4 Panel A lists our sample selection steps. Our sample of non-US firms comprises 28,531 unique firms and 226,522 firm-year observations.

For US firms, we rely on EDGAR and Compustat. Our starting point is the SEC EDGAR 10-K filings universe, which we restrict to the timeframe 2000–2019. Next, we match the remaining documents to Compustat using the Central Index Key (CIK) and the filing date retrieved from EDGAR's metadata. As in our non-US sample, we require documents to contain at least the first percentile (P1) of unique accounting terms within the US sample. Table 4 Panel B provides our sample selection steps. Our sample of US firms consists of 11,945 unique firms and 95,096 firm-

[11] We focus on annual filings as the most comprehensive output of the financial reporting system. Our thesaurus of accounting terminology will be equally applicable to other types of filings or communications where accounting information plays a role but may be supplemented by context-specific domain or jargon.
[12] The Perfect Information database stores documents using only firm names as identifiers. We develop an algorithm to match the firm name from the Perfect Information database with firm names from LSEG Datastream.

year observations. We repeat these steps to construct our sample of non-US cross-listed firms, which consists of 770 unique firms and 5,790 unique observations, as reported in Panel C.

Throughout the manuscript, we combine our non-US and US samples and refer to that as our main sample.

For our AI data analyst tests, we construct a sample of iXBRL filings from EDGAR. iXBRL documents are human-readable XBRL submissions predominantly available since 2020. Unlike standard XBRL, which provides only machine-readable tags, iXBRL embeds both tags and the human-readable labels associated with them, allowing us to more precisely link financial statement line items to their descriptive text. We parse every HTML table within the filing and compare its XBRL tags against the XBRL taxonomy to determine whether the table represents an income statement or a balance sheet. Next, we map XBRL tags within the statements to the data items in Compustat and Worldscope and apply a series of cleaning steps to the extracted tables to ensure the statements are suitable for analysis. After removing tables with formatting issues, duplicate tags, or ambiguous matches, we retain only statements with at least five unique line item matches, firms covered in both databases, and filings that include both an income statement and a balance sheet. This process yields 2,375 iXBRL 10-K filings, resulting in 4,750 financial statements. Table 4 Panel D summarizes all the steps for the construction of our iXBRL sample.

4. Empirical Analyses

4.1. Unstandardized Accounting Terminology in Reporting Practice

This section presents some selected descriptive statistics on accounting terms and concepts. We begin by assessing the importance of accounting terminology in our main sample of annual reports by calculating the percentage of accounting-related words.[13] On average, according to our term lists, accounting words represent around 15 percent to 20 percent of the text a firm uses in its

[13] We calculate the percentage of accounting words by dividing the number of words appearing in our accounting term lists by the total number of words in our main sample of annual reports. We exclude stop words and numbers from the computation.

annual report, and for an average concept in our lists, the likelihood that two firms within an industry-year use the same term for a concept is only between 50 percent and 60 percent, even when firms apply the same accounting standards, such as in the United States.[14]

Table 5 provides descriptives on terms' and concepts' usage at the document level. On average, firms use up to 1,851 (1,505) accounting terms corresponding to 482 (457) unique accounting terms in their reports, in our top-down (bottom-up) list. Terms that are part of a concept and have synonyms play a meaningful role. Firms report an average up to 200 (414) unique concepts within a given year in their annual reports (10-K). Overall, these percentages demonstrate the importance of accounting terminology and the use of synonyms in reporting practice.

Figure 1 illustrates that the most frequent accounting terms are the lower n-gram terms, and, interestingly, when we consider unique terms, most of the accounting terms that appear in firms' annual reports consist of more than one word. Table 6 Panel A reports the 25 most frequent terms using the bottom-up and top-down lists, respectively. According to these frequencies, some of the most used accounting terms in our list are also part of more general business vocabulary (e.g., "board of director," "corporate governance," "real estate," and "interest rate"), while we also observe more technical terms that one would find in a financial accounting textbook (e.g., "impairment loss" or "intangible asset"). In general, the higher the number of grams, the more technical in nature the terms get. These more technical and prescriptive terms need multiple words to describe the accounting concept they intend to capture and help distinguish the essence of different economic transactions. The observation that our term lists and documents' corpora include both a more general business vocabulary as well as more targeted accounting terminology is consistent with the fact that accounting emerges from business practice, and thus that the definition of an

[14] We calculate the likelihood of firms sharing the same term within a concept by randomly sampling pairs of firms within each industry-year-concept combination and determining whether they use identical terms. Next, we average this measure across all industry-year-concept groups in our sample.

accounting term, in practice, can have a broader or narrower scope (i.e., the boundaries are a gray zone). Taken together, our descriptives suggest that term lists that solely focus on 1-grams, such as the Loughran and McDonald (2011) dictionary, are ignoring a considerable portion of the accounting vocabulary, in particular, the more technical terms.

Next, we explore the extent to which we observe unstandardized terminology in reporting practice. Table 6 Panels B, C, and D report the five most frequent financial statement concepts by term list and provide illustrative examples of how synonyms emerge. Specifically, some synonyms result from linguistic causes because, for example, they contain synonyms within the English language, such as the terms "profit" and "income." Other synonyms are linked to the existence of different accounting standards, for instance, the terms "shareholder" and "stockholder," which come from the IFRS and US GAAP, respectively. Table 6 also provides examples of the similarity in wording across synonyms within a concept: in some cases, a synonym more precisely describes a concept, e.g., "income tax" versus "income tax expense." In other cases, terms differ in their entirety, e.g., "fixed asset" versus "long lived asset."

We also explore how the usage of synonyms evolves. Figure 2 reports two examples of different trends toward standardization that we observe for individual concepts when employing our top-down (Panel A) and bottom-up (Panel B) lists, respectively. In some instances, we see a clear shift toward one prevalent term reported, such as for "non-controlling interest." In other cases, we do not see standardization emerging for certain concepts, and several terms keep coexisting, such as for "unearned revenue."

Overall, these descriptives uncover novel macro-level insights on the magnitudes and different facets of unstandardized terminology that may moderate the effect on information acquisition: the number of synonyms used, the textual similarity, and the concentration within a concept. In the remainder of the manuscript, we test whether unstandardized terminology affects users.

4.2. Unstandardized Accounting Terminology and Information Processing Costs

This section evaluates the implications of unstandardized accounting terminology on information processing costs. We study different groups of users of financial reporting information: (i) data analysts at commercial providers, (ii) AI data analysts (GPT), and (iii) financial analysts.

4.2.1. Data Analysts at Commercial Providers

For our data providers' tests, we proxy for information processing costs by the incompleteness of a database, i.e., whether the provider reports a missing data item-firm-year observation, under the assumption that items that are costlier to collect are less likely to be collected by commercial providers facing cost-benefit trade-offs.[15] The granularity of our data and measurement allows us to conduct this analysis at a concept, i.e., data item, level. That is, we investigate whether the level of terminology standardization for a given concept in a firm-year is associated with the likelihood of data providers reporting a missing value for the data item. This direct mapping of our standardization measure and outcome variable strengthens identification.

To perform our tests, we focus on Compustat and Worldscope. We match our concept lists to the data items collected and distributed by the two data providers. Following Du et al. (2022), we focus on the data items in the balance sheet and income statement that are part of the balancing models of each provider. To determine which underlying accounting concepts are captured by these data items, we fuzzy match their labels to the synonyms in our concept lists and manually validate the matches, consulting their corresponding definitions. Following this procedure, we can match 136 (119) unique concepts to 139 (119) unique data items reported in Compustat

[15] Our tests rely on the existence of missing items in commercial databases. We observe considerable variation across data items. Three main reasons, unrelated to terminology standardization, explain missing items. First, as data collection is costly, providers prioritize the collection of items and firms with higher customer demand. Second, items may capture rarely disclosed information or information that does not apply to specific industries. Third, providers may follow different data aggregation and standardization procedures, depending on the industry. Our fixed effects and controls should mitigate the effect of these alternative explanations.

(Worldscope) for our top-down list and 107 (92) unique concepts to 95 (81) unique data items in Compustat (Worldscope) for our bottom-up list, respectively.[16]

Next, we estimate the following firm-concept-year level specification:

$$\text{Missing Item}_{i,c,t,d} = \beta_0 + \beta_1 \textit{CBS-Firm-Concept}_{i,c,t} + \sum \beta_k \text{Controls} + \sum \beta_l \text{Fixed Effects}, \quad (1)$$

where the dependent variable is *Missing Item*, an indicator variable capturing whether the data item corresponding to concept *c* for firm *i* at time *t* in database *d* is missing. The main explanatory variable is the concept-based standardization measure *CBS-Firm-Concept* of concept *c* for firm *i* at time *t*. This measure (based on the Manhattan distance) is bounded between 0 and 1 and captures the extent to which two speakers (here firms) use the same term when referring to a specific concept (see Appendix C for details).[17] Higher levels of this measure indicate higher terminology standardization between two firms for a given concept. The *CBS-Firm-Concept* measure we employ is the yearly average of the CBS calculated for each concept over each unique pair of firms in the respective sample that belong to the same industry (as defined by the three-digit ICB code) and region (United States versus Non-United States) of the focal firm. We compute this standardization measure within industry, as in our communications with data providers, they stated that their analysts are assigned to firms primarily by industry. We compute the CBS within regions,

[16] To put the number of matches into context, in their analysis of discrepancies between as-reported 10-K files of US issuers and Compustat items, Du et al. (2022) analyze 73 data items, and 55 of these appear in either the balance sheet or income statement. Note that, for the bottom-up lists, while each tag can only be matched to one database item, each database item can be matched to multiple tags. For example, the XBRL tags <us-gaap:receivablesnetcurrent> and <us-gaap:accountsreceivablenetcurrent> are both matched to the Compustat item RECT (Receivables – Total) and the Worldscope item ID02051: Receivables (Net). For the top-down list, a handful of concepts can also be matched to multiple database items. For instance, the concept "minority interest" can be found both in the balance sheet and the income statement of providers' balancing models. We retain all these matches for our analyses. In the online appendix Table OA4-1, we restrict our analyses to the one-to-one matches between the XBRL Taxonomy and Compustat from Du et al. (2022). Our results are robust to this filter.

[17] This measure seems particularly suitable to our context for two reasons. First, it was developed to study how different varieties of the same language (e.g., British and American English) converge over time by examining the extent to which different terms (i.e., synonyms) refer to the same concept. Second, the measure has been applied to languages for specific purposes (e.g., da Silva, 2010). This measure is also known as profile-based linguistic uniformity. For a detailed description, see Speelman et al. (2003) and Geeraerts et al. (2024).

as data providers employ distinct data collection procedures for U.S. and non-US firms; for instance, Compustat maintains separate databases (Compustat North America and Compustat Global) to account for these differences. We control for firm size as a proxy for market demand for information that we predict will affect the data analysts' information acquisition. We also control for other firm characteristics associated with firm information quality: return on assets, book-to-market ratio, leverage, age, and loss observations. We include firm, year, database, and concept fixed effects to account for firm, time-trend, database, and concept characteristics. We estimate our specification also controlling for the *general* level of *accounting* language similarity of the report by including the average cosine similarity at the firm-year level relative to all other firms in the same industry (Lang and Stice-Lawrence, 2015), computed using our entire accounting term lists.[18] While the cosine similarity captures *how similarly firms discuss their accounting*, the concept-based standardization (CBS), by grouping terms according to their meaning, *captures how similarly firms address the same accounting issue* (Ruette et al., 2014).[19] Finally, we also run a specification that includes firm-year fixed effects, which account for any time variant firm and report characteristics.

We present our analyses separately for our top-down and bottom-up lists. Additionally, we take advantage of our bottom-up lists and restrict the estimation only to firm concepts for which the firm has reported a specific tag in a given year, that is, where we are sure the item is non-missing in a firm's original filing. This approach addresses the concern that a data item may be

[18] Our results are insensitive to computing the cosine similarity using the Loughran and McDonald (2011) dictionary, which captures business language more broadly and speaks to *how similar firms talk in general*.

[19] Our approach follows Loughran and McDonald's (2016, p. 1215) suggestion of considering word meaning when measuring document similarity. Ruette et al. (2014) show how controlling for word meaning is necessary when studying different varieties of the same language, as the cosine similarity metric alone could also suggest that "different words are used to refer to different content." In our context, this implies that cosine similarity would treat synonyms as distinct terms, implying that firms discuss different economic constructs when, in fact, they do not. In contrast, our approach recognizes that different terms can convey the same meaning, revealing that it is the linguistic expression and not the underlying economic substance that varies across disclosures.

missing because the firm is *not reporting* a specific line item for that concept. We refer to this sample as "Has Tag."

Table 7 reports the descriptive statistics for the variables used in our data analysts' tests. We display statistics separately for our main and "Has Tag" samples when employing the top-down and bottom-up lists. On average, we observe between 4 percent and 15 percent missing data items, depending on the sample and the term list. Consistent with expectations, the "Has Tag" samples present the lowest percentage of missing items. The CBS of the concepts in our data provider tests ranges, on average, between 42 percent and 60 percent.

Table 8 Panel A Columns 1 to 4 report the results for estimating Equation 1 for the association between the level of terminology standardization and the likelihood of having a missing item in Worldscope or Compustat for our main sample when employing our top-down or bottom-up list, respectively. The negative coefficients of the *CBS-Firm-Concept* standardization measure suggest that a higher level of terminology standardization for a concept-firm-year observation is associated with a lower likelihood of the data item corresponding to that concept being missing in a commercial database. Results are robust to controlling for the document's general level of textual (accounting) similarity by including the cosine similarity. In Columns 5 to 8, we find that our results are robust when limiting the sample to data items where the firm is reporting a value in a specific year (i.e., the tag mapped to the data item is used in the report). To gauge economic significance, a one standard-deviation increase in CBS is associated with a 0.19 to 1.71 percentage-point reduction in the probability of a missing item, depending on the specification. This effect corresponds to approximately 1,240 to 11,132 fewer missing observations for an average data item in the databases. The *CBS-Firm-Concept* coefficients are virtually unchanged when we include firm-year fixed effects, control for accounting comparability, and test each database separately (see online appendix).

In Table 8 Panel B, we complement our baseline evidence by identifying concept-level characteristics that reflect the extent to which concepts are unstandardized and the sources of variation. We replicate our analysis by splitting data items according to concepts' characteristics that we predict to result in higher information processing costs and distinguishing between i) concepts with few (many) synonyms according to the number of terms within a concept, ii) concepts with a high (low) textual similarity as measured by the average cosine similarity of the terms within a concept and iii) concepts with a dominant (nondominant) term, as proxied by an Herfindal Hirschman Index (HHI) based on the relative frequencies of each term within a concept. A concept is coded as one (and zero otherwise) if it lies above (below) the respective median. Consistent with the intuition that standardization plays a bigger role for concepts that are harder to process (i.e., where more variation exists in the first place), we observe larger negative coefficients of *CBS-Firm-Concept* for concepts with many synonyms, lower textual similarity, and no dominant term. The difference in coefficient is negative and statistically significant (p-value < 0.01) in five out of the six splits/term list combinations.

Taken together, our data providers' tests show that unstandardized terminology impedes data collection, leading to a higher likelihood of missing observations in commercial databases, and that these frictions intensify for concepts characterized by greater linguistic variation.

4.2.2. AI Data Analysts

In this section, we investigate whether unstandardized terminology represents a friction for machines when processing accounting information. When selecting an appropriate AI model for this task, we considered several alternatives. While specialized encoder-only models like BERT could be fine-tuned for accounting terminology recognition, they require extensive task-specific training and therefore can only perform specific tasks. Instead, we employ the most advanced large language models (LLMs) from OpenAI, specifically GPT-4o-mini and GPT-5-mini. Unlike

encoder-only architectures, a decoder-only model like GPT can handle diverse tasks without extensive fine-tuning and pre-training (Levy, 2025; de Kok, 2025).

To conduct our investigation, we task GPT with identifying accounting items in firms' financial statements and mapping them to data items in Compustat and Worldscope, essentially mimicking the job of a data analyst. Our approach mirrors Shaffer and Wang (2025), who test GPT's ability to perform human tasks in the context of core earnings estimation. To closely resemble the task of a data analyst, we provide GPT with a large sample of real-world financial statements. We restrict this analysis to iXBRL filings, as their structure allows us to precisely identify the human-readable label associated with a specific tag.

We prompt GPT with the line items' information for one filing at a time and ask it to match them to the corresponding accounting data item in Compustat or Worldscope. As prompting influences GPT's ability to perform tasks (Levy, 2025; Shaffer and Wang, 2025), we develop six different prompts with variation along two dimensions ordered by increasing task complexity: (1) *context*, i.e., whether we only provide the list of line items for which there is a match (items only), the list of line items to match underneath the financial statement in which they appear (items with financial statement context), or only the financial statement (financial statement only); and (2) the *matching scope*, i.e., whether we provide the complete list of the database's data items (full) or the selected list of data items for which we know that a match exists (select). In Appendix D, we provide examples of the six possible prompts, with differences across prompts shown in blue. We then measure the error rate of GPT in matching the correct financial statement item to the corresponding database's data item. The different prompts allow us to gauge GPT's ability to identify accounting items across a spectrum of tasks, ranging from simplified to more realistic ones.

In addition to the *CBS-Firm-Concept* computed within the sample of iXBRL reports, we take advantage of the nature of the technology we are dealing with and compute three additional

measures of standardization.[20] These additional proxies aim at capturing what GPT would identify as the standard term within a concept and build on three core characteristics of LLMs: they are trained on massive text corpora, represent linguistic units in high-dimensional vector spaces, and generate text by predicting the most likely next token given its context. First, we expect that the frequency of appearance of the term within the training corpus will influence GPT's ability to recognize it as a synonym. As we do not have access to OpenAI's underlying training corpus, we construct a measure of the frequency of a term within a concept based on the iXBRL corpus as a proxy.[21] We label this measure *Relative Frequency*. Second, we compute the cosine similarity of each term to other terms within the same concept's vector space representation of OpenAI's text-embedding-3-large model. The idea behind this proxy is that concepts with a lower average cosine similarity contain terms whose vector space representation is farther apart, suggesting a lower likelihood of appearance in a similar context relative to other terms within the same concept. We refer to this proxy as *Embedding Similarity*. Third, we apply a masking approach. We provide GPT with financial statements from our iXBRL sample and mask one specific line item at a time. Subsequently, we ask GPT to estimate the probability that the term the firm uses for the masked line item is contextually appropriate within the statement. These probabilities reflect GPT's contextual judgment of how "natural" the term appears within the statement. We then use them to construct a measure which we label *GPT Masking Prediction*.

We then run the following regression model:

[20] The CBS calculation follows the approach outlined in Section 4.2.1, with one modification: instead of constructing the reference group within industry-year, we pool all firms in the iXBRL sample, regardless of the industry. Given the relatively short timeframe covered by our data, further splitting observations into smaller industry-year groups would leave too few peers for comparison. In untabulated results, we compute the CBS within industry and show that this alternative specification yields virtually unchanged inferences.

[21] Levy and Wang (2024) and Levy (2025) suggest that GPT is most likely not trained on financial corpora. The assumption in using our specialized corpus to proxy for the frequency of terms in the OpenAI training corpus is that, while not directly trained on it, the frequencies in our corpus are reflected in the more general content on which GPT's training occurs.

$$\text{Error}_{l,f,d} = \beta_0 + \beta_1 \textit{Standardization}_{l,f} + \sum \beta_k \text{Controls} + \sum \beta_j \text{Fixed Effects}, \quad (2)$$

where the dependent variable is the *Error*, an indicator variable capturing whether GPT failed to identify the relevant term for data item *l* in filing *f* for database *d*, and the main explanatory variable is one of the standardization measures outlined above for term *l* in filing *f*.[22] We control for two task characteristics that may affect GPT's information acquisition: the number of data items it must match *N (Data Items)* and the textual cosine similarity between the term the firm is using and the data item name in the database (*Pair Similarity*). *N (Data Items)* accounts for the increasing difficulty of matching as the number of possible pairs grows.[23] *Pair Similarity* captures the lexical similarity between terms in the financial statements and the data item name in the respective database. Higher similarity indicates greater textual overlap, which mechanically increases the probability of correct matching, even without deeper semantic understanding. Additionally, our experimental design allows us to include concept-fixed effects to account for systematic differences in matching difficulty across accounting concepts, database fixed effects to account for differences across data providers, firm, and year fixed effects, where the latter two control for firm's time-invariant reporting characteristics and time trends, respectively.

Table 9 reports descriptive statistics for all the variables used in our tests across the six possible prompts. The average *Error* is between 1 percent and 12 percent in the Selecting Matching specifications. In these specifications, GPT is provided with the firm's line items, along with a subset of database items that can be matched to the financial statements. By contrast, the error rate in the Full Matching specifications is substantially higher, ranging from 11 percent to 34 percent, where GPT is tasked with choosing from the entire set of variables available in the respective

[22] We treat every filing as an independent observation, as we do not provide GPT with information about the company or time, and each filing represents a new prompt.

[23] The probability of correct random matching decreases exponentially, following a 1/n! distribution (where n is the number of terms). For example, with just two terms, random guessing yields a 50% success rate, but this random chance drops sharply when introducing more terms. In our test, we require at least five terms, which reduces the probability of a correct match by chance to 0.83%.

database and financial statements. These statistics indicate that mistakes occur even under the most simplified task, when we instruct GPT to perform a one-to-one matching (select matching with no context). In addition, the substantially higher error rates we observe for the prompts that more closely resemble the task of a data analyst, i.e., the Full Matching ones, suggest that GPT is still incapable of handling more complex tasks that replicate the job of a human. While unstandardized terminology may contribute to errors in our Full Matching prompts, other factors related to the underlying functioning of GPT are also likely to drive these errors (e.g., the needle in the haystack problem). For this reason, to be conservative, we restrict the reporting of our regression analyses to the Selected Matching prompts, i.e., the setting in which GPT performs *best*. When it comes to the standardization measure, the mean *CBS-Firm-Concept* is 40 percent, in line with what we observe in the data analysts' tests. All standardization measures indicate that, on average, it is possible to identify a standard term within a concept. *Relative Frequency* suggests that the relative frequency of a term within a concept is, on average, 37 percent. The *Embedding Similarity* measure is on average 67 percent, indicating a high embedding similarity within a concept. The mean *GPT Masking Prediction* is 87 percent, indicating that GPT assigns, on average, an 87% likelihood that the firm's actual label is contextually appropriate when evaluated in its masked statement context.

Table 10 reports the results from estimating Equation 2 for our Selected Matching specifications. Panels A, B, and C display estimates for the three prompting variations along the "context" dimension. For each panel, we report results using both GPT 4o-mini and 5-mini.[24] Across all the different standardization proxies and under a restrictive set of fixed effects, we consistently observe negative, statistically significant coefficients, demonstrating that unstandardized terminology represents a friction also for GPT. For example, using our main measure (*CBS-Firm-Concept*)

[24] These estimations use a temperature of 1 as at the time of conducting our analysis; GPT-5 models do not allow for a zero temperature. We can run GPT-4o with a temperature of 0, and the coefficients are almost identical. In addition, we also provide results for GPT-5-nano. We report these regressions in the online appendix.

in Panel A, a one standard-deviation increase in standardization is associated with a 0.58 percentage point reduction in errors for GPT-4o-mini and 0.65 for GPT-5-mini, representing approximately 27 percent to 45 percent of the baseline error rate of 1 percent to 2 percent in this simplified task. Panel C exhibits the largest coefficient magnitudes—approximately three times those observed in the other panels. This panel reports results for the Financial Statement Only prompt, which best mirrors a data analyst's task by providing GPT with raw financial statements and a list of database items that can be mapped to individual line items but without any explicit hints as to which line items have a match (in contrast to Panels A and B). Among the proxies and across all panels, *Embedding Similarity* consistently exhibits the largest coefficient magnitudes, suggesting that synonyms that are more likely to appear in the same context within GPT's training corpus and therefore yield a higher embedding similarity are easier for the model to process.[25] This pattern is consistent with the architecture of GPT, where linguistic units are represented through high-dimensional vector spaces based on the context in which they appear. The coefficients of the task characteristics controls are also significant and comport with our expectations.

Overall, our results show that, when tasked with identifying relevant accounting information, even in simplified matching tasks, LLMs still cannot fully undo the complexity introduced by unstandardized terminology without further domain-specific fine-tuning. We deliberately use base models rather than fine-tuned versions for two reasons. First, fine-tuning introduces researchers' degrees of freedom that could hamper the generalizability of our findings (different fine-tuning approaches, datasets, and hyperparameters could yield vastly different results), impeding isolation of the effect of terminology standardization itself. Second, by documenting base model

[25] One potential concern is that the strong association between Embedding Similarity and the error rate may in part reflect mechanical similarity, since both measures rely on models from the same OpenAI and likely share overlapping training corpora. While text-embedding-3-large is not the underlying embedding layer of GPT, the two representations may be closely related. To address this concern, we replicate our analysis in untabulated results using FinBERT embeddings, a domain-specific model trained independently of GPT, and obtain qualitatively similar results, suggesting that our inference is not solely driven by model-related bias.

performance, we establish a benchmark that reveals the magnitude of the terminology standardization problem that fine-tuning would need to overcome.[26]

4.2.3. Financial Analysts

In the last step of our analyses, we study whether unstandardized terminology represents a friction for financial analysts. We examine the association between the level of terminology standardization and analysts' coverage and the properties of their forecasts—accuracy and dispersion—as measures of information processing cost (e.g., De Franco et al., 2011; Peterson et al., 2015). This setting comes at the cost of losing the direct mapping between our measure and outcome (i.e., we do not observe the analyst's output at the concept level); instead, we capture the output of the complete process of information collection, processing, and forecast production by an analyst for a firm year. Yet it equips us with established proxies for outcome variables and provides detailed information on an individual analyst's portfolio, enabling us to investigate whether standardization is more beneficial for less *experienced* analysts (capturing the notion that unstandardized terminology poses a greater challenge for users with less expertise) and whether the extent that firms standardize within an *individual user's portfolio* matters most (reflecting that what is relevant for a given user is standardization toward that user's reference group). To conduct our tests, we obtain information on consensus and individual analysts' EPS estimates from the unadjusted *summary* and *detail files* in the I/B/E/S database through WRDS. We merge this data with information on price and adjustment factors from Datastream. We adjust firms' actuals to account for stock splits

[26] Indeed, major database providers are investing substantial resources in developing domain-specific models. Wu et al. (2023) report that BloombergGPT, a 50-billion parameter model, was trained on a 363 billion token dataset based on Bloomberg's extensive data sources augmented with 345 billion tokens from general purpose datasets. Yet even these well-resourced efforts face fundamental limitations: fine-tuned models might suffer from catastrophic forgetting, where the model loses essential knowledge acquired during pretraining (Kirkpatrick et al., 2017; Luo et al., 2025). Our base model results thus identify the fundamental linguistic challenge that persists, even before considering the complexities and costs of fine-tuning.

(Payne and Thomas, 2003) and calculate the mean accuracy and dispersion at the firm-year level based on the mean forecast of each analyst.[27]

We then run the following regression model at the firm-year level for our top-down and bottom-up lists, respectively:

$$\text{Coverage/Accuracy/Dispersion}_{i,t} = \beta_0 + \beta_1 \textit{CBS-Firm}_{i,t} + \sum \beta_k \text{Controls} + \sum \beta_j \text{Fixed Effects}, \quad (3)$$

where for firm *i* in year *t*, the dependent variable is either *Coverage*, the natural logarithm of the number of analysts following a firm, *Accuracy*, the absolute value of the mean forecast error before the earnings announcement scaled by lag price and multiplied by -100, or *Dispersion*, the standard deviation of analysts' mean forecast error scaled by price and multiplied by 100. The main explanatory variable is the average CBS at the firm-year level computed as the average of the CBS across all concepts and firms within the same industry (see Appendix C for details). We control for firm characteristics shown to affect coverage and the properties of analysts' forecasts (De Franco et al., 2011): *Size* (log market value), *Loss* (*NI*), whether the firm's earnings are below the reported earnings a year ago (*Neg. Earnings*), unexpected earnings (*Unexp. Earnings*), earnings and return volatility, and the general level of accounting language similarity by including the cosine similarity. We also include firm and year fixed effects and cluster our standard errors at the firm level. If unstandardized terminology increases financial analysts' information processing costs, we expect to observe a positive association between our measure of standardization and coverage and accuracy. Conversely, the CBS should be negatively associated with dispersion.

We then exploit the granularity of individual analysts' forecasts in the I/B/E/S *detail file* to further explore whether (i) standardization is more beneficial for less experienced analysts (those

[27] We follow the procedure illustrated as Method 3 in the dedicated WRDS guide to make the adjustments (https://wrds-www.wharton.upenn.edu/documents/5/A_Note_on_IBES_Unadjusted_Data_pdf.pdf). We rely on Datastream instead of CRSP as our sample encompasses international firms.

who have been following the firm for a shorter period) and (ii) the standardization that matters to an individual analyst is the one toward that person's own reference group. In this setting, we posit that an individual financial analyst would benefit more from a focal firm standardizing its terminology toward the other firms *within that analyst's portfolio* (rather than *all* other firms in the industry). Here, we fully leverage the granularity of the analysts' data and calculate a version of our CBS (and cosine) for each analyst-firm-year combination at the individual analyst's portfolio level, i.e., employing as a peer group all the other firms within the focal analyst's portfolio (*CBS-Firm-Analyst*). This approach allows us to estimate a model with the strictest fixed effects possible (i.e., firm-year fixed effects), basically enabling us to control for any time-variant characteristic of a firm. In this second set of tests, we predict our CBS-Firm-Analyst measure to be positively associated with accuracy.

$$\text{Accuracy}_{i,t,a} = \beta_0 + \beta_1 \textit{CBS-Firm-Analyst}_{i,t,a} + \sum \beta_k \text{Controls} + \sum \beta_j \text{Fixed Effects}, \quad (4)$$

where for each firm *i*, year *t*, and analyst *a* observation, the dependent variable is *Accuracy*, the absolute value of the analyst's mean forecast error before the earnings announcement scaled by lag price and multiplied by -100. The main explanatory variable is the average CBS at the firm-year-analyst level. We control for firm characteristics that may affect the properties of analysts' forecasts. We also include firm, year, and analyst fixed effects and cluster standard errors at the firm level. To assess whether our results reflect within-analyst portfolio effects, we create a placebo version of the CBS-Firm-Analyst measure. Specifically, for each firm-year-analyst observation, we randomly assign the CBS-Firm-Analyst value computed on the basis of a different analyst's portfolio (covering the same focal firm in the same year) in place of the actual analyst's portfolio-based measure.

Table 11 reports the descriptive statistics for all the variables used in the financial analysts' tests. Panels A and B provide statistics for the summary and detail files for our main sample and top-down (bottom-up) lists, respectively. On average, seven analysts cover our sample firms on a firm-year basis. The mean forecast accuracy for our main sample is 3.79 percent of the share price, and the mean dispersion is 4.90 percent of the share price. The average CBS-Firm is on average 64 percent. Table 12 Panel A Columns 1 to 3 (4 to 6) display the coefficients for estimating Equation 3 for our top-down (bottom-up) lists. The average CBS for a firm is positively associated with analysts' coverage and forecast accuracy and negatively associated with their dispersion. These associations are significant across all specifications. Specifically, these coefficients imply that a one standard deviation change in the *CBS-Firm* increases Coverage by 1.4 (0.8) percent, increases accuracy by 0.28 (0.19) percentage points, and decreases dispersion by 0.19 (0.16) percentage points, depending on whether we use the top-down (bottom-up) list to compute our standardization measure.

Table 12 Panel B presents the results that leverage the detail file. Columns 1 and 2 (5 and 6) display the coefficients for estimating Equation 3 separately for more versus less experienced analysts, where the CBS is computed using our top-down (bottom-up) list. The magnitude and statistical significance of the CBS coefficients indicate that standardization is particularly beneficial for analysts who are newer to the firm and therefore have higher information processing costs. Columns 3 and 4 (7 and 8) report the results for estimating Equation 4 at the individual analyst level, comparing the placebo *CBS-Firm-Analyst* measure ("outside" portfolio) with the true within-portfolio CBS. We observe a larger and statistically significant coefficient for the CBS computed within the analysts' portfolio, while the placebo measure shows no significant association. Taken together, our findings consistently show that terminology standardization is associated with higher coverage and improved analysts' forecast properties, consistent with the idea that it reduces information processing costs. Our granular evidence indicates that these benefits are

driven primarily by less experienced analysts and by the extent of standardization within their portfolios, highlighting that the benefits of standardization depend on users' expertise and reference group.

5. Conclusions

Ninety years after Preinreich's seminal address lauding the "benefits from a standardization of accounting terminology," we provide the first evidence on the magnitude of unstandardized accounting terminology in reporting practice and its implications for users' information processing costs. We document considerable variation in firms' usage of accounting terms that relate to the same accounting concept, i.e., in synonyms or differing labels in higher n-gram terms. This variation has economic consequences for users; i.e., they need to understand highly technical and complex terms *and* cope with meaningful ambiguities about whether, in a given context, differing terms capture the same economic transactions (similar to the challenges we faced as researchers in dealing with and defending our concept lists). These costs add up when processing accounting information. We show that variation in technical terms hinders even the most sophisticated users studied in the literature (data, AI, and financial analysts).

While our study has clear messages for regulators, standard setters, and reporting entities about the consequences of word choices in disclosure regulation and reporting practice, it does not speak to the costs of standardization. We do not expect that standardizing terms would lead to information loss per se (i.e., not accurately reflect the economics), but we recognize that it will generate adjustment costs for preparers and users unaccustomed to a standard term or in settings where there is less institutional fit. Therefore, as with any language, the "language of business" continues to evolve, and standardization in accounting terminology continues to remain an ideal rather than a reality.

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... and McGrew, B. (2023). Gpt-4 technical report. Retrieved from https://arxiv.org/pdf/2303.08774.

AICPA (1953). Review and Résumé. *Accounting Terminology Bulletin, 1*.

Akbas, F., Markov, S., Subasi, M., and Weisbrod, E. (2018). Determinants and consequences of information processing delay: Evidence from the Thomson Reuters institutional brokers' estimate system. *Journal of Financial Economics, 127*(2), 366-388.

Allee, K. D., DeAngelis, M. D., and Moon Jr, J. R. (2018). Disclosure "scriptability". *Journal of Accounting Research*, *56*(2), 363-430.

Becker, K., Bischof, J., and Daske, H. (2021). IFRS: Markets, practice, and politics. *Foundations and Trends® in Accounting*, *15*(1–2), 1-262.

Belkaoui, A. (1978). Linguistic relativity in accounting. *Accounting, Organizations and Society, 3*(2), 97.

Bernard, D., Blankespoor, E., de Kok, T., and Toynbee, S. (2024). Using GPT models to measure the complexity of business transactions. *Available at SSRN 4480309*.

Blankespoor, E., deHaan, E., and Marinovic, I. (2020). Disclosure processing costs, investors' information choice, and equity market outcomes: A review. *Journal of Accounting and Economics, 70*(2-3), 101344.

Blankespoor, E., deHaan, E., Wertz, J., and Zhu, C. (2019). Why do individual investors disregard accounting information? The roles of information awareness and acquisition costs. *Journal of Accounting Research, 57*(1), 53-84.

Blankespoor, E., Miller, B. P., and White, H. D. (2014). Initial evidence on the market impact of the XBRL mandate. *Review of Accounting Studies, 19*(4), 1468-1503.

Bonsall IV, S. B., Leone, A. J., Miller, B. P., and Rennekamp, K. (2017). A plain English measure of financial reporting readability. *Journal of Accounting and Economics, 63*(2-3), 329-357.

Boritz, J. E., and No, W. G. (2020). How significant are the differences in financial data provided by key data sources? A comparison of XBRL, Compustat, Yahoo! Finance, and Google Finance. *Journal of Information Systems*, *34*(3), 47-75.

Bowers, R. (1944). Terminology and form of the income sheet. *The Accounting Review, 19*(3), 274-279.

Brown, L. D., Call, A. C., Clement, M. B., and Sharp, N. Y. (2015). Inside the "black box" of sell-side financial analysts. *Journal of Accounting Research, 53(1)*, 1-47.

Cabré Castellví, M. T. (1999). In Sager J. C. (Ed.), *Terminology*. John Benjamins Publishing Company, Amsterdam.

CFA Institute (2016). CFA institute member survey: XBRL (extensible business reporting language). Retrieved from https://www.cfainstitute.org/en/research/survey-reports/xbrl-member-survey-report-2016.

Challapally, A., Pease, C., Raskar, R., and Chari, P. (2025). The genai divide: State of ai in business 2025. Technical report, Massachusetts Institute of Technology (MIT), NANDA Initiative. Retrieved from http://artificialintelligence-news.com/ wp-content/uploads/2025/08/ai_report_2025.pdf.

Chen, S., Miao, B., and Shevlin, T. (2015). A new measure of disclosure quality: The level of disaggregation of accounting data in annual reports. *Journal of Accounting Research*, 53(5), 1017-1054.

Chychyla, R., and Kogan, A. (2015). Using XBRL to conduct a large-scale study of discrepancies between the accounting numbers in Compustat and SEC 10-K filings. *The Journal of Information Systems, 29*(1), 37-72.

da Silva, A. S. (2010). Measuring and parameterizing lexical convergence and divergence between european and brazilian portuguese. *Advances in Cognitive Sociolinguistics* (pp. 41-84). De Gruyter Mouton, Berlin, New York.

De Franco, G., Kothari, S. P., and Verdi, R. (2011). The benefits of financial statement comparability. *Journal of Accounting Research, 49*(4), 895-931.

de Kok, T. (2025). ChatGPT for textual analysis? How to use generative LLMs in accounting research. *Management Science*.

Dyer, T., Lang, M., and 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.

Du, K., Huddart, S., and Jiang, X. D. (2022). Lost in Standardization: Effects of Financial Statement Database Discrepancies on Inference. *Journal of Accounting and Economics*, 101573.

Egozi, O., Markovitch, S., and Gabrilovich, E. (2011). Concept-based information retrieval using explicit semantic analysis. ACM Transactions on Information Systems (TOIS), 29(2), 1-34.

ESMA (2013). Review of accounting practices comparability of IFRS financial statements of financial institutions in Europe. Retrieved from https://www.esma.europa.eu/document/review-accounting-practices-comparability-ifrs-financial-statements-financial-institutions.

Felber, H. (1980). International Standardization of Terminology: Theoretical and Methodological Aspects, *1980*(23), 65-80.

Fjeld, E. I. (1936). Classification and terminology of individual balance-sheet items. *The Accounting Review, 11*(4), 330-345.

Fuertes-Olivera, P. A., and Nielsen, S. (2011). The dynamics of terms in accounting: What the construction of the accounting dictionaries reveals about metaphorical terms in culture-bound subject fields. *Terminology, 17*(1), 157-180.

Geeraerts, D., Speelman, D., Heylen, K., Montes, M., De Pascale, S., Franco, K., and Lang, M. (2024). *Lexical variation and change: A distributional semantic approach*. Oxford University Press.

Haried, A. A. (1972). The semantic dimensions of financial statements. *Journal of Accounting Research, 10*(2), 376-391.

Hoitash, R., Hoitash, U., and Morris, L. (2021). eXtensible business reporting language (XBRL): A review and implications for future research. *Auditing: A Journal of Practice and Theory, 40*(2), 107-132.

Hoogervorst, H. (2019). South Korea's contribution to international standard-setting. Retrieved from https://www.ifrs.org/news-and-events/news/2019/09/koreas-contribution-to-international-standard-setting/.

IASB (2015). *APF11: Principles of disclosures*. Retrieved from https://www.ifrs.org/content/dam/ifrs/meetings/2015/june/iasb/disclosure-initiative/ap11f-principles-diclosure-comparability.pdf.

IASB (2019). Using the IFRS Taxonomy. A preparer's guide. https://www.ifrs.org/content/dam/ifrs/resources-for/preparers/xbrl-using-the-ifrs-taxonomy-a-preparers-guide-january-2019.pdf?la=en.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... and Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13), 3521-3526.

Kohler, E. L. (1935). Some principles for terminologists. *The Accounting Review*, *10*(1)*,* 31-33.

Koonce, L., Lipe, M. G., and McAnally, M. L. (2005). Judging the risk of financial instruments: Problems and potential remedies. *The Accounting Review*, *80*(3), 871-895.

Lang, M., and Stice-Lawrence, L. (2015). Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics, 60*(2-3), 110-135.

Lehavy, R., Li, F., and Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review*, 86(3), 1087-1115.

Levy, B. (2025). Caution Ahead: Numerical Reasoning and Look-ahead Bias in AI Models. *Working Paper.*

Levy, B., and Wang, S. (2024). BeanCounter: A low-toxicity, large-scale, and open dataset of business-oriented text. arXiv preprint arXiv:2409.17827.

Loughran, T., and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), 35-65.

Loughran, T., and McDonald, B. (2014). Measuring readability in financial disclosures. *The Journal of Finance, 69*(4), 1643-1671.

Loughran, T., and McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research, 54*(4), 1187-1230.

Luo, Y., Yang, Z., Meng, F., Li, Y., Zhou, J., & Zhang, Y. (2025). An empirical study of catastrophic forgetting in large language models during continual fine-tuning. *IEEE Transactions on Audio, Speech and Language Processing.*

Mills, P. A. (1989). Words and the study of accounting history. *Accounting, Auditing and Accountability Journal, 2*(1), 21-35.

Nobes, C., and Stadler, C. (2018). Impaired translations: IFRS from English and annual reports into English. *Accounting, Auditing and Accountability Journal, 31*(7), 1981-2005.

Patwardhan, T., Dias, R., Proehl, E., Kim, G., Wang, M., Watkins, O., Fishman, S.P., Aljubeh, M., Thacker, P., Fauconnet, L., Kim, N.S., Chao, P., Miserendino, S., Chabot, G., Li, D., Sharman, M., Barr, A., Glaese, A., & Tworek, J. (2025). GDPval: Evaluating AI Model Performance on Real-World Economically Valuable Tasks. Retrieved from https://arxiv.org/pdf/2510.04374

Payne, J. L., and Thomas, W. B. (2003). The implications of using stock-split adjusted I/B/E/S data in empirical research. *The Accounting Review*, *78*(4), 1049-1067.

Peterson, K., Schmardebeck, R., and Wilks, T. J. (2015). The earnings quality and information processing effects of accounting consistency. *The Accounting Review*, *90*(6), 2483-2514.

Preinreich, G. A. D. (1933). Accounting terminology. *The Accounting Review, 8*(2), 113-116.

Refinitiv (2020). Refinitiv Worldscope Fundamentals Database. Data Definitions Guide (Issue 16).

Riahi-Belkaoui, A. (1995). *The linguistic shaping of accounting.* Quorum Books, Westport.

Ruette, T., Speelman, D., and Geeraerts, D. (2014). Lexical variation in aggregate perspective. *Pluricentricity: Language variation and sociocognitive dimensions*, 103-126.

Schaub, N. (2018). The role of data providers as information intermediaries. *Journal of Financial and Quantitative Analysis, 53*(4), 1805-1838.

Shaffer, M., and Wang, C. C. (2025). Scaling Core Earnings Measurement with Large Language Models. Available at SSRN.

Silhan, P. A. (1978). The recurring problem of divergent terminology. *The Accounting Review, 53*(1), 179-181.

Singer, F. A. (1960). Needed: A glossary to accompany audit reports. *The Accounting Review, 35*(1), 90.

Speelman, D., Grondelaers, S., and Geeraerts, D. (2003). Profile-based linguistic uniformity as a generic method for comparing language varieties. *Computers and the Humanities, 37*(3), 317-337.

Tarca, A. (2020). *Digital reporting—questions for practitioners, standard-setters and researchers*. Retrieved from https://www.ifrs.org/news-and-events/news/2020/07/digital-reporting-questions/.

Temmerman, R. (2000). *Towards new ways of terminology description*. John Benjamins Publishing Company, Philadelphia.

Tenner, I. (1941). Difficulties of a terminologist. *The Accounting Review, 16*(4), 349-358.

Thießen, F., D'Souza, J., and Stocker, M. (2023). Probing Large Language Models for Scientific Synonyms. In SEMANTICS Workshops.

Wermuth, M. C., and Verplaetse, H. (2019). Medical terminology in the Western world. *Handbook of terminology*, 83-108.

Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., ... and Mann, G. (2023). Bloomberggpt: A large language model for finance. arXiv preprint arXiv:2303.17564.

TABLE 1

Lab Experiment

Panel A: All Statements					
	More Standardized	**Less Standardized**	**Difference**	**t-stat**	**p-value**
Mean Time	12.081	16.009	-3.929***	-7.461	0.000
Std	12.159	16.073			
Obs	1,440	1,480			
Mean Error	0.106	0.270	-0.165***	-11.677	0.000
Std	0.307	0.444			
Obs	1,440	1,480			
Panel B: Income Statement					
	More Standardized	**Less Standardized**	**Difference**	**t-stat**	**p-value**
Mean Time	16.440	20.200	-3.760***	-2.991	0.003
Std	17.553	21.062			
Obs	468	481			
Mean Error	0.152	0.301	-0.150***	-5.603	0.000
Std	0.359	0.459			
Obs	468	481			
Panel C: Balance Sheet - Asset Side					
	More Standardized	**Less Standardized**	**Difference**	**t-stat**	**p-value**
Mean Time	8.401	10.892	-2.490***	-4.481	0.000
Std	5.843	8.900			
Obs	360	370			
Mean Error	0.061	0.162	-0.101***	-4.398	0.000
Std	0.240	0.369			
Obs	360	370			
Panel D: Balance Sheet - Liabilities and Equity Side					
	More Standardized	**Less Standardized**	**Difference**	**t-stat**	**p-value**
Mean Time	10.912	15.815	-4.904***	-7.561	0.000
Std	8.292	13.925			
Obs	612	629			
Mean Error	0.096	0.310	-0.214***	-9.718	0.000
StdDev	0.295	0.463			
Obs	612	629			
Panel E: All Statements – Excluding terms corresponding to data item names					
	More Standardized	**Less Standardized**	**Difference**	**t-stat**	**p-value**
Mean Time	12.434	16.233	-3.799***	-6.735	0.000
StdDev	12.73	16.199			
Obs	1,188	1,443			
Mean Error	0.114	0.275	-0.161***	-10.741	0.000
StdDev	0.319	0.447			
Obs	1,188	1,443			
Panel F: Within Student					
	More Standardized	**Less Standardized**	**Difference**	**t-stat**	**p-value**
Mean Time	12.192	16.204	-4.013***	-5.121	0.000
StdDev	4.186	4.632			
Obs	57	57			
Mean Error	0.105	0.245	-0.140***	-5.865	0.000
StdDev	0.147	0.167			
Obs	57	57			

Table 1 reports the results of the experiment described in Section 2. Panel A displays results for all financial statements, Panel B, C, and D report them separately for the Income Statement, the Asset Side, and the Liabilities and Equity Side. Panel E replicates the results in Panel A, excluding those observations where the data item name corresponds to the term displayed in the financial statement. Panel F presents the results for a within-student specification. *Mean Time* is the average time in seconds a participant spent on an individual question. *Mean Error* represents the average error rate for an individual question.

TABLE 2

Accounting Thesaurus: Term Lists

Panel A: Top-Down								
Source	**Terms**	**Avg. N-Gram**	**1-Gram**	**2-Gram**	**3-Gram**	**4-Gram**	**5-Gram**	**> 5-Gram**
Dictionaries	20,492	3.42	214	7,252	5,515	3,247	1,938	2,326
Panel B: Bottom-Up								
Source	**Terms**	**Avg. N-Gram**	**1-Gram**	**2-Gram**	**3-Gram**	**4-Gram**	**5-Gram**	**> 5-Gram**
10-Ks	43,484	4.89	97	4,306	7,357	8,905	7,891	14,928
20-Fs	19,615	4.69	72	2,682	3,650	3,973	3,143	6,095

Table 2 reports statistics on our term lists. Panel A and B present summary statistics for our top-down and bottom-up term lists, including the number of terms, average n-gram length, and frequency distribution across n-gram sizes.

TABLE 3

Accounting Thesaurus: Concept Lists

Panel A: Top-Down

			Concept Characteristics														
	Concept Statistics		Size					Textual Similarity					Concentration				
Source	**Concepts**	**Synonyms**	**Mean**	**Std**	**Q1**	**Median**	**Q3**	**Mean**	**Std**	**Q1**	**Median**	**Q3**	**Mean**	**Std**	**Q1**	**Median**	**Q3**
Dictionaries	964	5,166	5.36	2.71	4.00	5	6	0.78	0.11	0.71	0.79	0.85	0.58	0.23	0.39	0.54	0.77

Panel B: Bottom-Up

			Concept Characteristics														
	Concept Statistics		Size					Textual Similarity					Concentration				
Source	**Concepts**	**Synonyms**	**Mean**	**Std**	**Q1**	**Median**	**Q3**	**Mean**	**Std**	**Q1**	**Median**	**Q3**	**Mean**	**Std**	**Q1**	**Median**	**Q3**
10-Ks	4,166	50,021	12.01	18.78	3.00	5	12	0.86	0.08	0.82	0.87	0.92	0.54	0.27	0.32	0.52	0.77
20-Fs	1,989	21,464	10.79	12.21	3.00	7	13	0.82	0.09	0.77	0.83	0.88	0.56	0.26	0.34	0.52	0.79

Table 3 reports statistics on our concept lists. Panel A and B present summary statistics for our top-down and bottom-up concept lists, including the number of concepts, synonyms, and descriptives on concepts' size, textual similarity within a concept as proxied by the cosine similarity of the synonyms within the concept, and concentration as proxied by the Herfindahl-Hirschman Index computed using the frequency distribution of the terms within a concept.

TABLE 4

Annual Reports: Sample Selection and Composition

Panel A: Sample Selection (Non-US Firms)		
	Obs	**Firms**
Non-US PDF annual reports in Perfect Information	**246,394**	**29,174**
Less: Reports outside 2000-2019	(8,347)	(199)
Less: No identifier in Worldscope	(7,330)	0
Less: Reports with less than 1st percentile unique accounting terms	(3,010)	(163)
Less: Countries with less than 100 reports	(1,185)	(281)
Non-US Sample	**226,522**	**28,531**
Panel B: Sample Selection (US Firms)		
EDGAR SEC 10-K Universe	**197,283**	**14,243**
Less: Observations outside 2000-2019	(41,706)	(1,824)
Less: No Compustat match on filing date	(59,523)	(454)
Less: Reports with less than 1st percentile unique accounting terms	(958)	(20)
US Sample	**95,096**	**11,945**
Panel C: Sample Selection (Non-US cross-listed Firms)		
EDGAR SEC 20-F Universe	**19,105**	**2,109**
Less: Observations outside 2000-2019	(4,837)	(574)
Less: No Compustat match on filing date	(7,895)	(743)
Less: Reports with less than 1st percentile unique accounting terms	(583)	(22)
Non-US cross-listed Sample	**5,790**	**770**
Panel D: Sample Selection AI Analyst (iXBRL)		
EDGAR SEC 10-K Universe	**30,556**	**6,001**
Less: No Compustat match on filing date	(8,316)	(1,012)
Less: Reports with less than 1st percentile unique accounting terms	(228)	(42)
Less: Reports with no parsed financial statement	(112)	(16)
iXBRL Sample with parsed Financial Statements (FS)	**21,900**	**4,931**
Less: Malformatted financial statements	(857)	(130)
Less: FS with duplicated tags, terms or database matches	(7,408)	(1,134)
Less: FS with less than five matches to accounting databases	(159)	(23)
Less: FS not matched to Worldscope and Compustat	(252)	(39)
Less: Reports with only a single financial statement	(10,849)	(2,704)
AI Analyst Sample	**2,375**	**901**

Table 4 provides details on our sample selection steps and sample composition as described in Section 3.2. Panel A and B report on our main sample. Panel C presents our 20-F sample of Non-US cross listed firms. Panel D reports on our iXBRL sample.

TABLE 5

Accounting Terms and Concepts within Annual Reports

Panel A: Top-Down

		Term Frequency					Unique Terms					Unique Concepts				
Source	**Obs**	**Mean**	**Std**	**Q1**	**Median**	**Q3**	**Mean**	**Std**	**Q1**	**Median**	**Q3**	**Mean**	**Std**	**Q1**	**Median**	**Q3**
Dictionaries	310,970	1,850.67	1,441.43	911	1,597	2,483	482.41	248.45	331	483	638	200.44	68.08	168	210	246

Panel B: Bottom-Up

		Term Frequency					Unique Terms					Unique Concepts				
Source	**Obs**	**Mean**	**Std**	**Q1**	**Median**	**Q3**	**Mean**	**Std**	**Q1**	**Median**	**Q3**	**Mean**	**Std**	**Q1**	**Median**	**Q3**
10-Ks	304,692	1,504.79	1,258.53	670	1,223	2,028	456.70	277.77	265	414	615	414.04	213.96	271	399	547
20-Fs	302,976	1,417.29	1,076.94	697	1,229	1,890	424.27	211.00	288	424	557	350.32	150.12	264	358	448

Table 5 reports statistics on the frequency of terms, unique terms, and unique concepts at the annual report level for our main sample and top-down and bottom-up lists, respectively.

TABLE 6

Most Common Terms and Concepts in Reporting Practice

Panel A: 25 Most Common Terms

	Top-Down (Dictionaries)			Bottom-Up (10-Ks)			Bottom-Up (20-Fs)		
Rank	**Term**	**Frequency**	**Obs**	**Term**	**Frequency**	**Obs**	**Term**	**Frequency**	**Obs**
1	Financial Statements	10,835,259	283,241	Financial Statements	13,684,688	284,211	Consolidated Financial State-ment	7,043,474	248,437
2	Annual Reports	8,814,596	272,960	Consolidated Financial Statements	7,860,468	248,654	Common Stock	6,130,434	117,321
3	Consolidated Financial Statements	7,872,624	248,637	Cash Flow	4,552,243	282,530	Recognized	6,010,578	144,215
4	Common Stock	6,262,972	117,346	Financial Assets	3,842,704	188,700	Board Of Director	5,913,246	273,993
5	Board Of Director	5,912,494	273,954	Balance Sheet	3,738,758	257,871	Cash Flow	4,483,412	282,554
6	Financial Year	4,962,129	159,206	Common Stock	3,583,063	113,562	Balance Sheet	3,734,189	259,237
7	Cash Flow	4,411,690	283,291	Fiscal Year	3,561,887	150,780	Interest Rate	3,153,620	250,226
8	Income Tax	3,619,149	262,564	Interest Rate	2,889,228	249,190	Income Tax	2,821,622	259,851
9	Fiscal Year	3,605,324	151,378	Results Of Operation	2,784,930	168,357	Per Share	2,646,455	263,589
10	Interest Rate	3,067,645	249,768	Risk Management	2,695,865	199,619	Real Estate	2,362,899	127,037
11	Financial Assets	3,015,513	188,912	Per Share	2,458,576	259,287	Financial Instrument	2,172,744	234,142
12	Results Of Operation	2,817,958	166,648	Financial Instrument	2,378,323	235,683	Financial Asset	2,159,680	161,986
13	Audit Committee	2,759,665	215,544	Income Tax	2,292,198	260,447	Risk Management	2,103,648	185,234
14	Corporate Governance	2,685,632	231,806	Ordinary Share	2,153,910	124,788	Accounting Standard	1,984,103	256,218
15	Company Act	2,417,500	113,324	Assets And Liabilities	2,085,639	267,118	Net Income	1,824,459	143,717
16	Financial Instruments	2,387,903	238,406	Cash And Cash Equivalents	1,901,378	249,057	Cash And Cash Equivalents	1,805,323	248,931
17	Cash And Cash Equivalents	2,353,404	252,787	Accounting Standards	1,900,774	255,359	Joint Venture	1,784,050	155,893
18	Balance Sheet	2,341,674	248,789	Accounting Policies	1,885,078	254,808	Assets And Liabilities	1,777,297	261,034
19	Stock Option	2,281,183	141,509	Joint Venture	1,883,061	160,980	Intangible Assets	1,684,445	195,874
20	Ordinary Shares	2,276,467	125,456	Impairment Loss	1,765,309	196,753	Natural Gas	1,681,254	43,861
21	Net Income	2,154,427	149,969	Credit Risk	1,612,344	201,202	Deferred Tax Asset	1,639,424	237,439
22	Non-Executive Director	2,135,152	125,546	Income Statement	1,601,288	138,414	Ordinary Share	1,589,687	120,575
23	Profit Or Loss	2,085,174	139,105	Deferred Tax Asset	1,553,110	234,875	Preferred Stock	1,571,458	84,742
24	Risk Management	2,035,548	186,387	Net Income	1,533,473	140,690	Accounting Policy	1,547,944	252,417
25	Intangible Asset	2,020,699	203,283	Intangible Asset	1,521,010	196,182	Impairment Loss	1,489,475	195,298

TABLE 6 (continued)

Panel B: 5 Most Common Concepts for Top-Down

	Income Statement Concepts			Balance Sheet Concepts		
Rank	**Term**	**Frequency**	**Obs**	**Term**	**Frequency**	**Obs**
1	**Concept: Results Of Operation**	**8,937,013**	**282,121**	**Concept: Common Stock**	**12,904,950**	**283,968**
	Results Of Operation	2,817,958	166,648	Common Stock	6,262,972	117,346
	Net Income	2,154,427	149,969	Ordinary Share	2,276,467	125,456
	Net Loss	1,249,233	155,683	Common Share	1,456,771	112,301
	All Other Terms (N: 16)	*2,715,395*		*All Other Terms (N: 7)*	*2,908,740*	
2	**Concept: Income Tax**	**5,107,326**	**269,218**	**Concept: Defer Tax Asset**	**4,073,562**	**260,425**
	Income Tax	3,619,149	262,564	Deferred Tax Asset	1,777,218	238,738
	Income Tax Expense	518,573	139,330	Deferred Tax	1,188,549	209,951
	Tax Expense	367,765	136,722	Deferred Tax Liability	600,603	187,949
	All Other Terms (N: 5)	*601,839*		*All Other Terms (N: 5)*	*507,192*	
3	**Concept: Net Sale**	**3,009,315**	**224,379**	**Concept: Net Asset**	**3,012,120**	**277,541**
	Net Sale	1,160,021	78,910	Net Asset	966,969	197,834
	Total Revenue	628,335	133,151	Shareholder's Equity	813,682	122,794
	Net Revenue	443,929	44,252	Stockholder's Equity	496,155	71,016
	All Other Terms (N: 8)	*777,030*		*All Other Terms (N: 17)*	*735,314*	
4	**Concept: Operating Result**	**2,980,365**	**249,753**	**Concept: Cash And Cash Equivalent**	**2,988,861**	**267,928**
	Operating Result	1,061,948	168,767	Cash And Cash Equivalent	2,353,404	252,787
	Operating Income	731,695	96,421	Bank Balance	304,319	76,942
	Operating Profit	661,738	110,291	Cash Balance	210,194	97,889
	All Other Terms (N: 7)	*524,984*		*All Other Terms (N: 3)*	*120,944*	
5	**Concept: Interest Expense**	**2,772,014**	**261,675**	**Concept: Trade Receivable**	**2,923,105**	**255,799**
	Interest Expense	1,391,583	206,863	Trade Receivable	985,272	142,429
	Financing Cost	561,768	127,009	Accounts Receivable	923,603	135,855
	Borrowing Cost	462,378	112,961	Trade And Other Receivables	491,503	90,104
	All Other Terms (N: 6)	*356,285*		*All Other Terms (N: 19)*	*522,727*	

TABLE 6 (continued)

Panel C: 5 Most Common Concepts for Bottom-Up (10-Ks)

	Income Statement Concepts			Balance Sheet Concepts		
Rank	**Term**	**Frequency**	**Obs**	**Term**	**Frequency**	**Obs**
1	**Concept: Net Income**	**4,873,178**	**275,860**	**Concept: Common Stock**	**8,208,766**	**280,768**
	Net Income	1,533,473	140,690	Common Stock	3,583,063	113,562
	Profit Loss	829,014	116,065	Ordinary Share	2,153,910	124,788
	Net Profit	739,164	129,288	Share Capital	1,397,053	162,352
	All Other Terms (N: 43)	*1,771,527*		*All Other Terms (N: 13)*	*1,074,740*	
2	**Concept: Net Sale**	**3,863,361**	**249,141**	**Concept: Net Asset**	**2,619,815**	**254,133**
	Net Sales	954,648	69,840	Net Asset	714,997	172,374
	Net Revenue	390,176	42,389	Shareholder's Equity	708,171	117,838
	Operating Revenue	259,587	34,850	Stockholder's Equity	427,350	67,827
	All Other Terms (N: 53)	*2,258,950*		*All Other Terms (N: 55)*	*769,297*	
3	**Concept: Operate Profit**	**2,705,101**	**251,644**	**Concept: Cash And Cash Equivalent**	**2,574,854**	**269,301**
	Operating Profit	638,972	107,868	Cash And Cash Equivalents	1,901,378	249,057
	Operating Income	636,095	91,554	Cash Balance	211,527	97,176
	Operating Loss	264,912	96,894	Cash And Bank Balance	142,982	45,086
	All Other Terms (N: 74)	*1,165,122*		*All Other Terms (N: 28)*	*318,967*	
4	**Concept: Income Tax Expense**	**2,614,973**	**254,386**	**Concept: Fix Asset**	**2,520,360**	**268,560**
	Income Tax Expense	500,075	140,625	Fixed Asset	1,106,211	138,569
	Tax Benefit	354,785	96,166	Plant And Equipment	350,828	107,957
	Tax Expense	329,340	127,052	Long Lived Asset	335,397	77,485
	All Other Terms (N: 139)	*1,430,773*		*All Other Terms (N: 94)*	*727,924*	
5	**Concept: Research And Development**	**1,749,532**	**161,462**	**Concept: Trade Receivable**	**2,257,187**	**239,998**
	Research And Development	766,192	115,113	Trade Receivable	1,041,473	144,318
	Product Development	313,286	87,183	Account Receivable	728,286	128,362
	Research And Development Expense	219,762	43,817	Trade Account Receivable	103,919	33,561
	All Other Terms (N: 32)	*450,292*		*All Other Terms (N: 61)*	*383,509*	

TABLE 6 (continued)

Panel D: 5 Most Common Concepts for Bottom-Up (20-Fs)

	Income Statement Concepts			Balance Sheet Concepts		
Rank	**Term**	**Frequency**	**Obs**	**Term**	**Frequency**	**Obs**
1	**Concept: Net Income**	**7,507,858**	**285,490**	**Concept: Property Plant And Equipment**	**3,933,898**	**271,636**
	Net Income	1,824,459	143,717	Property Plant And Equipment	1,353,681	156,751
	Net Loss	945,348	145,103	Fixed Asset	962,041	141,131
	Profit Or Loss	944,348	138,356	Property And Equipment	420,603	71,225
	All Other Terms (N: 66)	*3,793,703*		*All Other Terms (N: 34)*	*1,197,573*	
2	**Concept: Income Tax**	**5,969,135**	**279,178**	**Concept: Financial Position**	**3,674,205**	**282,482**
	Income Tax	2,821,622	259,851	Financial Position	1,461,080	265,329
	Tax Benefit	548,446	106,757	Shareholder's Equity	701,135	118,924
	Income Tax Expense	455,271	134,534	Company Share	503,842	148,068
	All Other Terms (N: 111)	*2,143,796*		*All Other Terms (N: 37)*	*1,008,148*	
3	**Concept: Operating Result**	**4,133,990**	**267,604**	**Concept: Term Loan**	**3,020,989**	**231,920**
	Operating Result	1,054,951	168,465	Term Loan	1,062,768	109,602
	Operating Income	617,884	90,039	Long Term Debt	791,108	102,826
	Operating Profit	531,736	93,917	Convertible Note	362,459	26,368
	All Other Terms (N: 59)	*1,929,419*		*All Other Terms (N: 68)*	*804,654*	
4	**Concept: Net Sale**	**3,258,787**	**247,636**	**Concept: Market Value**	**2,905,949**	**255,844**
	Net Sale	1,057,567	77,037	Market Value	1,362,206	221,401
	Net Revenue	431,326	43,506	Treasury Share	444,812	70,689
	Operating Revenue	273,439	35,468	Acquisition Cost	405,074	106,153
	All Other Terms (N: 37)	*1,496,455*		*All Other Terms (N: 22)*	*693,857*	
5	**Concept: Profit Or Loss**	**2,924,266**	**272,612**	**Concept: Cash And Cash Equivalent**	**2,586,396**	**267,800**
	Profit Or Loss	944,348	138,356	Cash And Cash Equivalent	1,805,323	248,931
	Taxable Income	654,578	179,822	Cash Balance	214,299	97,458
	Profit Before Tax	315,624	82,276	Cash And Bank Balance	143,545	44,972
	All Other Terms (N: 101)	*1,009,716*		*All Other Terms (N: 30)*	*423,229*	

Table 6 displays the most common terms and concepts in our main sample. Panel A presents the 25 most common terms by term list. Panels B, C, and D report the 5 most common concepts for the income statement and balance sheet by concept list.

TABLE 7

Data Analysts' Test – Descriptive Statistics

Panel A: Top-Down - Sample: Main								
Variable	**Obs**	**Mean**	**Std**	**Min**	**Q1**	**Median**	**Q3**	**Max**
Missing Item	19,606,919	0.15	0.35	0.00	0.00	0.00	0.00	1.00
CBS-Firm-Concept	19,606,919	0.60	0.27	0.00	0.40	0.62	0.83	1.00
Cosine Similarity	19,606,919	0.40	0.12	0.00	0.31	0.39	0.49	0.76
Size	19,606,919	19.09	2.44	0.18	17.31	19.07	20.85	31.92
Return on Asset	19,606,919	-0.08	0.51	-4.38	-0.03	0.02	0.06	0.32
Book-to-Market	19,606,919	0.80	1.10	-4.06	0.31	0.60	1.10	5.56
Leverage	19,606,919	0.25	0.27	0.00	0.04	0.19	0.36	1.80
Age	19,606,919	31.77	31.37	0.00	11.08	21.25	39.67	261.92
Loss (NI)	19,606,919	0.32	0.47	0.00	0.00	0.00	1.00	1.00
Panel B: Bottom-Up - Sample: Main								
Variable	**Obs**	**Mean**	**Std**	**Min**	**Q1**	**Median**	**Q3**	**Max**
Missing Item	18,436,459	0.12	0.32	0.00	0.00	0.00	0.00	1.00
CBS-Firm-Concept	18,436,459	0.44	0.27	0.00	0.23	0.43	0.65	1.00
Cosine Similarity	18,436,459	0.42	0.10	0.00	0.36	0.43	0.49	0.70
Size	18,436,459	19.13	2.44	0.18	17.35	19.11	20.88	31.92
Return on Asset	18,436,459	-0.09	0.53	-4.38	-0.03	0.02	0.06	0.32
Book-to-Market	18,436,459	0.78	1.10	-4.06	0.30	0.59	1.08	5.56
Leverage	18,436,459	0.25	0.28	0.00	0.04	0.19	0.36	1.80
Age	18,436,459	31.43	31.22	0.00	11.00	21.00	39.00	261.92
Loss (NI)	18,436,459	0.32	0.47	0.00	0.00	0.00	1.00	1.00
Panel C: Bottom-Up - Sample: Has Tag (10-Ks)								
Variable	**Obs**	**Mean**	**Std**	**Min**	**Q1**	**Median**	**Q3**	**Max**
Missing Item	1,641,194	0.04	0.19	0.00	0.00	0.00	0.00	1.00
CBS-Firm-Concept	1,641,194	0.45	0.26	0.00	0.26	0.44	0.65	1.00
Cosine Similarity	1,641,194	0.44	0.08	0.06	0.39	0.44	0.49	0.70
Size	1,641,194	20.17	2.48	2.03	18.50	20.43	21.90	27.90
Return on Asset	1,641,194	-0.18	0.74	-4.38	-0.08	0.02	0.06	0.32
Book-to-Market	1,641,194	0.43	0.91	-4.06	0.17	0.38	0.69	5.56
Leverage	1,641,194	0.30	0.34	0.00	0.04	0.23	0.41	1.80
Age	1,641,194	32.06	31.45	0.00	12.75	22.17	36.17	234.92
Loss (NI)	1,641,194	0.38	0.48	0.00	0.00	0.00	1.00	1.00
Panel D: Bottom-Up - Sample: Has Tag (20-Fs)								
Variable	**Obs**	**Mean**	**Std**	**Min**	**Q1**	**Median**	**Q3**	**Max**
Missing Item	13,535	0.08	0.27	0.00	0.00	0.00	0.00	1.00
CBS-Firm-Concept	13,535	0.42	0.27	0.00	0.21	0.40	0.61	1.00
Cosine Similarity	13,535	0.47	0.11	0.10	0.37	0.48	0.55	0.85
Size	13,535	22.15	2.56	13.21	20.46	22.71	24.08	26.35
Return on Asset	13,535	-0.07	0.40	-5.08	0.00	0.02	0.06	0.28
Book-to-Market	13,535	0.78	0.78	-1.22	0.33	0.63	1.07	4.41
Leverage	13,535	0.28	0.23	0.00	0.15	0.23	0.37	1.35
Age	13,535	49.15	41.11	1.00	20.58	32.67	68.50	241.92
Loss (NI)	13,535	0.26	0.44	0.00	0.00	0.00	1.00	1.00

Table 7 provides descriptive statistics for the variables used in the data analysts' tests reported in Table 8. Panels A to D report these statistics separately for the main sample, and has tag samples, where our textual measures are computed employing the different term lists as indicated in the Panel title. All variables are defined in Appendix B

TABLE 8

Data Analyst Test – Regression Analyses

Panel A: Baseline								
Dependent Variable:	**Missing Item**							
Term List:	**Top-Down**		**Bottom-Up**					
Sample:	**Main**		**Main (10-Ks)**		**Has Tag (10-Ks)**		**Has Tag (20-Fs)**	
Independent Variables:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
CBS-Firm-Concept	**-0.007*** (-9.164)**	**-0.008*** (-9.654)**	**-0.025*** (-29.667)**	**-0.025*** (-29.869)**	**-0.026*** (-13.820)**	**-0.026*** (-13.800)**	**-0.064*** (-3.945)**	**-0.064*** (-3.934)**
Cosine Similarity	0.018*** (5.331)		0.022*** (6.920)		-0.040* (-1.862)		-0.035 (-0.592)	
Size	-0.006*** (-27.405)		-0.005*** (-24.644)		-0.004*** (-4.237)		0.000 (-0.074)	
Observations	19,606,919	19,606,919	18,436,459	18,436,459	1,641,194	1,641,194	13,535	13,535
Adj. R-Squared	0.370	0.372	0.293	0.296	0.292	0.310	0.439	0.434
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Firm FE	Y	N	Y	N	Y	N	Y	N
Year FE	Y	N	Y	N	Y	N	Y	N
Firm-Year FE	N	Y	N	Y	N	Y	N	Y
Database FE	Y	Y	Y	Y	Y	Y	Y	Y
Concept FE	Y	Y	Y	Y	Y	Y	Y	Y
Clustering	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Number of Concepts	231	231	132	132	132	132	84	84
Number Firm-Years	252,449	252,449	252,720	252,720	31,985	31,985	329	329

Panel B: Concepts Characteristics

Dependent Variable:	Missing Item											
Term List:	Top-Down						Bottom-Up (10-Ks)					
Split:	Number of Synonyms		Textual Similarity		Dominant Term		Number of Synonyms		Textual Similarity		Dominant Term	
	Few	Many	High	Low	Yes	No	Few	Many	High	Low	Yes	No
Independent Variables:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
CBS-Firm-Concept	**0.033*** (15.362)**	**-0.017*** (-19.219)**	**-0.003** (-2.491)**	**-0.009*** (-7.821)**	**0.018*** (11.756)**	**-0.008*** (-8.716)**	**-0.002 (-0.603)**	**-0.026*** (-31.591)**	**-0.011*** (-7.564)**	**-0.033*** (-31.545)**	**-0.030*** (-15.671)**	**-0.023*** (-24.381)**
Cosine Similarity	0.018*** (3.202)	0.019*** (5.341)	0.042*** (10.461)	-0.004 (-0.933)	0.049*** (10.392)	0.002 (0.534)	-0.005 (-1.159)	0.022*** (7.040)	0.021*** (5.151)	0.022*** (6.478)	0.040*** (8.929)	0.018*** (5.215)
Size	-0.005*** (-15.657)	-0.007*** (-27.940)	-0.006*** (-25.311)	-0.006*** (-24.320)	-0.008*** (-26.715)	-0.006*** (-23.685)	-0.002*** (-4.506)	-0.006*** (-24.846)	-0.006*** (-20.555)	-0.005*** (-22.681)	-0.003*** (-9.776)	-0.006*** (-25.809)
Difference (P-Value)	0.050*** (P < 0.01)		0.006*** (P < 0.01)		0.027*** (P < 0.01)		0.024*** (P < 0.01)		0.023*** (P < 0.01)		-0.007*** (P < 0.01)	
Observations	3,840,966	15,766,518	9,650,123	9,957,396	6,385,700	13,221,841	803,902	17,633,899	5,673,421	12,764,776	3,295,079	15,143,103
Adj. R-Squared	0.538	0.273	0.417	0.340	0.510	0.256	0.812	0.289	0.291	0.302	0.276	0.300
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Firm-Year FE	N	N	N	N	N	N	N	N	N	N	N	N
Database FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Concept FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Number of Concepts	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Clustering	97	134	157	74	99	132	10	122	52	80	27	105
Number of Firm-Years	251,481	253,092	252,001	252,898	251,675	253,008	222,894	254,533	251,246	254,299	249,282	254,509

Table 8 reports the results of estimating Equation 1 for the association between terminology standardization and information processing costs for data analysts working at commercial providers as described in Section 4.2.1. The outcome variable is *Missing Item*. *CBS-Firm-Concept* is the concept-firm-year level concept-based standardization for each concept used in this analysis. Controls include *Cosine Similarity*, which measures the textual similarity of the annual report to all other firms in the same industry and year, *Size* is proxied by Log Market Value, and other controls include ROA, BTM, Leverage, Age, LossNI. Panel A presents the Baseline results. In Panel B, we provide some cross-sectional evidence based on Concept Characteristics. All variables are defined in Appendix B. Columns are estimated on different subsamples as indicated at the top of each Column. T-statistics reported in parentheses are based on standard errors clustered at the firm level. *, ** and *** indicate significance: * at the 10% level, ** at the 5% level and *** at the 1% level respectively.

TABLE 9

GPT Test – Descriptive Statistics

Panel A: Error Rate by Prompt Type								
Variable	**Obs**	**Mean**	**Std**	**Min**	**Q1**	**Median**	**Q3**	**Max**
Errors - Matching Scope: Select								
Items Only								
Model: GPT-4o-Mini	108,732	0.01	0.11	0.00	0.00	0.00	0.00	1.00
Model: GPT-5-Mini	108,732	0.02	0.15	0.00	0.00	0.00	0.00	1.00
Items with Financial Statement								
Model: GPT-4o-Mini	108,732	0.02	0.14	0.00	0.00	0.00	0.00	1.00
Model: GPT-5-Mini	108,732	0.01	0.10	0.00	0.00	0.00	0.00	1.00
Financial Statement Only								
Model: GPT-4o-Mini	108,732	0.12	0.32	0.00	0.00	0.00	0.00	1.00
Model: GPT-5-Mini	108,732	0.03	0.16	0.00	0.00	0.00	0.00	1.00
Errors - Matching Scope: Full								
Items Only								
Model: GPT-4o-Mini	108,732	0.11	0.31	0.00	0.00	0.00	0.00	1.00
Model: GPT-5-Mini	108,732	0.12	0.33	0.00	0.00	0.00	0.00	1.00
Items with Financial Statement								
Model: GPT-4o-Mini	108,732	0.11	0.32	0.00	0.00	0.00	0.00	1.00
Model: GPT-5-Mini	108,732	0.11	0.31	0.00	0.00	0.00	0.00	1.00
Financial Statement Only								
Model: GPT-4o-Mini	108,732	0.33	0.47	0.00	0.00	0.00	1.00	1.00
Model: GPT-5-Mini	108,732	0.34	0.47	0.00	0.00	0.00	1.00	1.00
Panel B: Main Independent and Control Variables								
CBS-Firm-Concept	108,732	0.40	0.34	0.00	0.07	0.35	0.76	1.00
Relative Frequency	108,732	0.37	0.33	0.00	0.06	0.29	0.73	1.00
Embedding Similarity	108,732	0.67	0.09	0.13	0.63	0.68	0.73	0.86
GPT Masking Prediction	108,732	0.87	0.14	0.00	0.85	0.90	0.95	1.00
Cosine Similarity	108,732	0.18	0.08	0.00	0.12	0.19	0.25	0.33
Pair Similarity	108,732	0.81	0.17	0.00	0.75	0.87	0.94	1.00
N (Data Items)	108,732	13.51	5.09	5.00	9.00	14.00	18.00	28.00

Table 9 reports the descriptive statistics for the analyses described in Section 4.2.2 and reported in Table 10.

TABLE 10

GPT Test – Regression Analyses

Panel A: Select Matching - Items Only								
Dependent Variable:	**Error**							
GPT Model:	**4o-mini (Temperature: 1)**				**5-mini (Temperature: 1)**			
Standardization Proxy:	CBS Firm-Concept	Relative Frequency	Embedding Similarity	GPT Masking Prediction	CBS Firm-Concept	Relative Frequency	Embedding Similarity	GPT Masking Prediction
Independent Variables:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Standardization	**-0.017*** (-4.936)**	**-0.019*** (-5.015)**	**-0.256*** (-8.030)**	**-0.021*** (-3.660)**	**-0.019*** (-5.388)**	**-0.021*** (-5.165)**	**-0.286*** (-9.249)**	**-0.010*** (-5.144)**
Cosine Similarity	0.037 (1.601)	0.033 (1.425)	0.037 (1.585)	0.016 (0.671)	-0.043 (-1.517)	-0.048* (-1.725)	-0.044 (-1.592)	-0.068** (-2.405)
Pair Similarity	-0.028*** (-5.520)	-0.025*** (-4.834)	-0.024*** (-4.942)	-0.031*** (-5.904)	-0.024*** (-4.160)	-0.019*** (-3.415)	-0.019*** (-3.532)	-0.027*** (-4.680)
N (Data Items)	0.001*** (3.428)	0.001*** (3.480)	0.001*** (3.768)	0.001*** (3.771)	0.003*** (6.401)	0.003*** (6.453)	0.003*** (6.699)	0.003*** (6.640)
Observations	108,732	108,732	108,732	108,732	108,732	108,732	108,732	108,732
Adj. R-Squared	0.080	0.080	0.086	0.080	0.062	0.062	0.066	0.062
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Database FE	Y	Y	Y	Y	Y	Y	Y	Y
Concept FE	Y	Y	Y	Y	Y	Y	Y	Y
Clustering	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Number of Firm-Years	2,375	2,375	2,375	2,375	2,375	2,375	2,375	2,375

TABLE 10 (Continued)

Panel B: Select Matching - Items with Financial Statement								
Dependent Variable:	**Error**							
GPT Model:	**4o-mini (Temperature: 1)**				**5-mini (Temperature: 1)**			
Standardization Proxy:	CBS Firm-Concept	Relative Frequency	Embedding Similarity	GPT Masking Prediction	CBS Firm-Concept	Relative Frequency	Embedding Similarity	GPT Masking Prediction
Independent Variables:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Standardization	**-0.017*** (-4.624)**	**-0.020*** (-5.031)**	**-0.223*** (-7.603)**	**-0.009 (-1.613)**	**-0.017*** (-5.795)**	**-0.018*** (-5.318)**	**-0.186*** (-7.505)**	**-0.008*** (-5.640)**
Cosine Similarity	0.015 (0.534)	0.012 (0.439)	0.012 (0.415)	-0.008 (-0.289)	0.030* (1.679)	0.025 (1.388)	0.023 (1.262)	0.008 (0.457)
Pair Similarity	-0.027*** (-4.880)	-0.023*** (-4.095)	-0.023*** (-4.369)	-0.030*** (-5.296)	-0.012*** (-3.496)	-0.009** (-2.458)	-0.010*** (-2.979)	-0.015*** (-4.288)
N (Data Items)	0.001*** (3.393)	0.001*** (3.422)	0.001*** (3.662)	0.001*** (3.609)	0.001*** (2.705)	0.001*** (2.770)	0.001*** (3.078)	0.001*** (3.030)
Observations	108,732	108,732	108,732	108,732	108,732	108,732	108,732	108,732
Adj. R-Squared	0.070	0.070	0.073	0.069	0.066	0.066	0.070	0.066
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Database FE	Y	Y	Y	Y	Y	Y	Y	Y
Concept FE	Y	Y	Y	Y	Y	Y	Y	Y
Clustering	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Number of Firm-Years	2,375	2,375	2,375	2,375	2,375	2,375	2,375	2,375

TABLE 10 (Continued)

Panel C: Select Matching - Financial Statement Only								
Dependent Variable:	**Error**							
GPT Model:	**4o-mini (Temperature: 1)**				**5-mini (Temperature: 1)**			
Standardization Proxy:	CBS Firm-Concept	Relative Frequency	Embedding Similarity	GPT Masking Prediction	CBS Firm-Concept	Relative Frequency	Embedding Similarity	GPT Masking Prediction
Independent Variables:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Standardization	**-0.084*** (-10.256)**	**-0.096*** (-10.174)**	**-0.793*** (-15.666)**	**-0.060*** (-5.752)**	**-0.047*** (-9.805)**	**-0.050*** (-9.011)**	**-0.441*** (-10.412)**	**-0.017*** (-6.961)**
Cosine Similarity	-0.170*** (-2.916)	-0.189*** (-3.265)	-0.219*** (-3.854)	-0.285*** (-4.990)	0.061* (1.771)	0.046 (1.346)	0.034 (0.972)	-0.001 (-0.035)
Pair Similarity	-0.204*** (-13.120)	-0.184*** (-11.926)	-0.195*** (-13.496)	-0.217*** (-14.544)	-0.047*** (-6.596)	-0.037*** (-5.230)	-0.042*** (-6.178)	-0.055*** (-7.893)
N (Data Items)	0.008*** (9.425)	0.008*** (9.542)	0.008*** (10.166)	0.008*** (10.142)	0.000 (-0.021)	0.000 (0.100)	0.000 (0.577)	0.000 (0.515)
Observations	108,732	108,732	108,732	108,732	108,732	108,732	108,732	108,732
Adj. R-Squared	0.183	0.183	0.190	0.181	0.178	0.177	0.186	0.176
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Database FE	Y	Y	Y	Y	Y	Y	Y	Y
Concept FE	Y	Y	Y	Y	Y	Y	Y	Y
Clustering	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Number of Firm-Years	2,375	2,375	2,375	2,375	2,375	2,375	2,375	2,375

Table 10 reports the results of estimating Equation 3 for the association between terminology standardization and information processing costs for AI analysts as described in Section 4.2.2. The outcome variable is *Error* as indicated at the top of the Columns. *Standardization* is one of the standardization proxies as indicated at the top of each Column. Controls include *Cosine Similarity*, *Pair Similarity*, and *N (Data Items).* All variables are defined in Appendix B. T-statistics reported in parentheses are based on robust standard errors. *, ** and *** indicate significance: * at the 10% level, ** at the 5% level and *** at the 1% level respectively.

TABLE 11

Financial Analysts – Descriptives

Main Sample – Top-Down								
Variable	**Obs**	**Mean**	**Std**	**Min**	**Q1**	**Median**	**Q3**	**Max**
Summary Files								
Coverage	105,553	2.39	0.31	0.00	2.40	2.48	2.48	2.56
Accuracy	105,553	-3.79	10.77	-97.98	-2.73	-0.98	-0.33	-0.01
Dispersion	105,553	4.90	10.79	-68.82	3.38	6.04	8.73	36.36
CBS-Firm	105,553	0.59	0.06	0.00	0.56	0.59	0.62	1.00
Cosine Similarity	105,553	0.30	0.09	0.00	0.25	0.29	0.35	0.67
Size	105,553	20.44	1.94	7.30	19.08	20.41	21.76	27.82
Loss (NI)	105,553	0.21	0.41	0.00	0.00	0.00	0.00	1.00
Neg. Earnings	105,553	0.42	0.49	0.00	0.00	0.00	1.00	1.00
Unexp. Earnings	105,553	5.07	66.96	0.00	0.01	0.05	0.39	2,401.51
Earnings Volatility	105,553	0.06	0.19	0.00	0.01	0.02	0.05	5.08
Return Volatility	105,553	0.18	0.43	0.04	0.08	0.12	0.16	5.59
Detailed Files								
Accuracy	922,062	-2.16	4.95	-39.68	-1.88	-0.70	-0.24	0.00
CBS-Firm-Analyst (Within Portfolio)	922,062	0.70	0.07	0.00	0.67	0.71	0.74	1.00
CBS-Firm-Analyst (Outside Portfolio)	922,062	0.70	0.07	0.00	0.67	0.71	0.74	1.00
Cosine Analyst (Within Portfolio)	922,062	0.51	0.14	0.00	0.44	0.53	0.61	1.00
Cosine Analyst (Outside Portfolio)	922,062	0.51	0.14	0.00	0.44	0.53	0.61	1.00
Ln(Days)	922,062	4.51	1.10	0.00	4.01	4.72	5.31	5.90
Ln(Firms)	922,062	2.04	0.65	0.69	1.61	2.08	2.48	4.83
Main Sample – Bottom-Up								
Variable	**Obs**	**Mean**	**Std**	**Min**	**Q1**	**Median**	**Q3**	**Max**
Summary Files								
Coverage	106,143	2.39	0.31	0.00	2.40	2.48	2.48	2.56
Accuracy	106,143	-3.79	10.76	-97.98	-2.74	-0.99	-0.33	-0.01
Dispersion	106,143	4.91	10.79	-68.82	3.39	6.05	8.75	36.36
CBS-Firm	106,143	0.50	0.06	0.00	0.46	0.49	0.53	1.00
Cosine Similarity	106,143	0.33	0.10	0.00	0.27	0.33	0.39	0.66
Size	106,143	20.44	1.94	7.30	19.08	20.41	21.76	27.82
Loss (NI)	106,143	0.21	0.40	0.00	0.00	0.00	0.00	1.00
Neg. Earnings	106,143	0.42	0.49	0.00	0.00	0.00	1.00	1.00
Unexp. Earnings	106,143	5.35	68.06	0.00	0.01	0.05	0.41	2,401.51
Earnings Volatility	106,143	0.06	0.19	0.00	0.01	0.02	0.05	5.08
Return Volatility	106,143	0.18	0.43	0.04	0.08	0.12	0.16	5.59
Detailed Files								
Accuracy	926,069	-2.17	4.96	-39.68	-1.88	-0.70	-0.24	0.00
CBS-Firm-Analyst (Within Portfolio)	926,069	0.59	0.09	0.00	0.52	0.58	0.65	1.00
CBS-Firm-Analyst (Outside Portfolio)	926,069	0.59	0.09	0.00	0.52	0.58	0.65	1.00
Cosine Analyst (Within Portfolio)	926,069	0.51	0.15	0.00	0.42	0.51	0.61	1.00
Cosine Analyst (Outside Portfolio)	926,069	0.51	0.15	0.00	0.42	0.51	0.61	1.00
Ln(Days)	926,069	4.51	1.10	0.00	4.01	4.72	5.31	5.90
Ln(Firms)	926,069	2.04	0.65	0.69	1.61	2.08	2.48	4.83

Table 11 reports the descriptive statistics for the analyses reported in Table 12. All variables are defined in Appendix B

TABLE 12

Financial Analysts – Regression Analyses

Panel A: Summary Files						
Term List:	**Top-Down**			**Bottom-Up (10-Ks)**		
Dependent Variable:	**Coverage**	**Accuracy**	**Dispersion**	**Coverage**	**Accuracy**	**Dispersion**
Independent Variables:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
CBS-Firm	**0.240*** (5.766)**	**4.619*** (4.654)**	**-3.210*** (-3.501)**	**0.125*** (3.550)**	**3.072*** (3.599)**	**-2.526*** (-3.211)**
Cosine Similarity	0.169*** (6.141)	1.385* (1.653)	-0.593 (-0.761)	0.183*** (8.213)	2.991*** (4.293)	-1.212* (-1.933)
Size	0.027*** (12.880)	2.678*** (25.788)	1.509*** (15.934)	0.027*** (12.971)	2.671*** (25.772)	1.511*** (15.985)
Loss (NI)	-0.008** (-2.318)	-4.177*** (-27.379)	-7.093*** (-48.339)	-0.008** (-2.158)	-4.199*** (-27.548)	-7.113*** (-48.512)
Neg. Earnings	0.015*** (8.149)	0.364*** (5.776)	-0.087 (-1.533)	0.015*** (7.776)	0.359*** (5.716)	-0.075 (-1.319)
Unexp. Earnings	0.000*** (-5.404)	-0.022*** (-10.666)	-0.001 (-0.371)	0.000*** (-5.467)	-0.021*** (-10.657)	-0.001 (-0.380)
Earnings Volatility	-0.052*** (-4.008)	-5.529*** (-8.475)	-4.431*** (-8.500)	-0.051*** (-3.964)	-5.514*** (-8.460)	-4.449*** (-8.523)
Return Volatility	-0.005 (-1.389)	-3.401*** (-12.114)	-0.732*** (-2.821)	-0.005 (-1.298)	-3.419*** (-12.167)	-0.749*** (-2.884)
Observations	105,553	105,553	105,553	106,143	106,143	106,143
Adj. R-Squared	0.246	0.345	0.528	0.247	0.346	0.528
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Clustering	Firm	Firm	Firm	Firm	Firm	Firm
Number of Firm-Years	105,553	105,553	105,553	106,143	106,143	106,143

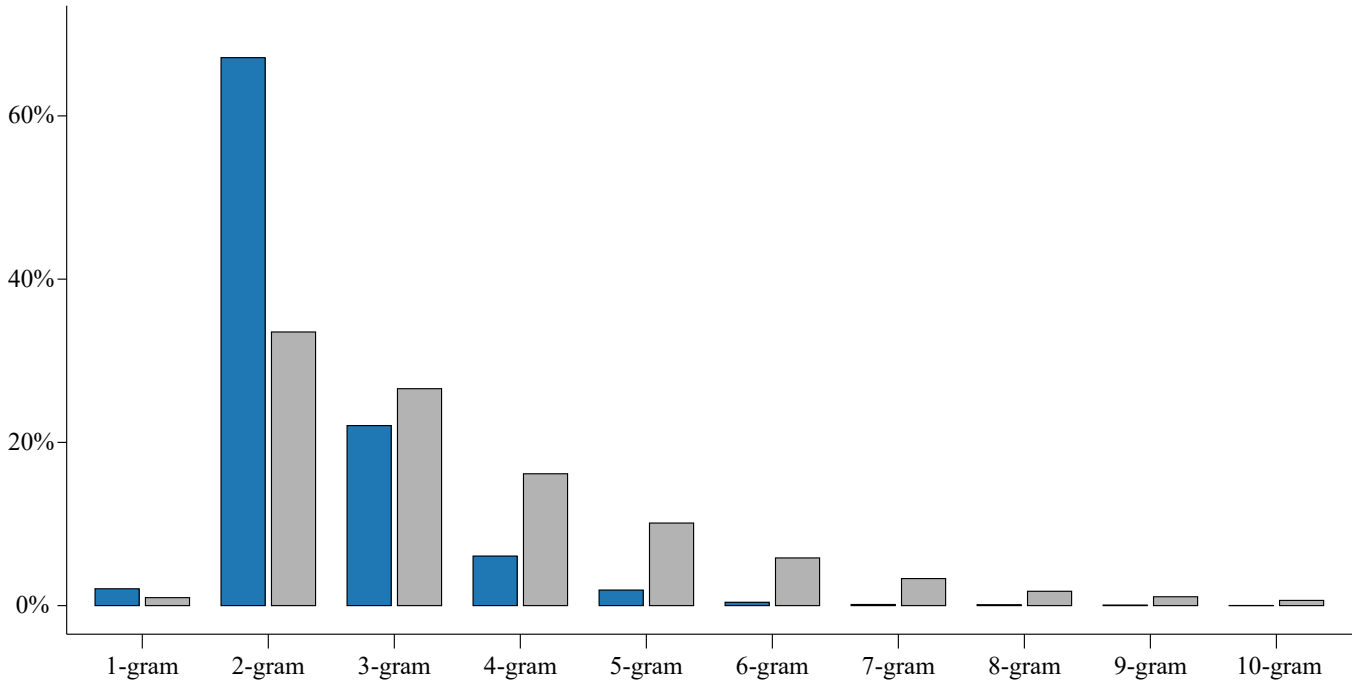
Panel B: Detail Files								
Dependent Variable:	**Accuracy**							
Term List:	**Top-Down**				**Bottom-Up (10-Ks)**			
	Analyst Experience		**Analyst Portfolio CBS**		**Analyst Experience**		**Analyst Portfolio CBS**	
	High	**Low**	**Outside**	**Within**	**High**	**Low**	**Outside**	**Within**
Independent Variables:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
CBS-Firm	**0.440**	**0.854***			**0.882**	**0.754***		
	(0.556)	**(1.918)**			**(1.279)**	**(1.780)**		
Cosine Similarity	0.856*	0.366			1.065***	0.948***		
	(1.724)	(0.875)			(2.788)	(2.785)		
CBS-Firm-Analyst			**0.067**	**0.163***			**0.031**	**0.170***
			(1.069)	**(1.702)**			**(0.438)**	**(1.870)**
Cosine Analyst			-0.002	0.153***			-0.029	0.188***
			(-0.056)	(2.933)			(-0.795)	(3.506)
Size	1.410***	1.222***			1.409***	1.220***		
	(21.388)	(25.601)			(21.393)	(25.627)		
Loss (NI)	-2.892***	-2.962***			-2.908***	-2.978***		
	(-23.781)	(-30.259)			(-23.895)	(-30.471)		
Neg. Earnings	0.184***	0.157***			0.183***	0.156***		
	(5.876)	(5.301)			(5.865)	(5.301)		
Unexp. Earnings	-0.015***	-0.011***			-0.015***	-0.012***		
	(-4.477)	(-8.284)			(-4.459)	(-8.175)		
Earnings Volatility	-3.642***	-2.790***			-3.642***	-2.787***		
	(-6.587)	(-5.536)			(-6.599)	(-5.537)		
Return Volatility	-1.709***	-1.425***			-1.713***	-1.424***		
	(-7.656)	(-9.676)			(-7.684)	(-9.667)		
Ln(Days)	-0.196***	-0.272***	-0.240***	-0.240***	-0.197***	-0.273***	-0.241***	-0.241***
	(-27.409)	(-35.513)	(-57.744)	(-57.673)	(-27.503)	(-35.723)	(-57.932)	(-57.878)
Ln(Firms)	-0.008	0.006	-0.005	-0.002	-0.011	0.005	-0.006	-0.002
	(-0.373)	(0.366)	(-0.621)	(-0.240)	(-0.494)	(0.308)	(-0.758)	(-0.289)
Difference	0.414				-0.128			
(P-Value)	(P = 0.18)				(P = 0.70)			
Observations	412,745	420,255	922,062	922,062	413,885	422,820	926,069	926,069
Adj. R-Squared	0.422	0.425	0.817	0.817	0.422	0.426	0.817	0.817
Firm FE	Y	Y	N	N	Y	Y	N	N
Year FE	Y	Y	N	N	Y	Y	N	N
Firm-Year FE	N	N	Y	Y	N	N	Y	Y
Analyst FE	Y	Y	Y	Y	Y	Y	Y	Y
Clustering	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
Number of Firm-Years	77,669	93,602	108,675	108,675	77,999	94,125	109,223	109,223

Table 12 reports the results of estimating Equations 2 and 3 for the association between terminology standardization and information processing costs for financial analysts as described in Section 4.2.3. Panels A and B are based on the I/B/E/S summary and detail file, respectively. The outcome variable is either *Coverage*, *Accuracy,* or *Dispersion,* as indicated at the top of each Column. *CBS-Firm* is the firm-year level concept-based standardization, and the *CBS-Firm-Analyst* is a version of the CBS-Firm estimated within an analyst's portfolio. All variables

are defined in Appendix B. T-statistics reported in parentheses are based on standard errors clustered at the firm level. *, ** and *** indicate significance: * at the 10% level, ** at the 5% level and *** at the 1% level respectively.

FIGURE 1

Panel A: n-grams Frequency and Unique Terms (Top-Down)



Panel B: n-grams Frequency and Unique Terms (Bottom-Up)

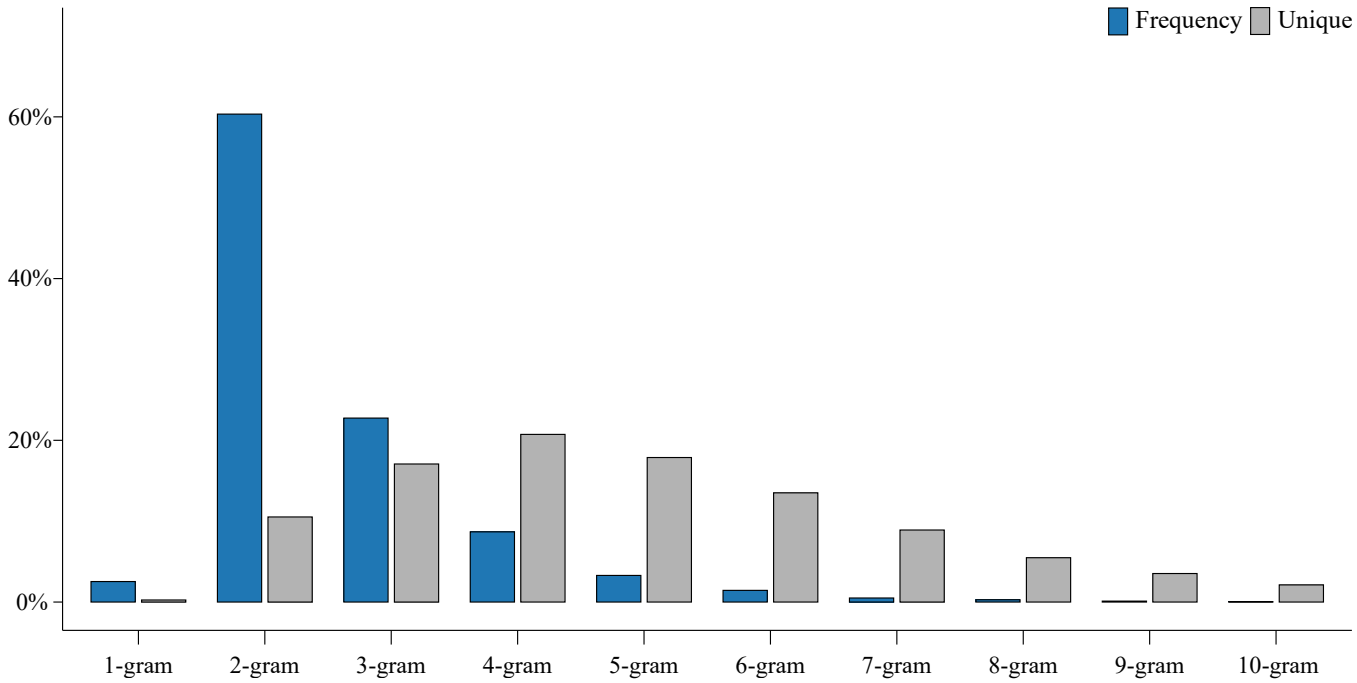
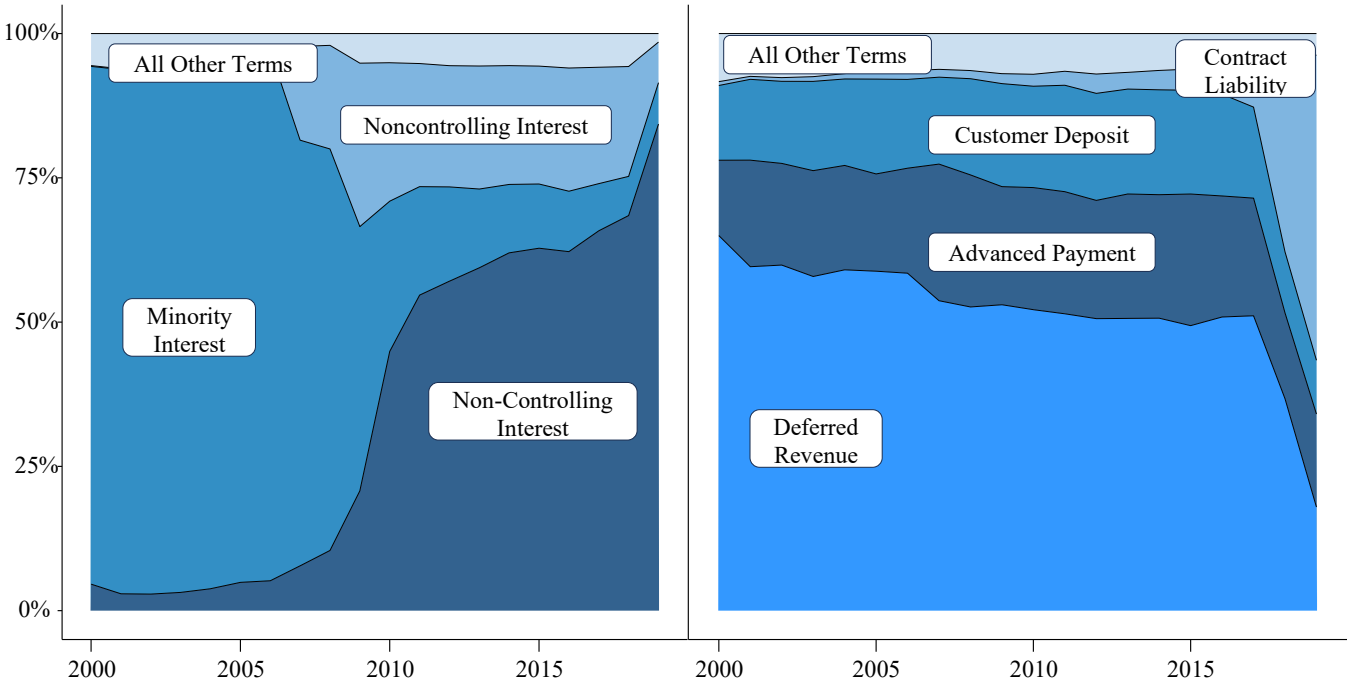


Figure 1 shows the distribution of n-grams across our different term lists. Panel A displays the n-gram distribution for the top-down list, Panel B shows the n-gram distribution for the bottom-up lists (we combine our bottom-up 10-K and 20-F lists). For each panel, the blue bars represent the relative frequency distribution, showing how often n-grams appear in our corpus, while the gray bars show the unique occurrences in our term lists.

FIGURE 2

Panel A: Concept-Development (Top-Down)



Panel B: Concept-Development (Bottom-Up)

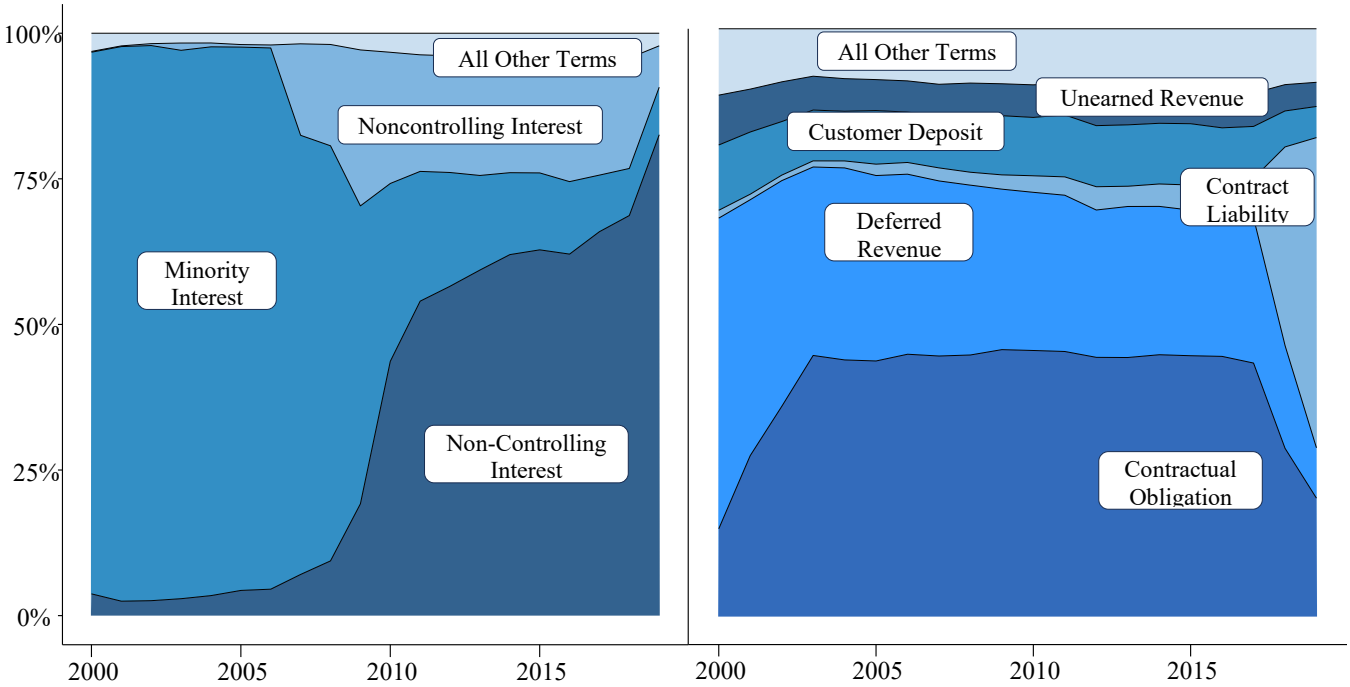


Figure 2 shows examples of the development of different accounting concepts over time. Panel A shows the distribution for our top-down list and Panel B for our bottom-up list (we combine our bottom-up 10-K and 20-F lists).

APPENDIX A

Panel A: Experimental Conditions – High (Low) Standardized Statements

Income Statement	
Low Standardization	**High Standardization**
REVENUE	NET SALES
COST OF PRODUCT REVENUE	COST OF GOODS SOLD
GROSS MARGIN	GROSS PROFIT
SELLING GENERAL AND ADMINISTRATIVE COSTS	SELLING GENERAL AND ADMINISTRATIVE EXPENSES
PROVISION FOR DEPRECIATION AND AMORTIZATION	DEPRECIATION AND AMORTIZATION
COST OF OPERATIONS	TOTAL OPERATING EXPENSES
EARNINGS FROM OPERATIONS	OPERATING INCOME
INVESTMENT INCOME	INTEREST AND OTHER INCOME
FINANCE EXPENSES NET	INTEREST EXPENSE
SUNDRY INCOME (EXPENSE) NET	OTHER INCOME (EXPENSE) NET
(LOSS) EARNINGS BEFORE PROVISION FOR INCOME TAXES	INCOME BEFORE INCOME TAXES
INCOME TAX EXPENSE	PROVISION FOR INCOME TAXES
NET PROFIT (LOSS)	NET INCOME

Balance Sheet - Asset Side	
Low Standardization	**High Standardization**
CASH AND CASH ITEMS	CASH AND CASH EQUIVALENTS
TRADE RECEIVABLES	ACCOUNTS RECEIVABLE
INVENTORIED COSTS	INVENTORIES
OTHER ASSETS CURRENT	OTHER CURRENT ASSETS
TOTAL ASSETS CURRENT	TOTAL CURRENT ASSETS
FIXED ASSETS	PROPERTY PLANT AND EQUIPMENT
INTANGIBLES	INTANGIBLE ASSETS
FUTURE INCOME TAX BENEFITS	DEFERRED TAX ASSETS
MISCELLANEOUS ASSETS	OTHER ASSETS
TOTAL ASSETS	TOTAL ASSETS
Balance Sheet - Liabilities and Equity Side	
Low Standardization	**High Standardization**
CURRENT BORROWINGS	SHORT TERM BORROWINGS
TRADE PAYABLES	ACCOUNTS PAYABLE
ADVANCES FROM CUSTOMERS	DEFERRED REVENUE
ACCRUED TAXES	INCOME TAXES PAYABLE
CURRENT INSTALLMENTS OF LONG TERM BORROWINGS	CURRENT PORTION OF LONG TERM DEBT
ACCRUED LIABILITIES	ACCRUED EXPENSES
CURRENT LIABILITIES TOTAL	TOTAL CURRENT LIABILITIES
DEBT DUE AFTER ONE YEAR	LONG TERM DEBT
OTHER OBLIGATIONS	OTHER LIABILITIES
TOTAL LIABILITIES	TOTAL LIABILITIES
ORDINARY SHARES	COMMON STOCK
CAPITAL IN EXCESS OF PAR VALUE	ADDITIONAL PAID IN CAPITAL
CUMULATIVE NET INCOME	RETAINED EARNINGS
CUMULATIVE OTHER COMPREHENSIVE EARNINGS (LOSS)	ACCUMULATED OTHER COMPREHENSIVE EARNINGS (LOSS)
REPURCHASED STOCK	TREASURY STOCK
TOTAL MEMBERS CAPITAL	TOTAL STOCKHOLDERS EQUITY
TOTAL LIABILITIES AND MEMBERS EQUITY	TOTAL LIABILITIES AND STOCKHOLDERS EQUITY

Panel B: Task Description and Example Question

Assume you have been hired by a data provider to collect accounting information from companies' financial statements. Your task is to match line items from a firm's balance sheet and income statement to a corresponding list of variables. Each line item from the variables' list must be matched to exactly one line item from the company's financial statement.

The company's income statetement is provided below

REVENUE
COST OF PRODUCT REVENUE
GROSS MARGIN
SELLING GENERAL AND ADMINISTRATIVE COSTS
PROVISION FOR DEPRECIATION AND AMORTIZATION
COST OF OPERATIONS
EARNINGS FROM OPERATIONS
INVESTMENT INCOME
FINANCE EXPENSES NET
SUNDRY INCOME (EXPENSE) NET
(LOSS) EARNINGS BEFORE PROVISION FOR INCOME TAXES
INCOME TAX EXPENSE
NET PROFIT (LOSS)

Your task is to match **REVENUE** to the list below

	REVENUE
INTEREST AND RELATED EXPENSE	○
NET INCOME (LOSS)	○
OPERATING INCOME AFTER DEPRECIATION	○
SALES TURNOVER (NET)	○
OPERATING EXPENSES TOTAL	○
GROSS PROFIT (LOSS)	○
INTEREST AND RELATED INCOME	○
NONOPERATING INCOME (EXPENSE) TOTAL	○
PRETAX INCOME	○
DEPRECIATION AND AMORTIZATION TOTAL	○
COST OF GOODS SOLD	○
SELLING GENERAL AND ADMINISTRATIVE EXPENSES	○
PROVISION FOR INCOME TAXES	○

APPENDIX B

Variables Definitions

Dependent Variables	**Definition**
Missing item	*Missing Item* is an indicator variable capturing whether the data item corresponding to concept *c* for firm *i* at time *t* is missing.
Error	*Error* is an indicator variable that takes the value of 1 if GPT fails to match a firm's line item to a concept *c* for firm *i* at time *t*. Where a concept *c* is either a Worldscope or Compustat database variable.
Coverage	*Coverage* is the natural logarithm of the number of analysts following a firm in a given year.
Accuracy	*Accuracy* is the absolute value of the last mean or individual forecast error before the earnings announcement for firm *i* at time *t* scaled by lag price and multiplied by -100.
Dispersion	*Dispersion* is the standard deviation of an analysts' mean forecast error scaled by price and multiplied by 100.
Independent Variables	**Definition**
CBS-Firm-Concept	*CBS-Firm-Concept* is the concept-based standardization of firm *i* at time *t* for concept *c* computed as the mean concept-based standardization between firm *i* and all firms *j*s in the same industry (three-digit ICB code) as described in Appendix C.
CBS-Firm	*CBS-Firm* is the concept-based standardization of firm *i* at time *t* computed as the mean concept-based standardization (based on CBS-Pair$_{cw}$) between firm *i* and all firms *j*s in the same industry (three-digit ICB code).
CBS-Firm-Analyst	*CBS-Firm-Analyst* is the concept-based standardization of firm *i* at time *t* for analyst *a*, computed as the mean concept-based standardization between firm *i* and all firms in analyst *a*'s portfolio in year *t*.
Embedding Similarity	*Embedding Similarity* is the cosine similarity of each term to other terms within the same concept's vector space representation of OpenAI's TextEmbedding3Large embedding model.
GPT Masking Prediction	*GPT Masking Prediction* is a variable between 0 and 1 that captures the percentage of times GPT predicts a certain term within a concept when presented with a sentence from a firm's 10-K Item 8 section, where the term is masked.
Relative Frequency	*Relative Frequency* is a variable between 0 and 1 that captures the relative frequency of a term within a concept based on our complete corpus.
Control Variables	Definition
Age	*Age* of a firm in years is defined as the difference between a firm's fiscal year-end [Worldscope Code: 05350] and the minimum date of i) a firms coverage date, ii) date of incorporation [Worldscope Code: 18273], and iii) founding date [Worldscope Code: 18272]
Book-to-Market	*Book-to-Market* ratio is the book value of equity [Worldscope Code: 07220] divided by the market value of firm [Worldscope Code: 07210] *i* at time *t*.
Cosine Similarity	*Cosine Similarity* is the cosine similarity of firm *i* at time *t* computed as the mean cosine similarity between firm *i* and all firms *j*s in the same industry (three-digit ICB code).
Cosine Analyst	*Cosine Analyst* is the cosine similarity of firm *i* at time *t* for analyst *a*, computed as the mean cosine similarity between firm *i* and all firms in analyst *a*'s portfolio in year *t*.
Earnings Volatility	*Earnings Volatility* is the rolling standard deviation of net income over the previous four fiscal years [Worldscope data: WC01551].
Leverage	*Leverage* is the debt-to-assets ratio of firm *i* at time *t*, it is computed using the sum of long-term debt [Worldscope Code: 03251] and short-term debt [Worldscope Code: 03051] divided by total assets [Worldscope Code: 02999].
Loss (NI)	*Loss (NI)* is an indicator variable capturing whether firm *i* incurred a loss [Worldscope data: 01551] at time *t*.
N (Data Items)	*N (Data Items)* measures the total number of terms with a matched database variable in a GPT API call.
Neg. Earnings	*Neg. Earnings* is an indicator variable capturing whether firm i net income in year *t* is less than its net income in year *t-1*.
Pair Similarity	*Pair Similarity* is the character level cosine similarity between the term the firm is using and the variable name in the database.
Return on Assets	*Return on Assets* is defined as the net income before extraordinary items and preferred dividends [Worldscope Code: 01551] divided by total assets [Worldscope Code: 02999] of firm *i* at time *t*.
Return Volatility	*Return Volatility* is the standard deviation of monthly stock returns over the previous 48 months.
Size	*Size* is the natural logarithm of the market value of firm *i* at time *t* [Worldscope Code: 07210].
Unexp. Earnings	*Unexp. Earnings* is the absolute change in net income from *t-1* to *t* scaled by lagged stock price [Worldscope Codes: 01551, 05017].

Appendix B provides an overview of the variables used throughout the manuscript and their definitions.

APPENDIX C

Concept-based Standardization

We employ techniques from computational linguistics (e.g., Jurafsky and Martin, 2021) and introduce a novel measure, i.e., concept-based standardization (e.g., Speelman et al., 2003), as our main proxy for the level of standardization of accounting terminology. While the cosine similarity (e.g., Lang and Stice-Lawrence, 2015) captures *how similarly firms discuss their accounting*, the concept-based standardization (CBS) captures *how similarly firms address the same accounting issue*. By grouping terms according to their meaning, we control for the underlying economic transaction (assuming our concept lists are precise).

Specifically, for a given concept C_k (e.g., minority interest), a set of n possible synonyms S_l (e.g, "minority interest" versus "non-controlling interest"), and a firm-pair *i* and *j* at time *t*, the concept-based standardization measure is defined as follows:

$$CBS_{C_k,i,j,t}=1-\frac{1}{2}\sum_{l=1}^{n}\left|R_{C_k,i,t}(S_l)-R_{C_k,j,t}(S_l)\right|,$$

where $R_{C_k,i,t}(S_l)$ and $R_{C_k,j,t}(S_l)$ are the relative frequencies of appearance of synonym S_l within concept C_k for firm *i* and *j* at time *t* in their text corpus (here: annual reports). The concept-based standardization measure (based on the Manhattan distance) is bounded between 0 and 1, where higher levels indicate higher terminology standardization between the two firms for a concept. We calculate the CBS over time for each unique pair of firms in our sample that belongs to the same industry (as defined by the three-digit ICB code) following the literature on comparability (e.g., De Franco et al., 2011; Lang and Stice-Lawrence, 2015). The resulting dataset contains the most granular version of our CBS measure (over 5 billion *firm-pair-year-concept* observations).

To test the association between terminology standardization and information acquisition costs for data analysts, we collapse the above CBS at the *firm-pair-year-concept* level within industry

into a *firm-year-concept* level proxy *CBS-Firm-Concept* (i.e., for a firm *i* in industry A, we compute the mean CBS, relative to all M firms *j*s in the same industry). The *CBS-Firm-Concept* captures the average *firm-year-concept* standardization relative to all other firms in the same industry.

$$\text{CBS-Firm-Concept}_{C_k,i,t}=\frac{1}{M}\sum_{j=1}^{M}\text{CBS}_{C_k,i,j,t}$$

We aggregate this measure to a firm-year level for the financial analysts' tests as follows. First, we move from a *firm-pair-year-concept* level to an aggregated measure across all M concepts:

$$\text{CBS-Pair}_{i,j,t}=\sum_{C_k=1}^{m}\left(\text{CBS}_{C_k,i,j,t}\right)W(C_k),$$

where each concept is weighted by its relative occurrence within an industry year in our sample (*CBS-Pair*$_{cw}$).[28] Second, we collapse the measure to a *firm-year* level (*CBS-Firm*) by taking the mean of the *CBS-Pair* computed across all firm pairs (1 to N).

$$\text{CBS-Firm}_{i,t}=\frac{1}{N}\sum_{j=1}^{N}\text{CBS-Pair}_{i,j,t}$$

Finally, we construct a version of the *CBS-Firm* at the *analyst-portfolio-firm-year* level by taking as an input only the CBS-Pairs computed for firm-pairs within the same analyst portfolio in a given year. Specifically, for a firm *i* covered by analyst *a* at time *t*, we collapse the measure to a *firm-analyst-year* level by taking the mean of the CBS-Pair computed across all P firms in analyst a's portfolio. We label this measure *CBS-Firm-Analyst*.

$$\text{CBS-Firm-Analyst}_{i,a,t}=\frac{1}{P}\sum_{j=1}^{P}\text{CBS-Pair}_{i,j,t}$$

[28] To illustrate, consider two concepts, A and B, within an industry-year (e.g., utilities in 2001) and, for simplicity, assume that the frequency of all other concepts is zero (i.e., they are not used within that industry-year). To calculate *CBS-Pair*$_{cw}$, we assign a weight to concepts A and B that reflects their relative usage by firms. If A is used by 25 firms and B is used by 50 firms, we weight B twice as much as A; i.e. A has a relative weight of 0.33 and B of 0.66.

APPENDIX D

Prompts – Illustrative Examples

Prompt Example: Items Only, Select Matching

TASK DESCRIPTION:

You are a financial analyst at WORLDSCOPE tasked with matching standardized line items from WORLDSCOPE to corresponding line items in a company's financial statement.
Below you find the line item names both for WORLDSCOPE and the company (Enclosed in curly brackets {}).

RULES:

1. Each WORLDSCOPE item must be matched to exactly one company item.
2. Matches must be exclusive (no two WORLDSCOPE items can match to the same company item).
3. Never Ask for Clarifications, complete the task as requested! If you are uncertain, use your expert knowledge to resolve it (while adhering to the specified output format)!

DATA TO BE USED FOR MATCHING TAKS:

Company Line Items {
- Income (loss) before nonoperating expenses
- Interest expense
- Income (loss) before income taxes
- Depreciation, depletion and amortization
- Selling, general and administrative expenses
- Income (loss) from operations
- Provision for (benefit from) income taxes
- Net income (loss)

}

WORLDSCOPE Line Items {
- Income Taxes
- Net Income- Bottom Line
- Gross Income
- Pre-tax Income
- Selling, General & Administrative Expenses
- Interest Expense on Debt
- Operating Income
- Depreciation, Depletion & Amortization

}

Prompt Example: Items with Financial Statement Context, Select Matching

TASK DESCRIPTION

You are a financial analyst at WORLDSCOPE tasked with matching standardized line items from WORLDSCOPE to corresponding line items in a company's financial statement.
Below you find the line item names both for WORLDSCOPE and the company as well as the Full Financial Statement (Enclosed in curly brackets {}).

RULES

1. Each WORLDSCOPE item must be matched to exactly one company item.
2. Matches must be exclusive (no two WORLDSCOPE items can match to the same company item).
3. Never Ask for Clarifications, complete the task as requested! If you are uncertain, use your expert knowledge to resolve it (while adhering to the specified output format)!

ADDITIONAL INFORMATION:

Revenues	1,467,592
Costs, expenses and other operating	
Cost of sales (exclusive of items shown separately below)	1,378,479
Depreciation, depletion and amortization	121,552
Accretion on asset retirement obligations	19,887
Change in fair value of coal derivatives and coal trading activities, net	5,219
Selling, general and administrative expenses	82,397
Costs related to proposed joint venture with Peabody Energy	16,087
Asset impairment and restructuring	221,380
Gain on property insurance recovery related to Mountain Laurel longwall	(23,518)
(Gain) loss on divestitures	(1,505)
Preference Rights Lease Application settlement income	—
Other operating income, net	(22,246)
Income (loss) from operations	1,797,732
Interest expense, net	(330,140)
Interest expense	(14,432)
Interest and investment income	3,808
Income (loss) before nonoperating expenses	(10,624)
Nonoperating (expenses) income	(340,764)
Non-service related pension and postretirement benefit (costs) credits	(3,884)
Net loss resulting from early retirement of debt and debt restructuring	—
Reorganization items, net	26
Income (loss) before income taxes	(344,622)
Provision for (benefit from) income taxes	(7)
Net income (loss)	(344,615)
Net income (loss) per common share	
Basic earnings (loss) per share	(22.74)
Diluted earnings (loss) per share	(22.74)
Weighted average shares outstanding	
Basic weighted average shares outstanding	15,153
Diluted weighted average shares outstanding	15,153

DATA TO BE USED FOR MATCHING TAKS:

Company Line Items {
- Income (loss) before nonoperating expenses
- Interest expense
- Income (loss) before income taxes
- Depreciation, depletion and amortization
- Selling, general and administrative expenses
- Income (loss) from operations
- Provision for (benefit from) income taxes
- Net income (loss)

}

WORLDSCOPE Line Items {
- Income Taxes
- Net Income- Bottom Line
- Gross Income
- Pre-tax Income
- Selling, General & Administrative Expenses
- Interest Expense on Debt
- Operating Income
- Depreciation, Depletion & Amortization

}

Prompt Example: Items with Financial Statement Context, Select Matching

TASK DESCRIPTION

You are a financial analyst at WORLDSCOPE tasked with matching standardized line items from WORLDSCOPE to corresponding line items in a company's financial statement.
Below you find the line item names both for WORLDSCOPE and the company as well as the Full Financial Statement (Enclosed in curly brackets {}).

RULES

4. Each WORLDSCOPE item must be matched to exactly one company item.
5. Matches must be exclusive (no two WORLDSCOPE items can match to the same company item).
6. Never Ask for Clarifications, complete the task as requested! If you are uncertain, use your expert knowledge to resolve it (while adhering to the specified output format)!

ADDITIONAL INFORMATION:

Revenues	1,467,592
Costs, expenses and other operating	
Cost of sales (exclusive of items shown separately below)	1,378,479
Depreciation, depletion and amortization	121,552
Accretion on asset retirement obligations	19,887
Change in fair value of coal derivatives and coal trading activities, net	5,219
Selling, general and administrative expenses	82,397
Costs related to proposed joint venture with Peabody Energy	16,087
Asset impairment and restructuring	221,380
Gain on property insurance recovery related to Mountain Laurel longwall	(23,518)
(Gain) loss on divestitures	(1,505)
Preference Rights Lease Application settlement income	—
Other operating income, net	(22,246)
Income (loss) from operations	1,797,732
Interest expense, net	(330,140)
Interest expense	(14,432)
Interest and investment income	3,808
Income (loss) before nonoperating expenses	(10,624)
Nonoperating (expenses) income	(340,764)
Non-service related pension and postretirement benefit (costs) credits	(3,884)
Net loss resulting from early retirement of debt and debt restructuring	—
Reorganization items, net	26
Income (loss) before income taxes	(344,622)
Provision for (benefit from) income taxes	(7)
Net income (loss)	(344,615)
Net income (loss) per common share	
Basic earnings (loss) per share	(22.74)
Diluted earnings (loss) per share	(22.74)
Weighted average shares outstanding	
Basic weighted average shares outstanding	15,153
Diluted weighted average shares outstanding	15,153

DATA TO BE USED FOR MATCHING TAKS:

WORLDSCOPE Line Items {
  Income Taxes
  Net Income- Bottom Line
  Gross Income
  Pre-tax Income
  Selling, General & Administrative Expenses
  Interest Expense on Debt
  Operating Income
  Depreciation, Depletion & Amortization
}