What Anchors Standard-setting Deliberations?*

Jie Cui†

September 2025

Abstract

This paper introduces an LLM-based framework that organizes International Accounting Standards Board (IASB) meeting deliberations into a hierarchical taxonomy, providing a transparent map from unstructured dialogue to interpretable constructs capturing stakeholder inclusivity (whose perspectives are invoked), evidentiary grounding (how arguments are substantiated), and topic specificity (which issues are emphasized). Using this taxonomy, I provide novel evidence on the procedural quality and effectiveness of standard-setting, including metrics of deliberation range, depth, balance and polarization, as well as measures of standard-setter style. Three salient patterns emerge. (i) No single stakeholder group, particularly the accounting profession, dominates the discourse; the mix of voices is broader than commonly presumed. (ii) Deliberations rely more on expert judgment (experiential, technical, and theoretical claims) than on objective factual assertions, aligning with an evidence-informed rather than purely evidence-based policymaking paradigm. (iii) Nontrivial between-speaker heterogeneity and persistent within-speaker patterns in deliberative communication reveal diverse and stable standard-setter styles. By presenting a replicable taxonomy and analysis pipeline, this study furnishes tools and benchmarks for subsequent work on the political economy of accounting standard-setting.

Keywords: LLMs; Taxonomy generation; IASB; Legitimacy and due process assessment; Natural language processing; Machine learning
JEL Codes: M41, M48, D72, C55

*I am deeply grateful to Laurence van Lent for his guidance. I am indebted to Katharina Hombach and Yuping Jia for their continuous feedback and support. I also thank Katja Kisseleva-Scherenberger, Yifei Liao, Matthias Mahlendorf, Clare Wang, Ruishen Zhang, and Menghan Zhu for their valuable comments. I gratefully acknowledge funding from the Deutsche Forschungsgemeinschaft Project ID 403041268—TRR 266.

†**Frankfurt School of Finance & Management**; Adickesallee 32-34, 60322 Frankfurt am Main, Germany; E-mail: j.cui@fs.de.

1. Introduction

Several questions motivate this paper. First, *how effective is the IASB standard-setting process?* Official meetings of the IASB are crucial in developing global accounting standards. Yet, systematic research on the content of deliberations, including their effectiveness and quality, remains scarce. The *Due Process Handbook* of IFRS Foundation emphasizes that the IASB's primary objective is to develop a set of globally accepted high-quality standards in the public interest, and that this should be done through open consultation with a wide range of global stakeholders (IFRS Foundation, 2020).[1] Accordingly, understanding how effectively the IASB addresses and balances the concerns of different stakeholders (e.g., whether there are dominant voices or underrepresented groups) is an important line of inquiry in itself.

Second, *to what extent do the IASB's deliberations incorporate objective, verifiable evidence that supports evidence-based policymaking paradigm?* Evidence-based policymaking, a term used to describe the need for more scientific and less ideological policymaking, is increasingly being advocated as best practice (OECD, 2021; European Commission, 2023). Calls from the OECD and the European Commission underscore the need to observe and measure deliberation itself, not just final output. Thus, deciphering the content of deliberations and linking it to the norms of evidence-based policymaking can better inform the procedural quality of standard-setting bodies.

Third, *are there stable differences (i.e., "styles") among standard setters during IASB meeting discussions?* The ideology theory posits that accounting standards are the combined result of standard setters' own ideologies regarding accounting principles and the lobbying efforts of interest groups, which may not necessarily be optimal in promoting efficient capital allocation (Kothari et al., 2010). Therefore, examining whether standard setters consistently exhibit different patterns, particularly in terms of whose perspectives they advocate and what evidentiary basis they rely on to justify their views, and whether this "style" has a substantive impact on deliberations, can help us understand the true nature of the standard-setting process.

Finally, *are the "unintended" consequences studied in previous research truly unintended?* Numerous studies have examined the unintended consequences of standards, regulations, or laws (Arya

[1] IFRS stands for International Financial Reporting Standards, a set of accounting standards issued by the International Accounting Standards Board (IASB) that promote consistency, transparency, and comparability in financial reporting worldwide. For background, see the IFRS Foundation website: https://www.ifrs.org/.

et al., 2005; Anderson, 2009; Mongrut and Winkelried, 2019). However, due to a lack of verifiable data, it is unclear whether these consequences are truly unintended (not anticipated or discussed by the standard setters) or simply the result of consensus reached by the standards-setting bodies. This study provides a reusable taxonomy and scalable measurement framework that future research can use to verify whether the consequences studied are truly unintended.

This paper introduces an LLM-based framework that organizes IASB meeting transcripts into a hierarchical taxonomy. This taxonomy encompasses stakeholder orientations, forms of expertise, and fine-grained topics, enabling researchers and policymakers to effectively examine the standard-setting process at multiple levels of detail. I apply the framework to 66 meetings (2013–2021) covering the Conceptual Framework, Fair Value Measurement, and Leases standards, yielding 8,758 substantive speech segments by 26 board members.[2,3] The analysis yields three main empirical patterns.

Breadth of voices. Board members regularly invoke multiple stakeholder perspectives. In the full sample, Preparers account for roughly 46% of substantive segments, Users about 28%, Regulators (≈10%), Accounting Professions (≈8%), Academics (≈3%), and Public Interest (< 1%). Although Preparers are most frequently invoked (given the deliberations center on preparer-facing technical judgments and implementation details), no single group dominates the discourse; multiple perspectives are consistently present. Complementing these shares, meeting-level inclusiveness and balance indices (based on normalized HHI) indicate that discussion is often distributed across several categories rather than concentrated in one.

Forms of expertise. After capturing whose interests are invoked, I further decompose the deliberations by source of authority/knowledge, drawing on argumentation scholarship about how appeals to evidentiary basis function in policy reasoning. In these data, factual claims comprise about 14% of substantive segments, whereas experiential, technical, and theoretical claims together account for about 55%. This pattern is consistent with an evidence-informed policymaking paradigm in

[2]In this paper, a *meeting* is defined as a unique Standard×Meeting-Date pair in the IASB's official schedule. The IASB typically holds its official meetings over several consecutive days within each month (except August) to discuss agenda items related to IFRS standards. Accordingly, *Meeting-Date* refers to the month of the meeting cycle (e.g., 2018-05 for May 2018), not a specific calendar day. A speech segment is defined as an uninterrupted speech by a speaker at an IASB meeting. All segments that are not substantive to the deliberations (e.g., greetings) are removed for this study.

[3]For brevity, I refer to the *Conceptual Framework* as a "standard" in this paper, even though it is not an IFRS standard. In all *standard-level* tables and figures, the Conceptual Framework is included as a separate standard category alongside other standards.

which expert judgment and reasoning, not only factual assertions, play a large role in deliberation.

Standard-setter style. Mixed-effects variance decompositions reveal economically meaningful between-speaker heterogeneity, especially for *Accounting Professions* (stakeholder orientation) and *Factual Claim* (form of expertise), where the speaker component accounts for $\approx$ 13–14% of total variance. Meanwhile, substantive standard and meeting-date effects were observed, indicating the combined influence of individual style and contextual factors. A complementary fixed-effects ΔR^2 ladder corroborates that style is primarily a stable across-topic difference rather than a purely topic-contingent rhetorical device. Individual speaker-level displays (e.g., context-adjusted speaker fixed-effects coefficients) show persistent nonzero deviations for several members, and indicate that these tendencies port across standards. Pairwise correlations further reveal clusters of similar emphasis and, in forms-of-expertise mixes, several clear contrasts (negative correlations), signaling divergent styles.

To summarize, the evidence confirms that IASB deliberations are pluralistic but structured: multiple stakeholder perspectives are present in most meetings; expert judgment (experiential, technical, theoretical) is more prevalent than purely factual assertions; and standard-setter style varies meaningfully across individuals, with stable within-speaker profiles and clusterable between-speaker differences.

The paper contributes in three ways. First, it offers the first systematic, multi-resolution mapping of IASB standard-setting deliberations, covering the breadth of stakeholder orientations, the balance between factual and expert-judgment claims, the distribution of fine-grained topics, and the standard- and speaker-level structure of variation. These outputs extend prior work on standard-setting politics by providing process-level, quantitative evidence on the quality and effectiveness of standard-setting: who is heard (stakeholder mix), what types of evidentiary basis are deployed (forms of expertise), and how emphasis varies across standards, meetings, and speakers, thereby establishing benchmarks and stylized facts that subsequent theory-driven and identification-based studies can test and build upon (Kothari et al., 2010). A further motivation for constructing this structured taxonomy stems from evidence that disclosure processing costs impede the diffusion of information; as a result, disclosures may not be as "public" as traditionally assumed but can function as costly private information (Blankespoor et al., 2020). Since not all market participants or IFRS stakeholders have the resources and capabilities to mine and analyze large volumes of

standard-setting deliberation texts, the mapping in this study can help reduce those search and classification costs, thereby promoting transparency in policymaking.

Second, the paper provides a reusable taxonomy and a scalable measurement framework that transform unstructured deliberation text into a structured and meaningful form, organized by useful category labels, enabling investors, researchers, and policymakers to navigate the information landscape more effectively. Methodologically, the approach aligns with emerging LLM text-mining that automates label generation and assignment at scale, and with research on principled prompt optimization, that is, systematic, data-driven procedures for designing and tuning prompts (e.g., search and iterative refinement with quantitative feedback) to improve classification quality and reproducibility (Wan et al., 2024; Zhang et al., 2025). This design choice is also motivated by the argumentation literature, which calls for large, reliably annotated corpora of real-world argumentative discourse because insight into how argumentation is actually used in practice is essential (Visser et al., 2020). The study responds by delivering (i) a sizable corpus annotated with theory-grounded constructs and (ii) a quantitative pipeline that can be reused and extended to other institutional settings (e.g., corporate board meetings, earnings calls, or central-bank committees).

Third, by converting deliberations into searchable, fine-grained topics linked to stakeholder appeals and forms of expertise, this study provides an empirical basis to verify whether a post-implementation outcome of standards is an unanticipated side effect (i.e., a genuinely "unintended" consequence) or a trade-off explicitly accepted by consensus. The framework allows researchers to assess whether and, with what prominence and evidentiary posture, the issue was anticipated, debated, and weighed during standard-setting, thereby adding process-level evidence and methodologies to the literature on unintended consequences (Brüggemann et al., 2013). The framework can also support analogous audits in expert bodies beyond the IASB.

The remainder of the paper proceeds as follows. Section 2 provides conceptual underpinnings and literature review. Section 3 details the taxonomy construction and data sources. Section 4 describes the deliberation metrics (range, depth, and balance/polarization) and statistics of the taxonomy layers (stakeholder orientations; forms of expertise, and fine-grained topics). Section 5 reports the main descriptive results, including standard-setter style analysis, pairwise correlation patterns among standard setters, and association tests between stakeholder orientations and forms of expertise. Section 6 presents robustness and validation. Section 7 concludes.

2. Theory and Related Literature

In this section, I explain why standard-setting deliberations, particularly those of the IASB, matter for global financial markets, and how this motivates a multi-level taxonomy of the discourse (stakeholder orientations, forms of expertise, fine-grained topics). I then develop why the styles of individual standard-setters and potential coalitions or cleavages among them are theoretically salient for procedural quality and effectiveness.

2.1. Standard-setting Deliberations

Committee deliberations are the locus where policies are framed, evidence and judgment are gathered, and trade-offs are reasoned through; in expert bodies, the mix of information revelation, persuasion, and reputation concerns can shape deliberation patterns, committee decisions, and downstream legitimacy (Levy, 2007; Visser and Swank, 2007; Swank and Visser, 2023). Research on monetary-policy committees shows that transparency measurably changes members' behavior: "opening the black box" of discussion alters communication styles and reveals both discipline and conformity forces, with career/audience incentives affecting voting and information revelation (Hansen et al., 2018). In technical rule-making domains such as accounting, often characterized as "thin political markets," with specialized expertise, complex interest cleavages, and persuasion within bureaucracies, these forces channel how stakeholder pressures and expert judgments enter deliberation and are translated into rules (Kothari et al., 2010; Ramanna, 2015; Vogel, 2022). Taken together, this literature motivates direct analysis of deliberations in standard-setting meetings: what is argued, how it is argued, and for whom.

Since its creation in 2001, the IFRS Foundation (through the IASB) has reshaped the global reporting landscape: IFRS Accounting Standards are, in effect, a global financial language used or required in more than 140 jurisdictions, enabling cross-border investment and fostering capital-market efficiency and stability around the world (IFRS Foundation, 2025). IFRS adoption has measurable real consequences in the global economy. For example, an early cross-country study documents economically meaningful effects around mandatory IFRS adoption: on average, higher market liquidity, lower costs of capital, and higher valuations, though magnitudes vary with institutions and enforcement (Daske et al., 2008). Moreover, DeFond et al. (2011) show that when

mandatory IFRS in the EU credibly increases comparability via greater uniformity and credible implementation, foreign mutual fund ownership rises, consistent with lower information frictions for international investors. This evidence makes clear that IFRS has a measurable, real impact on global capital markets.

The due process of IFRS is expressly built on transparency, full and fair consultation, and accountability, making the formal IASB meeting deliberations a central component of procedural legitimacy (IFRS Foundation, 2020). In other words, the IASB's deliberations are not merely ceremonial but the central mechanism by which the IFRS standards are debated, justified, and refined. Understanding the content and structure of those deliberations is therefore integral to assessing procedural quality and, ultimately, the credibility of the standards.

These considerations motivate a hierarchical taxonomy of IASB discourse at three levels, designed to map observable speech to recognized drivers of procedural quality and regulatory effectiveness:

Stakeholder orientations. Consistent with the IFRS due-process requirement to engage affected constituencies (e.g., users, preparers, regulators, accounting professions), tracking which audiences are invoked or addressed in Board dialogue provides evidence on participation breadth, responsiveness, and balance, which are key facets of input and throughput legitimacy, and align with "better regulation" frameworks that stress inclusive consultation and reason-giving across the policy cycle (IFRS Foundation, 2020; European Commission, 2023). The stakeholder-orientations layer of the taxonomy is designed to make this dimension measurable.

Forms of expertise. Classifying arguments by evidentiary basis (i.e., factual claims versus experiential, technical, or theoretical reasoning) links Board discourse to evidence-based policymaking norms. Regulatory guidelines and the policy-making literature emphasize using the "best available evidence," while recognizing that expert judgment and professional reasoning remain central when evidence is incomplete (OECD, 2021; Head, 2016). Tracking the relative mix of factual versus expert-judgment claims in deliberations therefore operationalizes a core procedural quality dimension. The forms-of-expertise layer of the taxonomy mirrors this plural conception, letting us assess whether the discourse leans more toward empirical grounding or professional/theoretical reasoning.

Fine-grained topics. Identifying specific issues under discussion illuminates agenda formation and problem framing, and supports ex-ante impact analysis (e.g., diagnosing the problem and its

drivers, defining objectives, and mapping options) and ex-post evaluation (e.g., checking effectiveness against the intervention logic and agreed indicators), which are central pillars of contemporary regulatory governance (European Commission, 2023). Topic resolution at this level also helps detect whether downstream analyses regarding "unintended consequences" of standards are truly unintended or were anticipated and debated at the time of standard-setting.

The final motivation is grounded in normative rationale and institutional context. There is no single, context-free yardstick for "accounting standard quality." Classic work on the impossibility of a universal normative ranking of accounting alternatives shows why attempts to order standards without embedding user preferences and institutional context will be incomplete or inconsistent (Demski, 1973). In light of this challenge, I focus on procedural indicators rooted in deliberations: consultation breadth, evidentiary grounding, and specificity of problem framing. These yield measures that are observable, comparable across projects, and theoretically linked to effectiveness and legitimacy. In parallel, standard-setting research and political-economy studies further underscore how international constituency structure, legitimacy concerns, and thin-market politics shape both process and outcomes, reinforcing the value of opening the "black box" of deliberation (Ramanna, 2015; Camfferman and Zeff, 2018).

2.2. Standard-setter Style

Foundational work in behavioral economics shows that individual choices reflect individual preferences, and that preferences themselves can be context-dependent, implying that how evidence is framed and weighed will vary across decision-makers and situations (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). This underscores why individual behavior matters for decision making and varies even under common institutional procedures. Empirically, individual decision-maker effects are large in organizational and policy settings. A classic firm-level study documents sizable manager fixed effects on investment, financing, and organizational policies, demonstrating that "managing with style" can shape outcomes beyond observables (Bertrand and Schoar, 2003). A macro-level research using quasi-random leader transitions shows that national leaders can affect policy and growth, especially when constraints are weak, again pointing to durable individual heterogeneity with real consequences (Jones and Olken, 2005).

Within expert committees, reputational incentives further differentiate behavior. Theory shows

that internal reputations (within-committee standing) and external reputations (outside audiences) create strategic complementarities in preparation and participation, with internal concerns typically stronger and more persistent as committees grow. These forces shape who acquires information, who speaks, and how arguments are presented (Swank and Visser, 2023). Taken together, this literature implies persistent heterogeneity in how individuals marshal and present arguments, even in a common institutional environment, and shows how this individual heterogeneity can affect decision making, reinforcing the need for direct analysis of individual "styles" in standard-setting deliberations.

If individual preferences and reputational incentives operate through speech, then stable speaking styles—what topics board members emphasize, how they justify positions, and which audiences they invoke—are theoretically expected and substantively meaningful. Such styles can persist across projects and align into coalitions or cleavages (e.g., clusters that systematically stress user concerns vs. preparer costs; or empirical evidence vs. conceptual reasoning), providing a micro-foundation for voting blocs and agenda-setting dynamics. As these individual styles interact and endure, system-level patterns emerge. In technical rule-making domains such as accounting, where interests are complex and not purely two-sided, alignments tend to organize around shared expertise and common argumentative frames rather than simple industry blocs. This pattern is consistent with the notion of accounting standard-setting as a "thin political market" in which expertise is concentrated, interests are dispersed, and persuasion within bureaucracies is crucial (Vogel, 2022; Ramanna, 2015). These features help explain how coalitions emerge and cleave within bodies like standard setters.

These insights motivate measuring individual-level discourse features and testing for similarity structures among members. If style is more than situational rhetoric, we should observe (i) persistent, non-zero speaker deviations in stakeholder orientation and in evidentiary posture, and (ii) clusters with similar profiles or clear contrasts. Such evidence links micro-speech to macro-questions of due-process quality and effectiveness in a global standard setter: who gets addressed, what kinds of evidentiary bases are foregrounded, and how coalitions shape outcomes under a legitimacy-oriented due process (Camfferman and Zeff, 2018; OECD, 2021).

3. Methodology

Analyzing automated speech recognition (ASR) transcripts, such as the IASB meeting data used in this study, poses several challenges. The transcripts contain an overwhelming volume of content, often characterized by hesitations, repetitions, fragments, and grammatical errors. To address these issues, this study develops an LLM-powered framework designed to systematize the noisy nature of spoken language through the automated construction of a taxonomy of meeting deliberations. The taxonomy functions as a hierarchical classification system that improves the accessibility and navigability of meeting discussions for users by categorizing claims (i.e., the informational content board members provide to justify or support their positions during IASB meetings) into broad, medium, and detailed levels of granularity.

The framework first applies an LLM-based "gate" to filter out non-substantive content from the IASB meeting transcripts, subsequently infers stakeholder orientations and forms of expertise. It then generates use-case summaries of meeting discussions to seed and iteratively refine a taxonomy of deliberations, and assigns final, fine-grained taxonomy labels across the full sample. The outcome is a structured taxonomy that provides an organized lens through which to examine the informational content of IASB deliberations. In this section, I outline the procedures for my LLM-powered taxonomy generation and text classification.

3.1. Data Source, Unit of Analysis, and Cleaning

The data are drawn from a contemporaneous research (Cui et al., 2025), which employs a novel methodology to construct a comprehensive dataset of IASB deliberations. The authors collect and process all available audio recordings from the IASB's official website, yielding a dataset of 898 recordings that document discussions among 28 board members between 2013 and 2021. The dataset is organized at the level of individual speech segments, defined as uninterrupted speech by a single speaker. Each segment includes the speaker's identity, timestamp, transcript, and other relevant meta data. This granular structure enables precise measurement of participation patterns and speaking dynamics, such as who speaks, when, for how long, and in response to whom, thereby offering unique insights into how expert committees deliberate regulatory policies.

For this study, I focus on transcripts from three pilot topics, namely, Conceptual Framework,

Leases, and Fair Value Measurement, as these represent some of the most central agenda items discussed by the IASB during the sample period (Cui et al., 2025).

The unit of analysis in this study is the individual speech segment. This choice is motivated by the hierarchical design of the taxonomy, which operates across multiple levels of granularity: broad (stakeholder orientations), medium (forms of expertise), and detailed (fine-grained labels). Using a coarser unit, such as aggregated, speaker-meeting level contributions, risks conflating multiple stakeholder orientations or forms of expertise within a single text unit, while also limiting the ability to capture meaningful fine-grained distinctions. Accordingly, each speech segment (i.e., each row in the transcripts) is treated as a separate input record.

Preprocessing involves light text cleaning: removing duplicated words (e.g., "the the"), eliminating common fillers (e.g., "um," "uh," "ah"), normalizing punctuation, and collapsing whitespace. I also exclude very short segments (fewer than 15 words) and remove a small subset of non-target speakers (e.g., technical staff). The output of this step is a segment-level dataset containing both raw and cleaned text, ordered chronologically within meeting topics and dates.

3.2. Filtering for Substantive Content

Policy decision-makers, whether in political or administrative settings, have frequently expressed support for *evidence-based* policy-making (Head, 2013). Accordingly, identifying whether and, to what extent, IASB discussions contain claims that can be objectively checked or verified (e.g., empirical evidence, statistics, or regulatory facts) is itself an interesting and important line of inquiry. At the same time, many professionals prefer the term *evidence-informed* policy-making, emphasizing that decisions are rarely derived solely from objective science. Instead, they often rely on reasoned argumentation that incorporates professional judgments, stakeholder interests, and political contexts (Head, 2013). In this regard, claims grounded in professional or personal experience, or expert judgment, while not strictly verifiable, are equally important to standard-setting. This is particularly true in "thin political markets" such as accounting standard-setting, where expert perspectives are central to deliberations. Capturing both factual claims and expert opinions therefore provides a more complete representation of substantive discussions, better reflecting the reality of standard-setting debates.

Notably, IASB meetings also contain a considerable amount of non-substantive information,

such as procedural matters, turn-taking, greetings, and scheduling. Given that the unit of analysis in this study is the individual speech segment, it is important to account for this feature, as some segments consist entirely of non-substantive talk and would introduce noise if included in the taxonomy construction.

Taking these considerations into account, I employ GPT-4.1 for the first round of claim detection, filtering out non-substantive speech. Specifically, each speech segment is processed by the LLM, which classifies the segment into one of three categories:

1. **Factual claim** (verifiable)
2. **Expert opinion** (non-verifiable but relevant)
3. **Other** (non-substantive)

Prompt templates are provided in Online Appendix Table 1. As a validation step for the LLM's factual claim detection, I apply ClaimBuster, a pioneering model for detecting check-worthy factual claims (Hassan et al., 2017). This model assigns each sentence or paragraph a score indicating its likelihood of being a check-worthy factual claim, with higher scores reflecting stronger check-worthiness. Overall, this filtering stage reduces the dataset to a more manageable subset while ensuring that retained segments are more likely to contain substantive information relevant for taxonomy construction.

3.3. Stakeholder Orientations

An important dimension of IASB deliberations concerns whose perspective or interest is being voiced during the discussions. Theories of regulatory capture and standard-setting politics emphasize that accounting standard setters operate in an environment shaped by the competing interests of distinct stakeholder groups, such as preparers, users, auditors, and regulators (Kothari et al., 2010). Empirical evidence further shows that board deliberations often reflect these stakeholder concerns (Cui et al., 2025). Identifying such orientations is therefore central to understanding how different interests influence the IASB's deliberative process.

The IFRS *Due Process Handbook* explicitly states that the primary objective of the IFRS Foundation is to develop, in the public interest, a single set of high-quality and globally accepted financial

reporting standards. It further specifies that the IASB should operate on principles that encourage stakeholder engagement in the due process, that matters raised by stakeholders are addressed satisfactorily, and that perspectives of a wide range of global stakeholders are fully and fairly considered (IFRS Foundation, 2020). From this perspective, ensuring that various stakeholders' concerns are properly addressed is essential both for the quality and for the legitimacy of the due process.

Motivated by this insight, to capture stakeholder orientations quantitatively, I prompt the LLM to classify each substantive speech segment (i.e., factual claims and expert opinions) into one of six stakeholder groups:

1. **Regulators** (e.g., national standard setters, government agencies)
2. **Accounting professions** (e.g., auditors, and their professional bodies)
3. **Users** (e.g., investors, analysts, creditors)
4. **Preparers** (e.g., companies, CFOs, executives)
5. **Academics** (e.g., professors, researchers)
6. **Public interest organizations** (e.g., NGOs, charities)

The classification prompt instructs the model to identify the primary stakeholder orientation implied in each segment. To evaluate the reliability of these LLM-based classifications, I manually annotate a stratified subsample of 100 segments and compare the results with the model outputs, achieving substantial agreement. The outcome of this step is a stakeholder-oriented mapping of claims across all substantive segments, representing the broad level of my hierarchical taxonomy. This enables subsequent analyses of how particular stakeholder concerns emerge, disappear, or dominate during board discussions.

3.4. Forms of Expertise

While stakeholder orientation captures whose interests are invoked, another critical aspect of IASB deliberations concerns the source of authority or knowledge, that is, what evidentiary basis speakers primarily rely on in their reasoning. Whereas "factual claims" are self-explanatory as a source of authority, "expert opinion" is a broader and less clearly defined category, since expertise can

be acquired in different ways. Inspired by prior work in argumentation theory and accounting standard-setting research (Wagemans, 2011; Bradbury and Harrison, 2014), I therefore further classify "expert opinion" into sub-categories based on the forms of expertise underpinning board members' justifications (i.e., personal experience, technical knowledge, or theoretical background). Distinguishing among these subtypes of expert opinion sheds light on how board members establish credibility and justify their positions.

A further rationale for this design choice derives from recent calls in the argumentation literature. Visser et al. (2020) emphasize that identifying conventional patterns of reasoning is essential for interpreting and evaluating argumentation, and that gaining insight into its actual use in communicative practice is crucial. Yet, large and systematically annotated corpora of argumentative discourse remain scarce. Addressing this gap, the design adopted here provides a large corpus of actual argumentative discourse annotated with theoretically grounded concepts, and an LLM-powered, quantitative approach that can be extended to other institutional settings.

To implement this step, I prompt the LLM to classify each "expert opinion" segment into one of four sub-categories:

1. **Experiential claim** (based on prior experience or past practices)
2. **Technical claim** (based on technical knowledge or domain-specific know-how)
3. **Theoretical claim** (based on accounting theory, the conceptual framework, or higher-level principles)
4. **Other claim** (where no clear source of authority is present)

To validate the results, I manually annotate a subsample and compare these annotations with the LLM outputs, finding that the model reliably distinguishes between experiential, technical, and theoretical reasoning. The outcome of this stage is a medium-level taxonomy of claim types, which complements the stakeholder dimension by clarifying the forms of expertise through which board members ground their arguments. Taken together, the two dimensions of stakeholder orientations and forms of expertise structure the informational content of IASB deliberations into an interpretable framework.

3.5. Fine-grained Topics

Building on the broad- and medium-level taxonomy developed in the preceding steps, this stage focuses on generating the detailed, fine-grained topics. Transforming unstructured text into structured categories is a key step in text mining. Yet most existing methods for building label taxonomies and classifiers depend heavily on domain expertise and manual curation, making them costly and time-consuming (Wan et al., 2024). These difficulties are especially pronounced when the label space is unclear, as in the IASB context examined here.

To address these challenges, I adopt the TnT-LLM framework, a novel approach that employs LLMs to automate end-to-end label generation and assignment with minimal human input (Wan et al., 2024). Specifically, I use a zero-shot, multi-stage reasoning procedure, which enables LLMs to iteratively generate and refine the label taxonomy, to construct a fine-grained taxonomy of IASB meeting discourse.

3.5.1. Use-case Summary

As noted earlier, a persistent challenge with ASR transcripts is their inconsistency: speakers often repeat the same idea, embed multiple thoughts within a single utterance, or express themselves in vague terms. To confront this problem, I generate concise and informative summaries of each substantive segment. This step of the TnT-LLM framework is inspired by the classic mixture-model clustering process (McLachlan and Basford, 1988), but implemented in a prompt-based manner. This stage reduces both the size and variability of the input segments while extracting the aspects most relevant to the use case, which is particularly important when label spaces are not evident from surface-level semantics (Wan et al., 2024).

Concretely, the LLM is prompted to produce a short summary of each substantive segment, with explicit instructions regarding the intended use case and a target summary length. Deterministic decoding (temperature = 0.001) and a structured output schema are employed to ensure consistent parsing. The prompt emphasizes extracting the key informational content, that is, what the segment is about and how the speaker justifies their views. This summarization step is analogous to the featurization stage in classic machine learning, where raw text inputs are projected into a vector space via a feature transformation such as an embedding model. Here, the output summary of each segment functions as a concise and informative feature representation of the original text.

The summaries serve two primary purposes. First, they normalize noisy speech into concise and homogeneous inputs suitable for clustering and category definition. Second, by focusing the model on grounds and rationales (rather than surface phrasing), they mitigate spurious topical drift during taxonomy generation and refinement.

3.5.2. Taxonomy Creation, Iterative Update, and Review

I next create and refine a label taxonomy using the summaries from the previous step, following the TnT-LLM framework. Specifically, I begin by splitting the summary corpus into equal-sized minibatches of 400 summaries each, balancing the desired granularity of the taxonomy against the token limits of the API. I then process these minibatches with three types of zero-shot LLM reasoning prompts in sequence. First, an initial generation prompt takes the first minibatch and produces an initial label taxonomy. Second, a taxonomy update prompt iteratively refines the intermediate taxonomy with subsequent minibatches. In each update step, the model performs three tasks: (i) evaluating the taxonomy against the new data, (ii) identifying issues and suggesting improvements, and (iii) modifying the taxonomy accordingly. I conduct 10 such updates, ensuring that the sample (approximately 50% of the corpus) is large enough to capture the diversity of the corpus while avoiding unnecessary computational costs.

This taxonomy creation and update stage of the TnT-LLM framework parallels prompt optimization using Stochastic Gradient Descent (Pryzant et al., 2023): the generation prompt initializes the taxonomy, which is then iteratively "optimized" through the chain of update prompts. In each round, the fit of the taxonomy to the minibatch is assessed, errors are analyzed, and refinements are introduced in a process analogous to backpropagation. Third, after the specified number of updates, an independent review prompt evaluates the candidate taxonomy for formatting and quality, and where appropriate, suggests minor edits to improve clarity and reduce overlap. In all three prompt types, I supply detailed use-case instructions specifying the goals and desired output format (e.g., number of labels, target length of label descriptions) together with the minibatch.

Finally, I conduct a human-in-the-loop calibration of the reviewed taxonomy. This step harmonizes near-synonymous labels, resolves residual overlaps, and clarifies ambiguous descriptions. Edits are intentionally conservative and fully documented (with versioning of category tables and change notes). The outcome is a finalized detailed-level taxonomy that is sufficiently granular for content analysis while remaining practical for large-scale assignment (see Online Appendix Table 2).

3.6. Label Assignment Using the Detailed-level Taxonomy

With the detailed-level taxonomy finalized, I assign a single primary label (i.e., the fine-grained topic) to each substantive segment. Specifically, for each segment the LLM receives (i) the segment summary and (ii) the full taxonomy table. The instruction explicitly prohibits multi-labeling, since prior research shows that LLMs perform best on single-choice tasks, in which they must indicate a clear preference, whereas they often struggle with multiple-choice settings (Wan et al., 2024). If no reasonable match is found, the model returns a sentinel value, which I use diagnostically during spot checks. Deterministic decoding (temperature = 0.001) is applied throughout to ensure reproducibility.

Unlike Wan et al. (2024), I do not train a lightweight classifier for label assignment. Instead, I rely directly on LLMs to produce the labeled corpus for the full sample, for three reasons. First, preliminary experiments with embedding-based clustering revealed strong sensitivity to "noise," and a lightweight classifier built on such representations risks propagating this instability.[4] Second, the corpus size in this study makes end-to-end LLM labeling cost-feasible. Third, recent studies demonstrate the effectiveness of LLMs as text annotators (Gilardi et al., 2023; Lee et al., 2023).

After assigning fine-grained labels to the full set of IASB meeting transcripts, I perform several quality checks. I manually audit a stratified subsample of assignments to assess disagreement patterns and test stability under small prompt perturbations and across random seeds (with decoding held deterministic). These diagnostics indicate strong internal consistency of the labeled corpus.

With this fine-grained label assignment complete, I finalize the hierarchical taxonomy of IASB deliberations, which then provides the basis for the descriptive analyses presented in the next section.

[4]In preliminary experiments with the dataset, I observed that under traditional embedding-based clustering approaches, a substantial number of speech segments were excluded as "noise." Upon inspection, this was largely due to the noisy nature of spoken language in ASR transcripts, where variations in wording, phrasing, length, or idiosyncratic claims made segments appear "unique" to the embedding models. By contrast, LLM-generated use-case summaries serve as feature representations of the original segments. This summarization step is analogous to the featurization process in classic machine learning, but it performs more effectively here, as the LLM can normalize noisy speech into concise, homogeneous inputs suitable for clustering.

4. Statistics and Metrics

As a private organization, the legitimacy of the input, that is, whether the input received reflect the views of all stakeholders involved, is a central issue for the IASB's acceptance as a global standard setter (Jorissen et al., 2013). Since the content of deliberations is the information that flows into the official IASB meetings (i.e., the input), a clear map of the discourse—who is invoked, which forms of expertise/claims are mobilized, and which topics are addressed—together with meeting-level deliberation metrics (range, depth, and balance/polarization) provides interpretable benchmarks for the procedural quality and effectiveness of the standard-setting process. This section first reports descriptive statistics for the three taxonomy layers and then defines the meeting-level deliberation metrics.

4.1. Stakeholder Orientations

The stakeholder-orientation layer in the taxonomy identifies whose perspective or interest is being invoked in board deliberations, a central element of the IASB's due process and a long-standing theme in the literature on the politics of standard-setting (Camfferman and Zeff, 2007; Gipper et al., 2013; Flores et al., 2025). It provides a direct way to assess whether the discourse appears pluralistic or skewed toward particular constituencies.

The statistics in Table 1 show that board members regularly invoke multiple stakeholder perspectives. In the full sample, for instance, Preparers account for 4,048 of 8,758 substantive segments (≈46%), Users 2,428 (≈28%), Regulators 844 (≈10%), Accounting Professions 731 (≈8%), Academics 272 (≈3%), and Public Interest 18 (< 1%), with a residual Others category (417; ≈5%). Preparers are the modal orientation, reflecting the technical nature of IASB deliberations, which center on preparer-facing technical judgments and implementation details. Yet no single stakeholder group, including the accounting professions, dominates the discourse; the mix of voices is broader than commonly presumed (see Figure 1). Meeting-level and speaker-level distributions, together with speaker-meeting cells, show substantial dispersion rather than concentration (Table 2 and 3), and the time-series panels indicate that inclusiveness fluctuates but is persistent over calendar time (Panel A of Figure 4).

While prior literature often emphasizes the prominent role of accounting professions (e.g., large

public accounting firms) and related regulatory actors in accounting standard-setting networks and governance (Zeff, 2005; Allen, 2018), the content-based evidence here does not indicate single-group dominance within board-room discourse.

4.2. Forms of Expertise

The form-of-expertise layer in the taxonomy captures the source of authority or knowledge (the evidentiary basis) invoked in board deliberation, i.e., factual assertions versus expert judgment (experiential, technical, theoretical), a distinction emphasized in the policy and argumentation literature as central to how deliberation justifies choices (Head, 2008; Wagemans, 2011). These labels allow the analysis to assess whether the discourse skews toward evidence presentation or toward professional judgment and reasoning, a key aspect of evidence-based (or evidence-informed) policymaking (Head, 2013).

The statistics in Table 1 indicate that expert-judgment claims are more common than purely factual assertions. In the full sample, factual claims constitute roughly 14% of substantive segments (1,194), whereas experiential, technical, and theoretical claims together comprise about 55% (1,051; 1,532; 2,190, respectively), with the remainder coded as Other claim (2,791). Figure 2 shows that composition differs across standards: for example, theoretical claims are relatively more frequent in the Conceptual Framework, whereas technical claims feature more prominently in Leases. Meeting-, speaker-, and speaker-meeting level summaries further indicate wide variation in expertise mixes across individuals and contexts (Table 4 and 5); time-series views show sustained reliance on expert judgment relative to purely factual assertions (Panel B of Figure 4).

Taken together, these facts suggest an evidence-informed process in which professional judgment and structured reasoning play a larger role than purely factual recitation in the IASB standard-setting deliberations.

4.3. Fine-grained Topics

The fine-grained topic layer describes the substantive issues the Board addresses within meetings (e.g., definitional clarity, recognition/measurement choices, disclosure objectives, scope and transition). This layer is useful for characterizing thoroughness and for contextualizing how stakeholder appeals and forms of expertise are deployed across concrete sub-issues.

The statistics in Table 1 show that prominent categories in this layer include Definitional clarity and ambiguity, Measurement-basis selection, Information completeness and transparency, and Process efficiency and simplification. In the full sample, Definitional clarity and ambiguity accounts for 2,264 of 8,758 substantive segments (≈25.9%), Measurement-basis selection 536 (≈6.1%), Information completeness and transparency 345 (≈3.9%), and Process efficiency and simplification 317 (≈3.6%). Figure 3 shows the Top-20 fine-grained topics, which together accounts for approximately 71% of all labeled segments, indicating a focused yet diverse agenda.

A further motivation for this layer is to inform debates about so-called "unintended" consequences of standards (e.g., see Brüggemann et al. (2013) for a review). Much empirical work attributes certain post-implementation outcomes to unintended effects, but without verifiable deliberation records it is hard to know whether those outcomes were unanticipated ex ante or instead acknowledged trade-offs accepted during standard-setting. By organizing meeting transcripts into a hierarchical taxonomy with multiple levels of detail, this study provides an empirical basis to audit the agenda: researchers and reviewers can check whether specific concerns were raised, how prominently, and in what argumentative form (factual vs. experiential/technical/theoretical), thereby distinguishing unanticipated side effects from explicitly debated compromises.

Concretely, Online Appendix Table 2 lists the full dictionary of detailed topics covered in the sample IASB meetings. This resource enables targeted verification, for example, whether issues related to transition relief, non-controlling interests, or measurement-basis selection were discussed for a given project and time window, and with what frequency.[5]

4.4. Deliberation Metrics—Range, Depth, and Balance (Polarization)

This section formalizes three complementary meeting-level metrics that summarize *how many* perspectives/issues are substantively present and *how evenly* they are treated. The measures follow standard practice in diversity and concentration analysis (Laakso and Taagepera, 1979).

Range (inclusiveness). This metric reflects the breadth of viewpoints brought into the deliberations, and is defined as the *count of categories* (of stakeholder orientations and forms of expertise)

[5]Upon completion of the project, the three-layer taxonomy used in this paper (i.e., stakeholder orientations, forms of expertise, and fine-grained topics) will be integrated into the IASB-deliberations dataset (Cui et al., 2025) and deposited on the Open Science Framework (OSF). Researchers will be able to query the corpus to verify whether specific issues were discussed, when they were discussed, and in what argumentative form.

whose within-meeting share meets or exceeds a prevalence threshold, $\tau = 10\%$:

$$\text{Range}_\tau \;=\; \sum_{k=1}^{K} \mathbf{1}\{p_k \geq \tau\}.$$

where K denote the *number of categories represented in a meeting* (e.g., Accounting Professions, Factual Claims), and p_k denote the *within-meeting share of segments* assigned to category $k \in \{1, \ldots, K\}$. The 10% rule ensures that categories are *substantively represented*, rather than appearing only sporadically. Panels A–B of Figure 5 plot the resulting time series of inclusiveness metrics for stakeholder orientations and forms of expertise, respectively.

Depth (thoroughness). This metric captures the depth of exploration within technical sub-issues addressed in a meeting. It is measured as the *count of distinct fine-grained topics* that appear at least once on that meeting. Panel C of Figure 5 reports the resulting time series of this metric: larger values indicate that the Board canvassed a wider set of technical issues within the session.

Balance (and polarization). To characterize how evenly discussion is distributed across categories, I use the Herfindahl–Hirschman Index (HHI), also known as the Simpson's concentration index. This index is computed from the within-meeting shares p_k:

$$\text{HHI} = \sum_{k=1}^{K} p_k^2.$$

Because the number of potential categories differs across classification systems (e.g., six stakeholder groups vs. four expertise types) and some meetings do not feature all categories, I normalize HHI to enable comparability:

$$\text{HHI}^* = \frac{\text{HHI} - \frac{1}{K}}{1 - \frac{1}{K}},$$

where K is the number of *represented* categories in that meeting.[6] I then report:

$$\text{Balance} = 1 - \text{HHI}^* \quad \text{and} \quad (\text{Polarization} = \text{HHI}^*).$$

[6]This linear rescaling sets the *even case* (equal shares across K represented categories, where $\text{HHI} = 1/K$) to 0 and the *monopoly case* ($\text{HHI} = 1$) to 1, yielding a unit-free $[0, 1]$ scale: higher HHI^* indicates greater concentration ("polarization"), while $1 - \text{HHI}^*$ indicates greater evenness ("balance"). The edge case $K = 1$ (only one represented category) is handled by definition with $\text{HHI}^* = 1$, hence Balance $= 0$. This normalization is conventional in applied work; see Owen and Owen (2022).

Panels A–B of Figure 6 plots balance indices for stakeholder orientations and forms of expertise, respectively. Sustained high balance signals multi-sided discussion, whereas temporary dips (spikes in polarization) indicate sessions where one perspective or claim type dominates.

Overall, Figure 5 shows that IASB meetings typically include multiple stakeholder perspectives and claim types and cover several fine-grained issues per session; Figure 6 indicates that within-meeting representation is often balanced, with episodic concentration around specific perspectives or forms of expertise. This pattern aligns with a pluralistic but structured deliberative process documented elsewhere in the paper.

5. Descriptive Evidence

This section documents three sets of descriptive results. First, I assess cross-speaker "style" in how board members allocate discussion across stakeholder orientations and forms of expertise. Second, I compare speakers' distributional profiles using pairwise correlations. Third, I examine the association between stakeholder orientations and forms of expertise. Taken together, these exercises yield a compact map of (i) stable individual communication patterns that matter in deliberative committees, (ii) areas of similarity and contrast that can foreshadow coalitions or cleavages, and (iii) audience-targeted argumentative choices that bear on due-process legitimacy and responsiveness. This approach builds on prior evidence that textual features of committee deliberations are informative about individual tendencies and group dynamics and that similarity in topic profiles is a useful proxy for latent alignment (Hansen et al., 2018), and on work showing that argumentation is shaped by the audiences it seeks to address (Palmieri and Mazzali-Lurati, 2016).

5.1. Standard-setter Style

I begin with a variance-components linear mixed model (LMM) to capture latent standard-setter "style" in IASB deliberations.[7] This model allows me to assess whether and, by how much, there is

[7] Variance-components LMMs model the outcome as a fixed-effects component plus random effects associated with grouping factors. The variances of these random effects—together with the residual variance—are the "variance components." Estimating these variance components (commonly via Restricted Maximum Likelihood, REML) partitions total variability across the modeled sources and is well defined for unbalanced designs (Corbeil and Searle, 1976; Harville, 1977).

stable between-speaker heterogeneity in how board members allocate attention across stakeholder orientations and forms of expertise.

Concretely, I aggregate speech segments to the Speaker×Standard×Meeting-Date unit.[8] For each category (e.g., Accounting Professions; Factual Claim), I estimate a separate model in which the outcome is that category's share within the unit (its segment count divided by the unit's total segments). Each model includes (i) a fixed intercept capturing the overall mean of the category share and (ii) random intercepts for the grouping factors—Speaker, Standard, and Meeting-Date.[9] Thus, for each category, the model partitions between-unit variability in the category share into components attributable to Speaker, Standard, Meeting-Date, and the residual.

For a given category c (e.g., *Accounting Professions* or *Factual Claim*), let $y_u \in [0,1]$ be the share of that category in unit $u \equiv (i,s,d)$ (Speaker i × Standard s × Meeting-Date d). Stacking the n outcomes into $\mathbf{y} \in \mathbb{R}^n$, I estimate the linear mixed model

$$\underbrace{\mathbf{y}}_{n\times 1} = \underbrace{\mathbf{X}\beta}_{\text{fixed (intercept)}} + \underbrace{\mathbf{Z}_S\mathbf{b}_S + \mathbf{Z}_T\mathbf{b}_T + \mathbf{Z}_D\mathbf{b}_D}_{\text{random effects}} + \boldsymbol{\varepsilon},$$

with independent Gaussian random effects

$$\mathbf{b}_S \sim \mathcal{N}(\mathbf{0}, \sigma_S^2\mathbf{I}), \qquad \mathbf{b}_T \sim \mathcal{N}(\mathbf{0}, \sigma_T^2\mathbf{I}), \qquad \mathbf{b}_D \sim \mathcal{N}(\mathbf{0}, \sigma_D^2\mathbf{I}), \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2\mathbf{I}_n).$$

Here $\mathbf{X} = \mathbf{1}_n$ is the $n \times 1$ vector of ones (a fixed intercept, i.e., the grand mean); $\mathbf{Z}_S, \mathbf{Z}_T, \mathbf{Z}_D$ are incidence matrices for *Speaker*, *Standard*, and *Meeting-Date*; $\mathbf{I}$ denotes the identity matrix of appropriate dimension for each random-effect block, and $\mathbf{I}_n$ is the $n \times n$ identity for the residuals. The variance components $\{\sigma_S^2, \sigma_T^2, \sigma_D^2, \sigma_\varepsilon^2\}$ are estimated by REML and reported as shares of total variance.[10]

[8] Units with fewer than five segments are excluded. To improve precision, I restrict to the top 10 speakers by total segment count, covering 86.2% of all segments.

[9] In variance-components linear mixed models, each random factor is modeled as an independent, mean-zero Gaussian source of variation. When factors are algebraically related (e.g., an interaction/product of others) or their crossings are sparsely replicated, the corresponding variance components are only weakly identified; ML/REML fits then frequently allocate variance to one component while driving others to boundary (≈0) estimates, producing unstable and hard-to-interpret partitions (Searle et al., 2009; Bates et al., 2015). For interpretability, I therefore focus on Speaker, Standard, and Meeting-Date in the main REML specification; richer structures (e.g., Speaker×Standard, Standard×Meeting-Date) are examined in the fixed-effects ΔR^2 diagnostics in Section 6.

[10] REML is an estimation method for linear mixed models that adjusts for the degrees of freedom used by the fixed effects, thereby reducing the small-sample downward bias that ML exhibits for variance components (Patterson and Thompson, 1971).

Table 6 reports the estimated results. Across categories, the Speaker component is economically meaningful in two cases: *Accounting Professions* (13.36%) and *Factual Claim* (13.74%); the remainder is absorbed mainly by the Residual, with smaller contributions from Standard and Meeting-Date.[11] By contrast, *Preparers* and *Theoretical Claim* are topic-driven (Standard share ≈45.2% and ≈42.9%) with relatively small Speaker components. These patterns indicate that individual style is most evident for *Accounting Professions* and *Factual Claim*, after modeling context (Standard and Meeting-Date), which motivates using these two categories as running examples in the individual-level analyses that follow.

To illustrate individual speaker-level tendencies, I begin with no-pooling caterpillar plots of raw speaker shares (the fraction of each speaker's segments that fall into the category), with Wilson 95% intervals. As shown in Figure 7, several speakers sit materially above or below the grand mean (the overall average share across all speakers) in both *Accounting Professions* and *Factual Claim*; intervals are wider for low-volume speakers, as expected from finite-sample variability. Taken together, these patterns are consistent with between-speaker heterogeneity rather than complete homogeneity.

I then consider partial pooling by estimating the BLUPs value for each speaker. In a linear mixed model, a BLUP is the best linear unbiased predictor of a random effect. In this application, using the same REML-fitted specification as above, BLUPs are speaker-specific deviations from the model baseline (the overall mean of the category share) estimated via partial pooling.[12,13] In brief, REML provides population-level variance components (one per grouping factor), whereas BLUPs provide group-specific random-effect predictions, i.e., one for each observed level of a grouping factor (e.g., a per-speaker deviation). Thus, REML answers "how much of the total variance is attributable to each source?", while the speaker BLUP answers "how much does a given speaker

[11] Each component's variance is reported as a share of total variance (row sums=100%). Shares correspond to: Speaker (stable between-speaker differences—"style"); Standard (differences across standards); Meeting-Date (calendar-time shocks common to all speakers and standards on that date); and Residual (everything else, including idiosyncratic noise and finite-sample variability). As a rule of thumb, larger shares indicate the component explains more variation in the outcome.

[12] With variance components estimated by REML, the resulting predictions are often called empirical BLUPs (Montesinos López et al., 2022).

[13] Partial pooling models each speaker with its own random effect drawn from a common distribution; estimates are shrunk toward the model baseline, with stronger shrinkage for speakers with fewer observations. This makes BLUP estimates *more conservative*—they temper extremes from sparsely observed speakers—so reported style patterns are less likely to be noise. By contrast, no pooling estimates each speaker entirely from its own data (e.g., raw proportions with Wilson intervals), so intervals are wider for low-volume speakers (Robinson, 1991; Brown et al., 2001; Gelman and Hill, 2007).

differ from the model baseline?".

Figure 8 reports these speaker random-effect estimates. Each point is a speaker's BLUP, an estimated speaker-specific deviation from the model baseline. Accordingly, the zero line denotes no speaker-specific deviation; positive (negative) values indicate that the speaker tends to emphasize the category more (less) than the average speaker.[14] For example, a BLUP of +0.02 means the speaker's expected share in the category is two percentage points higher than the model baseline (e.g., if the baseline is 27%, the prediction is 29%). Despite partial pooling, several speakers exhibit persistent nonzero deviations, suggesting that the differences are not purely sampling noise.

Finally, the heatmaps in Figure 10 (column-wise z-scores of category shares at the level of Speaker×Standard) show that speaker-style differences are often not confined to a single standard: some members place persistently higher emphasis on certain categories across multiple standards, while others are frequently below their peers. This pattern aligns more closely with portable style than with single-topic idiosyncrasy. For example, in Panel B, Sue Lloyd's z-scores exceed +1.5 across all displayed columns; because standardization is by column, this means that within each Standard, relative to the distribution of all speakers, she incorporates objective, verifiable evidence in her deliberations more than 1.5 standard deviations above the average speaker. This demonstrates a consistently fact-forward style across standards.

While the caterpillar plots (raw shares and BLUPs) document individual speaker style, the heatmaps complement them by showing how these tendencies extend across standards—that is, who stands out on which standard. Together, these displays complement the mixed-model variance decomposition by revealing the patterns underlying the reported variance shares. Overall, the evidence indicates (i) non-trivial between-speaker heterogeneity in several categories (e.g., *Accounting Professions* and *Factual Claim*), and (ii) portable within-speaker style across standards.

5.2. Correlations Among Standard Setters

Do speakers resemble each other in their usage patterns of stakeholder appeals and forms of expertise? Prior work shows that text-based profiles of deliberation can proxy latent alignment within committees and legislatures, helping map blocs and internal dynamics (e.g., topic/speech-based

[14] By construction of the random-intercept model in this section, $\mathbb{E}[b_S] = 0$, where b_S denotes the speaker random effect. Hence the model's average speaker effect equals zero, and the zero line corresponds to that average (i.e., no speaker-specific deviation).

similarity and positioning, see Lauderdale and Herzog (2016) and Hansen et al. (2018)).

To gauge the potential alignment, I compute pairwise Pearson correlations between each speaker's distributional profile: vectors of shares (not counts) across the stakeholder-orientation categories and, separately, across the forms-of-expertise categories.[15] Correlations on shares capture mix similarity rather than speaking volume: values near +1 indicate very similar mixes; values near 0 little relation; negatives indicate opposing emphasis patterns.

Stakeholder orientations. Table 8 shows consistently high positive correlations (many near 0.9–1.0), indicating that most speakers allocate attention across stakeholder groups in very similar proportions. In other words, at the coarse level of "whose interests are invoked?", speakers largely track one another.[16]

Forms of expertise. By contrast, Table 9 exhibits greater dispersion, including several moderate negatives, which suggests that there are meaningful differentiation in how arguments are substantiated (e.g., heavier vs. lighter reliance on factual claims, technical analysis, experiential justification, or theoretical argument).

Overall, high alignment on stakeholder targeting suggests shared normative expectations about to whom arguments should be addressed, whereas dispersion in forms of expertise points to heterogeneity in evidence-weighting, i.e., how members substantiate claims (fact-forward, technical, experiential, or theoretical).

5.3. Stakeholder Orientations versus Forms of Expertise

Do particular stakeholder appeals tend to be paired with particular forms of expertise? Argumentation research emphasizes that speakers tailor arguments to stakeholder audiences in public/organizational communication, and such tailoring bears on perceived legitimacy and responsiveness of the process (Palmieri and Mazzali-Lurati, 2016).

To test for the association, Table 10 cross-tabulates raw counts (Stakeholder Orientation $\times$ Forms of Expertise) and reports a Pearson χ^2 test of independence. The association is statistically

[15] Speaker profiles with zero variance (all mass in a single category) yield undefined pairwise correlations; those entries are set to zero ("no measurable similarity"). I exclude the categories "Others" (Stakeholder Orientations) and "Other Claim" (Forms of Expertise) and, for precision, restrict to the Top-10 speakers by segment count.

[16] These correlations are computed on speaker-level compositions and thus reflect overall mix similarity; they can remain high even when within-context differences in specific categories are economically meaningful (documented in Section 5.1)

significant ($\chi^2 = 635.642$, df $= 15$, $p \approx 9.4 \times 10^{-126}$). Because χ^2 scales with sample size, I also report Cramér's $V = 0.187$ (bias-adjusted), which indicates a moderate association by conventional benchmarks (rule of thumb for a 6×4 table: small ≈ 0.06, medium ≈ 0.17, large ≈ 0.29, see MRC Cognition and Brain Sciences Unit (2021)).[17]

Significance alone does not reveal which pairs drive dependence. Figure 11 therefore plots a heatmap of standardized Pearson residuals from the cross-tabulation of stakeholder orientations (rows) by forms of expertise (columns). Several patterns stand out: *Preparers* are strongly over-paired with *Technical* claims ($R = +9.75$) and under-paired with *Theoretical* ($R = -11.53$); *Users* and *Academics* are strongly over-paired with *Theoretical* ($R = +9.69$ and $R = +9.35$) and under-paired with *Technical* ($R = -7.29$ and $R = -6.85$). *Regulators* are modestly more factual/theoretical and less technical.[18]

To sum up, while the global Cramér's V is modest, specific combinations show pronounced over- or under-representation. In terms of due process, such audience-specific evidentiary choices are consistent with responsive, stakeholder-aware deliberation (e.g., technical detail when addressing Preparers) and align with the audience-design logic highlighted in the argumentation literature (Palmieri and Mazzali-Lurati, 2016).

6. Robustness

This section evaluates the robustness of the main descriptive evidence and modeling choices. First, I replicate the "standard-setter style" analysis using a fixed-effects (FE) variance decomposition via a sequential ΔR^2 ladder, which provides an order-conditional allocation of in-sample fit by blocks of factors. Second, I turn to individual-level checks using an alternative display and specification. Together, these exercises assess whether speaker-specific tendencies remain visible under alternative model specifications and after absorbing increasingly rich contextual structure.

[17] Under H_0 (independence), χ^2 has mean df and SD $\sqrt{2\,\text{df}}$; with df $= 15$, $\chi^2 = 635.642$ lies ~ 116 SD above the mean, making independence implausible (Agresti, 2007). Because χ^2 is sensitive to N, I also report Cramér's V, a normalized effect-size measure based on χ^2; it ranges from 0 (no association) to 1 (perfect association). I exclude the categories "Others" (Stakeholder Orientations) and "Other Claim" (Forms of Expertise).

[18] Standardized residuals flag cells that deviate most from independence; values with $|R| \gtrsim 2$ are typically noteworthy for lack of fit of H_0 in that cell (with the direction indicated by the sign) (Agresti, 2007).

6.1. Standard-setter Style—Variance Decomposition with Fixed Effects

To complement the REML variance-components results in Section 5.1, I estimate a sequence of weighted least squares (WLS) fixed-effects models that decompose variation in the outcome using a sequential ΔR^2 ladder.[19] The outcome is the category share at the Speaker × Standard × Meeting-Date unit (the category's segment count divided by the unit's total segments), and weights equal the unit's segment count.

I add FE blocks in the following order: Standard → Meeting-Date → Standard×Meeting-Date → Speaker → Speaker×Standard. Standard FEs capture between-standard (topic) differences in the mix of stakeholder appeals and forms of expertise (e.g., *Theoretical Claim* may systematically feature more in Conceptual Framework than in Leases). Meeting-Date FEs absorb date-specific shocks common across standards discussed on the same meeting date. The Standard×Meeting-Date interaction captures time-specific deviations by standard (topic-by-date shocks). The remaining speaker variation is then added in two steps: Speaker FEs (stable differences across individuals, i.e., a style component that does not depend on the topic) and Speaker×Standard FEs (speaker-specific departures that are topic-contingent). This order-conditional decomposition separates overall stable speaker style from topic-contingent speaker effects.

Table 7 reports the incremental R^2 at each step (M1–M5) and the remaining unexplained share ("Residual"). For stakeholder orientations, most explanatory power is absorbed by context (Standard, Meeting-Date, and their interaction), while the Speaker blocks add small-to-moderate increments in selected categories (e.g., *Preparers*, *Users*, *Regulators*, *Accounting Professions*). For *Accounting Professions*, cumulative fit reaches $\approx 36.7\%$ by M3 (Standard + Meeting-Date + Standard×Meeting-Date); the Speaker block (M4) adds $\approx$ 5.1 percentage points and Speaker×Standard (M5) adds $\approx$ 2.5 percentage points, leaving $\approx 55.6\%$ unexplained. For forms of expertise, *Theoretical Claim* is strongly standard-driven (M1 $\approx 42.8\%$), whereas *Factual Claim* shows a sizeable Speaker increment (M4 adds $\approx$ 17.1 percentage points), echoing the REML finding that speaker-level style matters for factual assertions. Across categories, the Speaker block contributes roughly 2.7–17.1 percentage points, and Speaker×Standard contributes about 1.2–6.3

[19] The REML variance-components model asks, "How much of the total variance is attributable to each random factor (e.g., speaker, standard, date)?"—a partitioning of total variance, see Harville (1977), whereas the FE ΔR^2 ladder asks, "How much additional variation is explained after adding a given block of dummies?"—a sequential in-sample fitting, see Grömping (2006). The two views are complementary.

percentage points, indicating that the style component is primarily a stable across-topic difference rather than a purely topic-contingent rhetorical device.

To summarize, the FE decomposition aligns with the REML variance-components picture: context (topic and date) explains a large share of variation, but a stable, speaker-specific style remains visible in several categories even after absorbing rich contextual structure.

6.2. Standard-setter Style—Individual-Level Analysis

To complement the individual-level displays (raw proportions and BLUPs) in Section 5.1, I estimate context-adjusted speaker tendencies using a fixed-effects specification that absorbs meeting context and yields directly comparable speaker coefficients, following the practice in legislative-speech literature (Osnabrügge et al., 2021). Specifically, for each category, the outcome is the category share at the Speaker × Standard × Meeting-Date unit, and I fit a weighted least squares (WLS) regression with weights equal to the unit's segment count. The model includes fixed effects for Standard, Meeting-Date, and their interaction (Standard×Meeting-Date), as well as Speaker fixed effects.

Figure 9 plots the estimated *Speaker coefficients*. For interpretability, I re-center the coefficients so that the dashed vertical line at zero marks the *precision-weighted mean across speakers* after absorbing the contextual fixed effects. Because all Speaker effects are estimated relative to this single common baseline, values are directly comparable across individuals: positive (negative) coefficients indicate greater (less) emphasis on the category than the average speaker, conditional on topic and timing. For example, a value of 0.10 in Panel B means the speaker allocates 10 percentage points more of their speech to *Factual Claim* than the average speaker, holding the Standard×Meeting-Date context fixed.

Several speakers exhibit material deviations in both Panel A (*Accounting Professions*) and Panel B (*Factual Claim*). These patterns align with the variance-components results (Table 6) and the FE ΔR^2 ladder (Table 7): even after absorbing rich contextual structure, a stable, speaker-specific style remains visible. The direction of the speaker effects also lines up with the raw proportions (Figure 7) and with the partially pooled BLUPs (Figure 8): speakers who sit above (below) the average in the raw display typically remain above (below) after conditioning on topic and timing, and partial pooling shrinks but does not erase those differences.

7. Conclusions

IFRS Accounting Standards function as a global financial language, used or required in more than one hundred jurisdictions. Given that the IFRS's due process renders formal meeting deliberations central to procedural legitimacy, the official IASB meetings are crucial for shaping global accounting standards, thereby influencing capital markets worldwide. Yet systematic, process-level evidence on the quality and effectiveness of standard-setting deliberations (e.g., *what is argued, how, and for whom*) has been scarce. I address this gap by constructing a multi-level taxonomy of IASB discourse, encompassing stakeholder orientations, forms of expertise, and fine-grained topics. The taxonomy provides a transparent map from unstructured dialogue to interpretable constructs that speak to input breadth, evidentiary grounding, and specificity of problem framing.

Methodologically, the study delivers a reusable, scalable framework that transforms meeting transcripts into structured data. I first filter for substantive claims, then classify each segment by stakeholder orientation and by evidentiary basis, and finally assign detailed topic labels using an LLM-guided, iteratively refined taxonomy. The pipeline yields a labeled corpus and a set of meeting-level metrics (range, depth, and balance/polarization) that can be replicated and extended to other deliberative settings.

Three salient patterns stand out. First, deliberations are pluralistic in audience: users, preparers, regulators, and the accounting professions are all regularly invoked; no single group dominates the room. Meeting-level measures show that multiple perspectives typically exceed substantive thresholds within a session, and balance indices indicate that representation is often diffuse rather than concentrated. Second, the evidentiary posture is best described as evidence-informed: factual assertions comprise roughly one-seventh of substantive segments, whereas experiential, technical, and theoretical claims together account for a majority, with patterns varying across standards (e.g., relatively more theoretical argument in the Conceptual Framework). Third, there is economically meaningful heterogeneity across speakers. Variance-components models attribute $\approx$ 13–14% of the variation in two focal categories (*Accounting Professions* and *Factual Claim*) to stable between-speaker differences, with complementary fixed-effects ladders showing sizeable speaker increments after absorbing topic- and time-specific structure. Individual-level displays (context-adjusted speaker coefficients and BLUPs) reveal persistent non-zero deviations for several

members; pairwise comparisons uncover clusters and clear contrasts in forms-of-expertise mixes, consistent with durable "styles" that travel across topics.

These results have two implications. Substantively, they provide process-level, quantitative evidence on who is addressed and how arguments are substantiated during agenda-defining episodes of global standard-setting. Methodologically, they establish a blueprint for turning rich, messy deliberation text into analyzable evidence, with outputs (labels, metrics, and stylized facts) that subsequent identification-based work can build upon.

In closing, I want to note the following caveat. This study analyzes official IASB meetings for three central projects over 2013–2021; while these are canonical agenda items, coverage is not exhaustive, and automated labeling, despite rigorous validation, may still misclassify edge cases. Conclusions should therefore be treated as directionally informative rather than exhaustive. Even so, the contribution is practical as well as conceptual: the taxonomy and pipeline are reusable. Researchers, investors, and policymakers can leverage the labeled corpus and metrics to navigate deliberations more efficiently, audit whether specific concerns (and their evidentiary bases) were raised, and examine how style and coalition patterns evolve. Beyond the IASB, the framework can be ported, with modest adaptation, to other deliberative venues such as corporate boards, earnings calls, and central-bank or regulatory committees, enabling cumulative and comparative research on how expert bodies deliberate in public and supporting assessments of their procedural quality and effectiveness.

References

Agresti, A. (2007). *An Introduction to Categorical Data Analysis* (2 ed.). Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons.

Allen, A. M. (2018). Agenda setting at the FASB: Evidence from the role of the FASAC. Working Paper 15-042, Harvard Business School Accounting & Management Unit.

Anderson, K. (2009). The rise of international criminal law: Intended and unintended consequences. *European Journal of International Law 20*(2), 331–358.

Arya, A., J. Glover, B. Mittendorf, and G. Narayanamoorthy (2005). Unintended consequences of regulating disclosures: The case of regulation fair disclosure. *Journal of Accounting and Public Policy 24*(3), 243–252.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting linear mixed-effects models using `lme4`. *Journal of Statistical Software 67*, 1–48.

Bertrand, M. and A. Schoar (2003). Managing with style: The effect of managers on firm policies. *The Quarterly Journal of Economics 118*(4), 1169–1208.

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

Bradbury, M. E. and J. Harrison (2014). An analysis of the FASB's dissenting opinions and the conceptual framework. Available at SSRN: https://ssrn.com/abstract=2464862.

Brown, L. D., T. T. Cai, and A. DasGupta (2001). Interval estimation for a binomial proportion. *Statistical Science 16*(2), 101–133.

Brüggemann, U., J.-M. Hitz, and T. Sellhorn (2013). Intended and unintended consequences of mandatory IFRS adoption: A review of extant evidence and suggestions for future research. *European Accounting Review 22*(1), 1–37.

Camfferman, K. and S. A. Zeff (2007). *Financial Reporting and Global Capital Markets: A History of the International Accounting Standards Committee, 1973–2000*. Oxford: Oxford University Press.

Camfferman, K. and S. A. Zeff (2018). The challenge of setting standards for a worldwide constituency: Research implications from the IASB's early history. *European Accounting Review 27*(2), 289–312.

Corbeil, R. R. and S. R. Searle (1976). Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics 18*(1), 31–38.

Cui, J., L. van Lent, and M. Zhu (2025). Quantifying standard-setting deliberations. Working Paper.

Daske, H., L. Hail, C. Leuz, and R. Verdi (2008). Mandatory ifrs reporting around the world: Early evidence on the economic consequences. *Journal of accounting research 46*(5), 1085–1142.

DeFond, M., X. Hu, M. Hung, and S. Li (2011). The impact of mandatory IFRS adoption on foreign mutual fund ownership: The role of comparability. *Journal of Accounting and Economics 51*(3), 240–258.

Demski, J. S. (1973). The general impossibility of normative accounting standards. *The Accounting Review 48*(4), 718–723.

European Commission (2023). Better regulation toolbox. Accessed via European Commission website.

Flores, E., B. R. Monsen, E. Shafron, and C. G. Yust (2025). "No comment": Language frictions and the IASB's due process. *Contemporary Accounting Research 42*(1), 446–489.

Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Gilardi, F., M. Alizadeh, and M. Kubli (2023). Chatgpt outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences 120*(30), e2305016120.

Gipper, B., B. J. Lombardi, and D. J. Skinner (2013). The politics of accounting standard-setting: A review of empirical research. *Australian Journal of Management 38*(3), 523–551.

Grömping, U. (2006). Relative importance for linear regression in R: The package `relaimpo`. *Journal of Statistical Software 17*(1), 1–27.

Hansen, S., M. McMahon, and A. Prat (2018). Transparency and deliberation within the FOMC: A computational linguistics approach. *The Quarterly Journal of Economics 133*(2), 801–870.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association 72*(358), 320–338.

Hassan, N., F. Arslan, C. Li, and M. Tremayne (2017). Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1803–1812. ACM.

Head, B. W. (2008). Three lenses of evidence-based policy. *Australian Journal of Public Administration 67*(1), 1–11.

Head, B. W. (2013). Evidence-based policymaking: Speaking truth to power? *Australian Journal of Public Administration 72*(4), 397–403.

Head, B. W. (2016). Toward more "evidence-informed" policy making? *Public Administration Review 76*(3), 472–484.

IFRS Foundation (2020). Due Process Handbook. IFRS Foundation, August 2020.

IFRS Foundation (2025). Ifrs foundation annual report 2024. Published March 31, 2025.

Jones, B. F. and B. A. Olken (2005). Do leaders matter? national leadership and growth since world war ii. *The Quarterly Journal of Economics 120*(3), 835–864.

Jorissen, A., N. Lybaert, R. Orens, and L. Van der Tas (2013). A geographic analysis of constituents' formal participation in the process of international accounting standard setting: Do we have a level playing field? *Journal of Accounting and Public Policy 32*(4), 237–270.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–291.

Kothari, S. P., K. Ramanna, and D. J. Skinner (2010). Implications for GAAP from an analysis of positive research in accounting. *Journal of Accounting and Economics 50*(2–3), 246–286.

Laakso, M. and R. Taagepera (1979). "effective" number of parties: A measure with application to west europe. *Comparative Political Studies 12*(1), 3–27.

Lauderdale, B. E. and A. Herzog (2016). Measuring political positions from legislative speech. *Political Analysis 24*(3), 374–394.

Lee, D.-H., J. Pujara, M. Sewak, R. W. White, and S. K. Jauhar (2023). Making large language models better data creators. *arXiv preprint arXiv:2310.20111*.

Levy, G. (2007). Decision making in committees: Transparency, reputation, and voting rules. *American Economic Review 97*(1), 150–168.

McLachlan, G. J. and K. E. Basford (1988). *Mixture Models: Inference and Applications to Clustering*, Volume 84 of *Statistics: Textbooks and Monographs*. New York: Marcel Dekker.

Mongrut, S. and D. Winkelried (2019). Unintended effects of IFRS adoption on earnings management: The case of latin america. *Emerging Markets Review 38*, 377–388.

Montesinos López, O. A., A. Montesinos López, and J. Crossa (2022). *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer Nature.

MRC Cognition and Brain Sciences Unit (2021). Rules of thumb on magnitudes of effect sizes. Accessed via MRC CBU Statistics Wiki.

OECD (2021). Organisation for economic co-operation and development (OECD) regulatory policy outlook. Accessed via OECD website.

Osnabrügge, M., S. B. Hobolt, and T. Rodon (2021). Playing to the gallery: Emotive rhetoric in parliaments. *American Political Science Review 115*(3), 885–899.

Owen, P. D. and C. A. Owen (2022). Simulation evidence on herfindahl–hirschman measures of competitive balance in professional sports leagues. *Journal of the Operational Research Society 73*(2), 285–300.

Palmieri, R. and S. Mazzali-Lurati (2016). Multiple audiences as text stakeholders: A conceptual framework for analyzing complex rhetorical situations. *Argumentation 30*(4), 467–499.

Patterson, H. D. and R. Thompson (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika 58*(3), 545–554.

Pryzant, R., D. Iter, J. Li, Y. Lee, C. Zhu, and M. Zeng (2023). Automatic prompt optimization with "gradient descent" and beam search. *arXiv preprint arXiv:2305.03495*.

Ramanna, K. (2015). *Political Standards: Corporate Interest, Ideology, and Leadership in the Shaping of Accounting Rules for the Market Economy*. University of Chicago Press.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science 6*(1), 15–32.

Searle, S. R., G. Casella, and C. E. McCulloch (2009). *Variance Components*. John Wiley & Sons.

Swank, O. H. and B. Visser (2023). Committees as active audiences: Reputation concerns and information acquisition. *Journal of Public Economics 221*, 104875.

Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

Visser, B. and O. H. Swank (2007). On committees of experts. *The Quarterly Journal of Economics 122*(1), 337–372.

Visser, J., J. Lawrence, C. Reed, J. Wagemans, and D. Walton (2020). Annotating argument schemes. In *Argumentation through Languages and Cultures*, pp. 101–139. Cham: Springer Nature Switzerland.

Vogel, S. K. (2022). The politics of accounting standards: A comment on Ramanna's "Unreliable accounts: How regulators fabricate conceptual narratives to diffuse criticism". *Accounting, Economics, and Law: A Convivium 12*(2), 247–252.

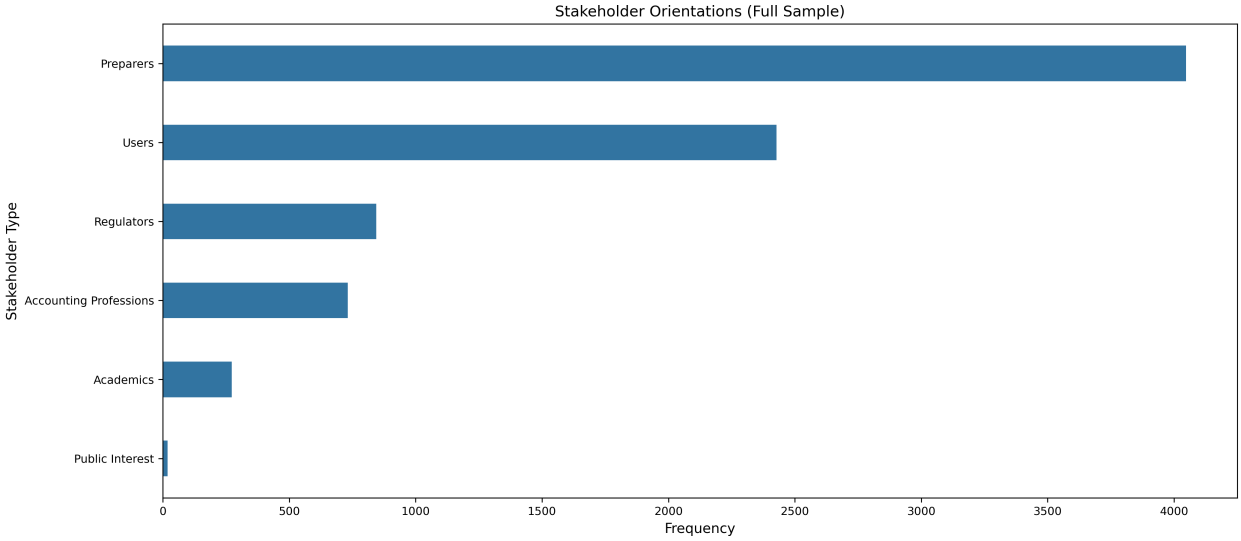
Wagemans, J. H. (2011). The assessment of argumentation from expert opinion. *Argumentation 25*(3), 329–339.

Wan, M., T. Safavi, S. K. Jauhar, Y. Kim, S. Counts, J. Neville, S. Suri, C. Shah, R. W. White, L. Yang, R. Andersen, G. Buscher, D. Joshi, and N. Rangan (2024, August). Tnt-llm: Text mining at scale with large language models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5836–5847. ACM.

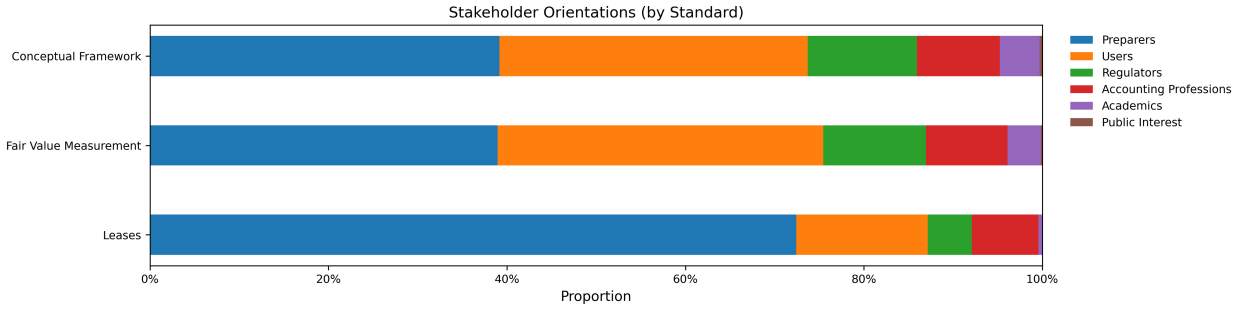
Zeff, S. A. (2005). The evolution of US GAAP: The political forces behind professional standards. *The CPA Journal 75*(2), 18–28.

Zhang, H., Z. Zhu, Z. Zhang, and C. Li (2025). LLMTaxo: Leveraging large language models for constructing taxonomy of factual claims from social media. *arXiv preprint arXiv:2504.12325*.

Figure 1. Meeting Composition — Stakeholder Orientations



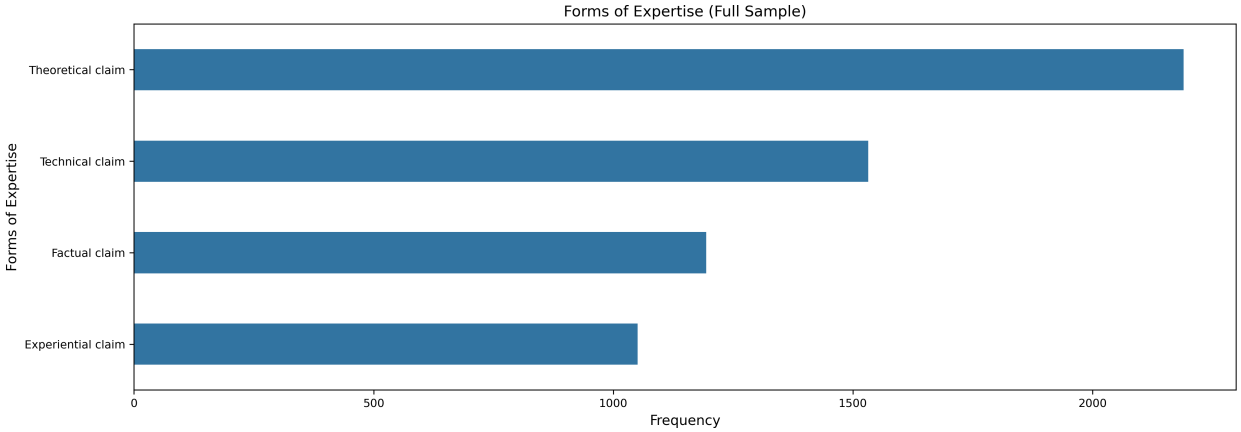
Panel A: Full Sample



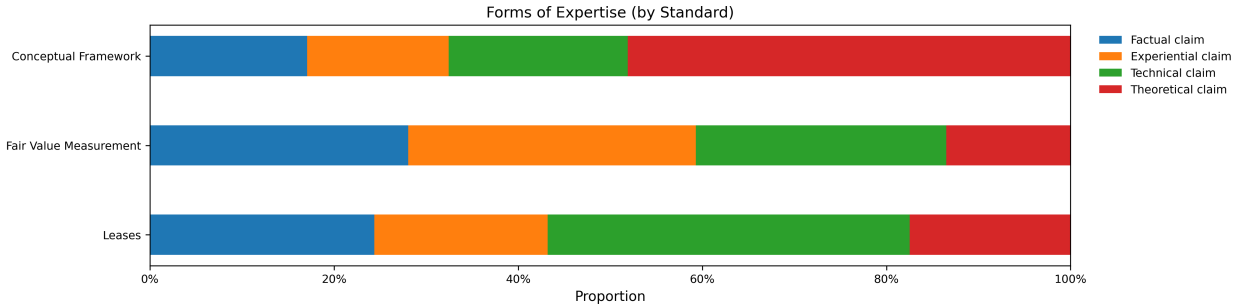
Panel B: By Standard

Notes: This figure presents the composition of stakeholder orientations in IASB meeting deliberations. **Panel A** reports full-sample counts of substantive speech segments coded to each stakeholder group (Preparers, Users, Regulators, Accounting Professions, Academics, Public Interest). **Panel B** shows, by standard (Conceptual Framework, Fair Value Measurement, Leases), the within-standard composition of stakeholder orientations (100% stacked bars). In both panels, categories are assigned at the segment level after omitting non-substantive segments; the "Others" category is excluded from the plots.

Figure 2. Meeting Composition — Forms of Expertise



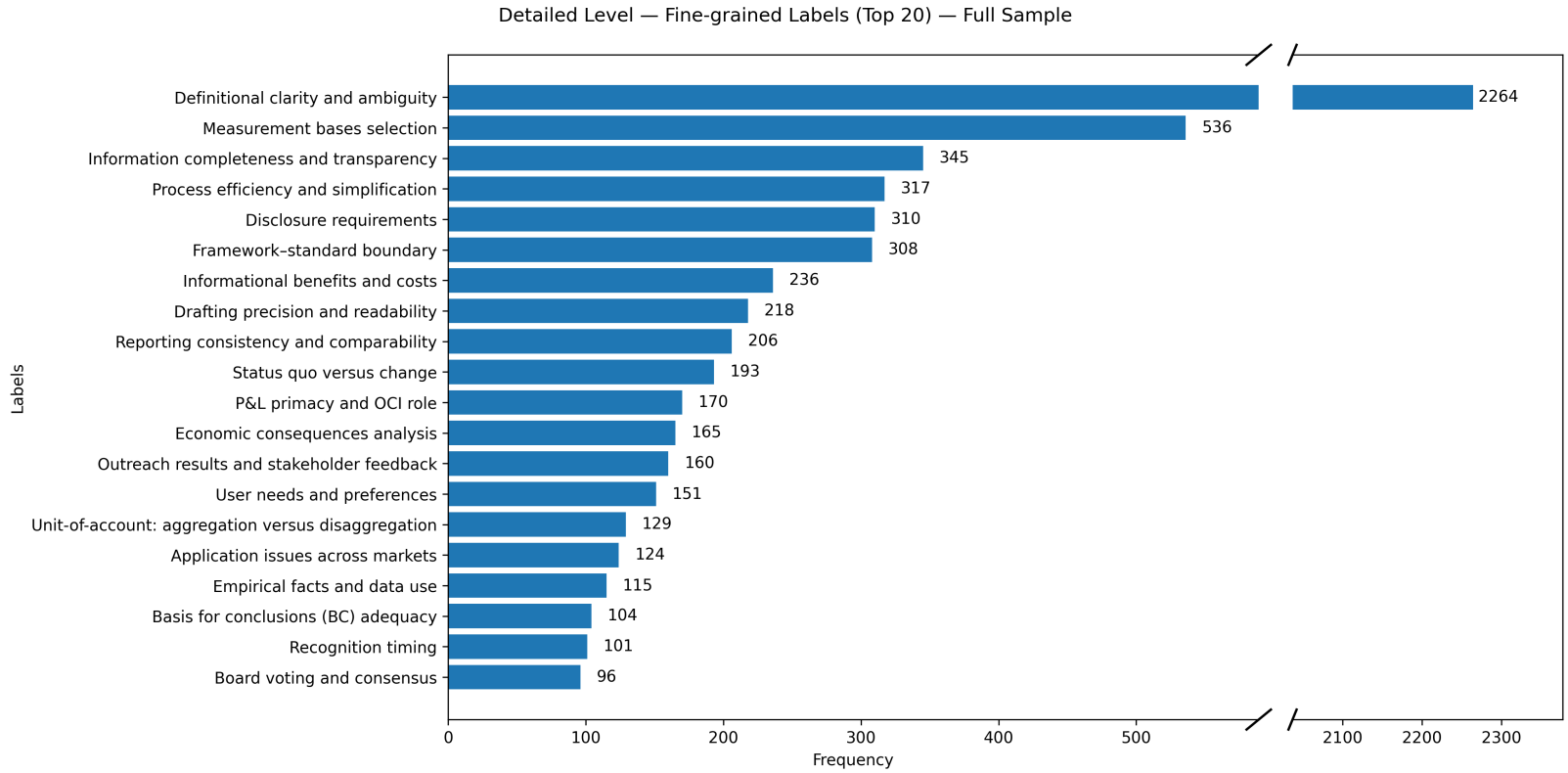
Panel A: Full Sample



Panel B: By Standard

Notes: This figure presents the composition of forms of expertise in IASB meeting deliberations. **Panel A** reports full-sample counts of substantive speech segments coded to each form of expertise (Factual Claim, Experiential Claim, Technical Claim, Theoretical Claim). **Panel B** shows, by standard (Conceptual Framework, Fair Value Measurement, Leases), the within-standard composition of forms of expertise (100% stacked bars). In both panels, categories are assigned at the segment level after omitting non-substantive segments; the "Other Claim" category is excluded from the plots.

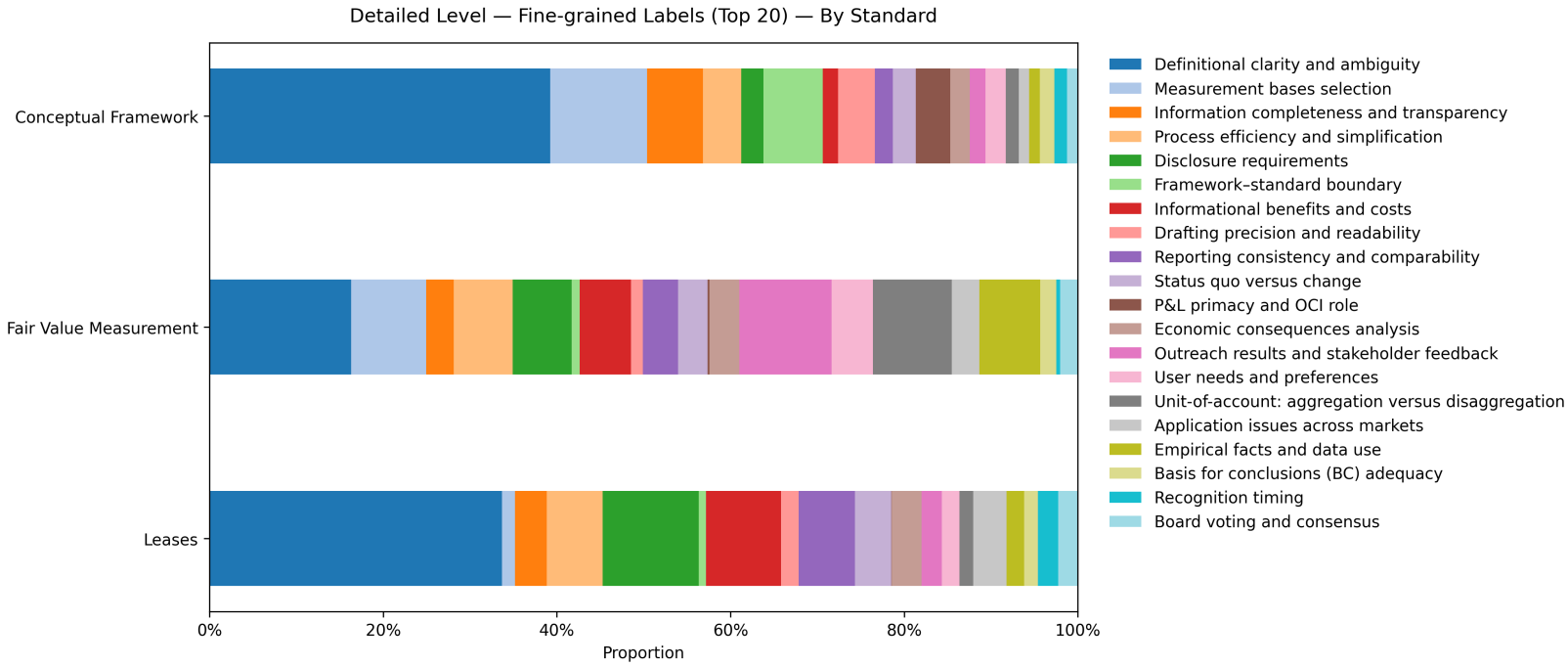
Figure 3. Meeting Composition — Fine-grained Topics (Top 20)



Panel A: Full Sample

Notes: This figure presents the distribution of fine-grained topics in IASB deliberations based on the study's detailed-level taxonomy; the global Top-20 topics account for approximately 71% of substantive segments. **Panel A** reports full-sample frequencies (horizontal bars), ranked by volume. **Panel B** displays, by standard (Conceptual Framework, Fair Value Measurement, Leases), the within-standard composition of the same Top-20 topics (100% stacked bars). In both panels, categories are assigned at the segment level after excluding non-substantive segments.

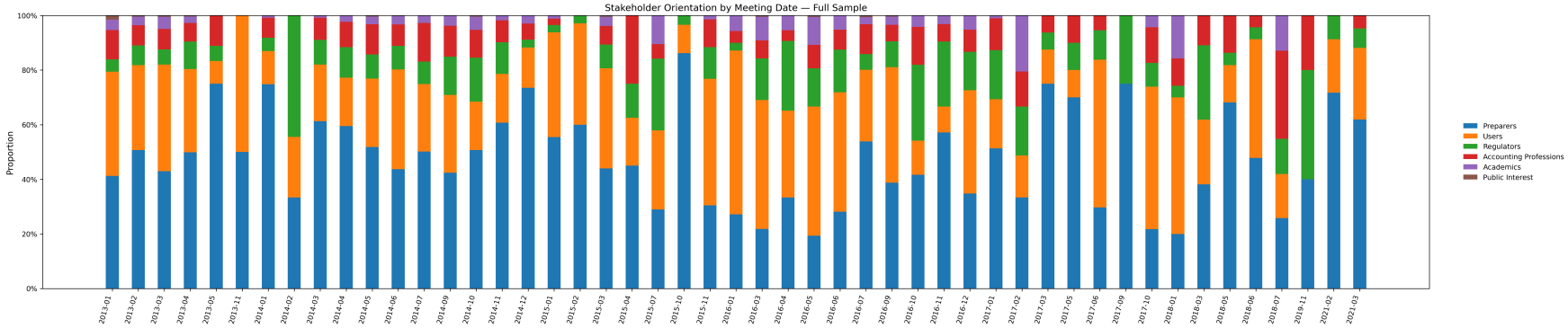
Figure 3. Meeting Composition — Fine-grained Topics (Top 20) (C'd)



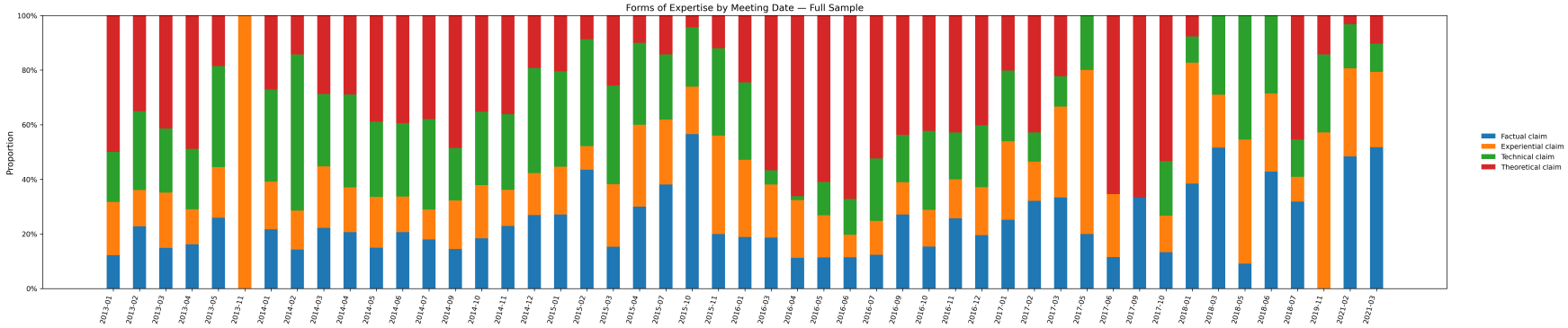
Panel B: By Standard

Notes: This figure presents the distribution of fine-grained topics in IASB deliberations based on the study's detailed-level taxonomy; the global Top-20 topics account for approximately 71% of substantive segments. **Panel A** reports full-sample frequencies (horizontal bars), ranked by volume. **Panel B** displays, by standard (Conceptual Framework, Fair Value Measurement, Leases), the within-standard composition of the same Top-20 topics (100% stacked bars). In both panels, categories are assigned at the segment level after excluding non-substantive segments.

Figure 4. Meeting Composition Over Time



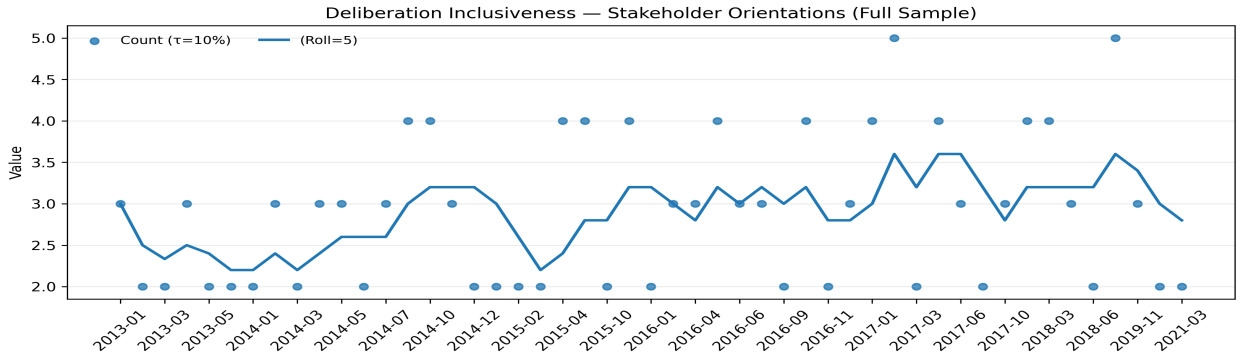
Panel A: Stakeholder Orientations



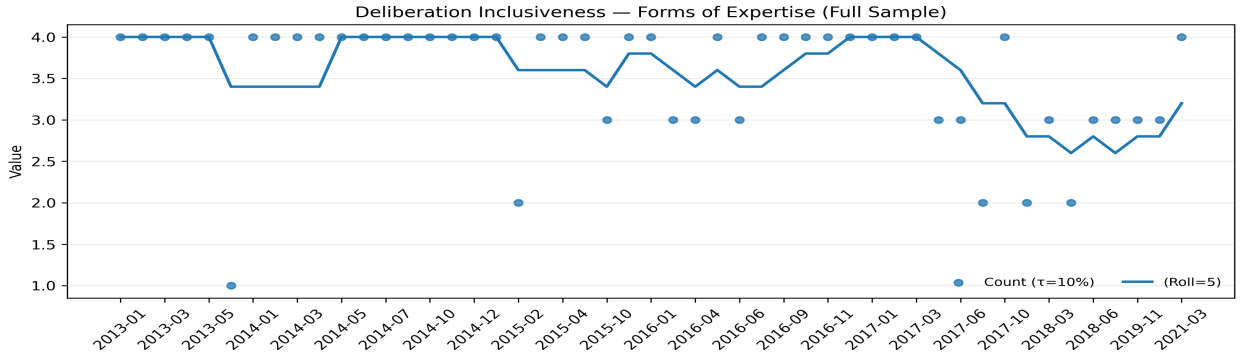
Panel B: Forms of Expertise

Notes: This figure presents the composition of IASB deliberations over time in the full sample. **Panel A** displays stakeholder orientations by Meeting-Date as 100% stacked bars. **Panel B** shows, using the same layout, the forms of expertise. In both panels, categories are assigned at the segment level after excluding non-substantive segments; the categories "Others" (stakeholder orientations) and "Other Claim" (forms of expertise) are excluded from the plots.

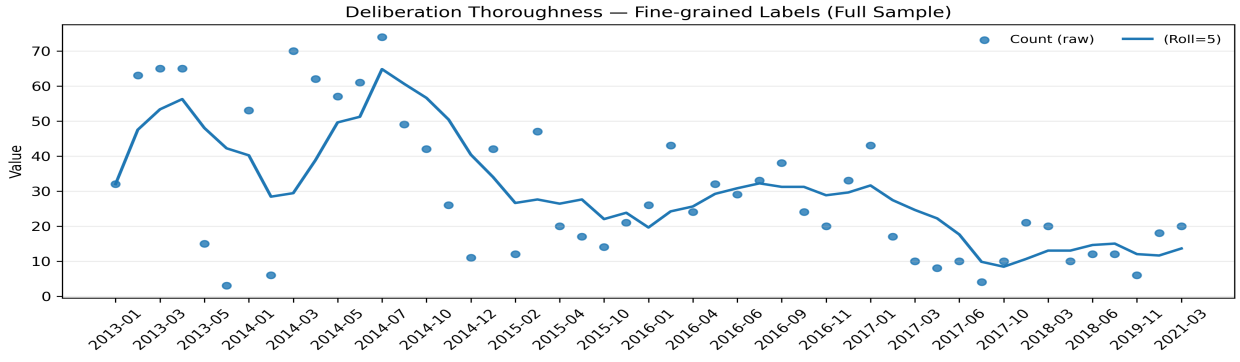
Figure 5. Deliberation Range and Depth



Panel A: Deliberation Range (Inclusiveness) – Stakeholder Orientations



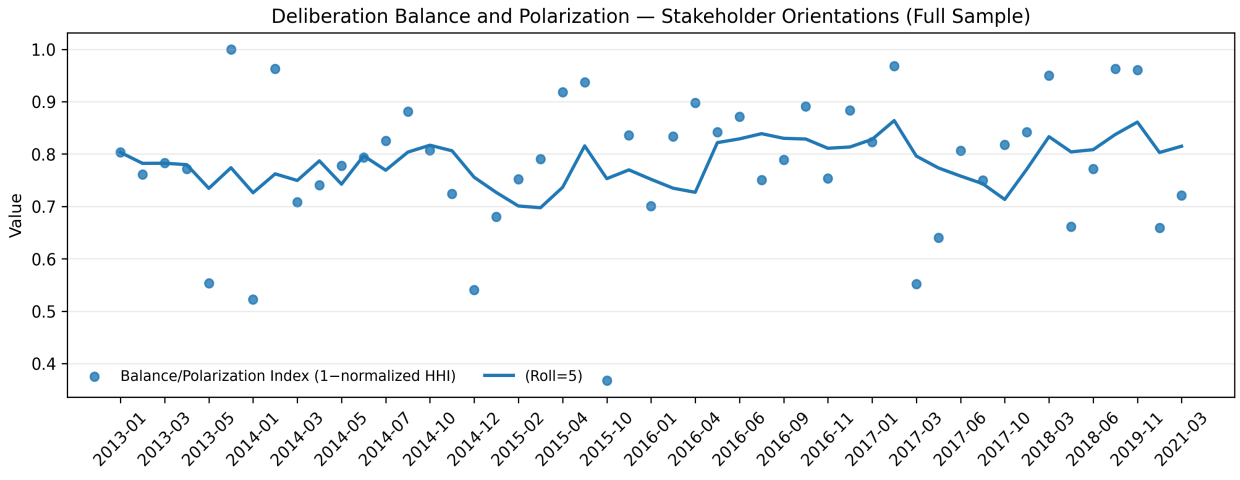
Panel B: Deliberation Range (Inclusiveness) – Forms of Expertise



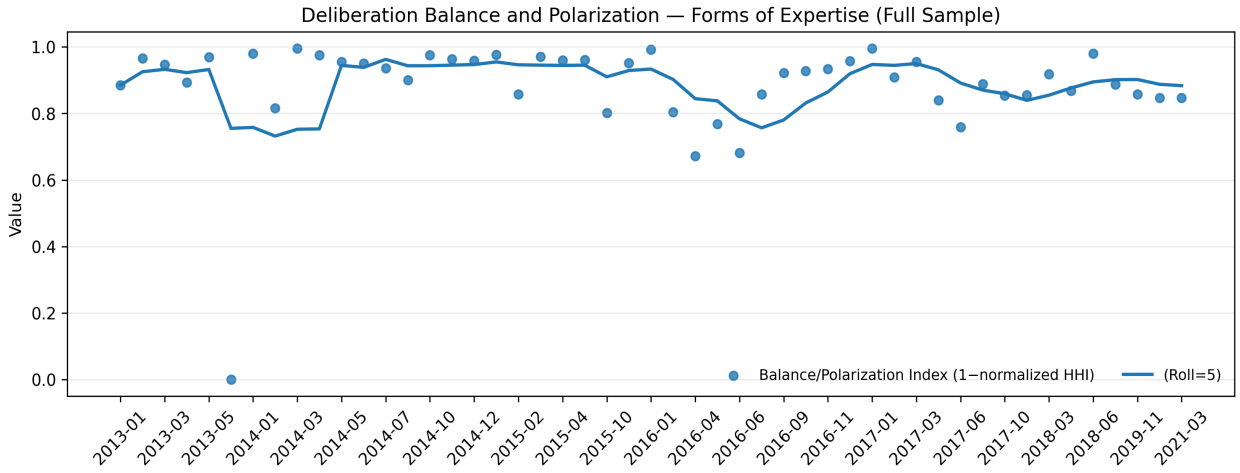
Panel C: Deliberation Depth (Thoroughness) – Fine-grained Labels

Notes: This figure presents standard-setting deliberation metrics that capture the *range* and *depth* of IASB board discussions. **Panels A** and **B** show the number of distinct stakeholder orientations and forms of expertise represented in each meeting date, respectively, which I interpret as *deliberation inclusiveness*. In these panels, inclusiveness is measured as the number of categories that exceed a prevalence threshold of $\tau = 10\%$ of all segments in a meeting date; this rule ensures that only substantively represented categories are counted, rather than those appearing sporadically. **Panel C** shows the count of fine-grained detailed labels per meeting date, interpreted as *deliberation thoroughness* (greater counts indicate that a wider set of technical issues and sub-topics were covered). Both raw series (dots) and smoothed rolling averages (lines, window = 5 meeting dates) are displayed. Together, the measures provide complementary perspectives: inclusiveness reflects the *breadth of viewpoints* brought into deliberation (stakeholder types, expertise forms), while thoroughness reflects the *depth of exploration* within technical sub-topics. Analysis excludes the categories "Others" (Stakeholder Orientations) and "Other Claim" (Forms of Expertise).

Figure 6. Deliberation Balance and Polarization



Panel A: Stakeholder Orientations



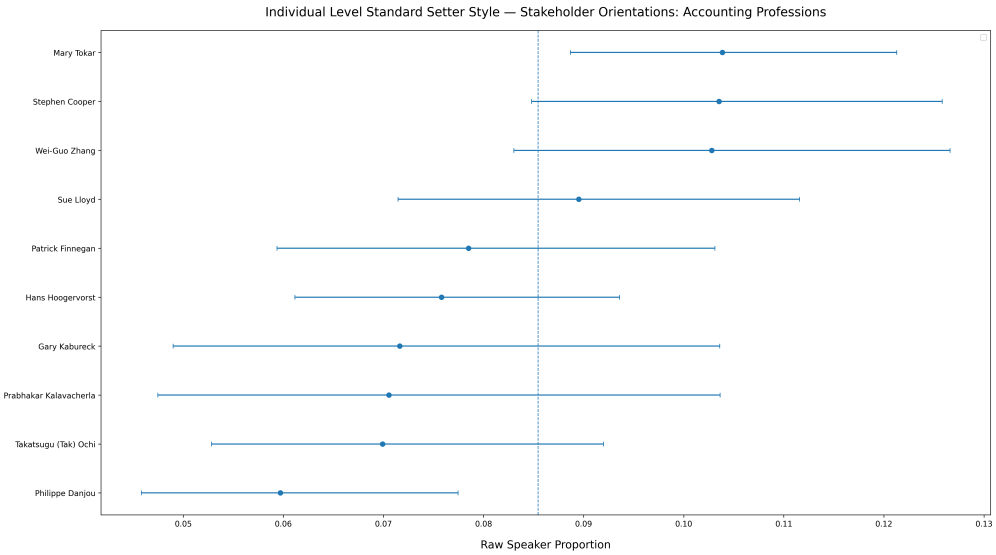
Panel B: Forms of Expertise

Notes: The figure presents indices of deliberation balance (and polarization) for IASB meetings. I interpret these indices as measures of deliberation balance (whether multiple categories are represented proportionally) and polarization (whether discussion is dominated by a single category). **Panel A** shows the index across Stakeholder Orientations and **Panel B** across Forms of Expertise. The underlying metric is the Herfindahl–Hirschman Index (HHI). The HHI is computed as the sum of squared shares of each category within a meeting date. To enable comparability across different classification systems (e.g., six stakeholder groups vs. four expertise types, or meetings where not all categories appear), I normalize HHI to the $[0, 1]$ interval:

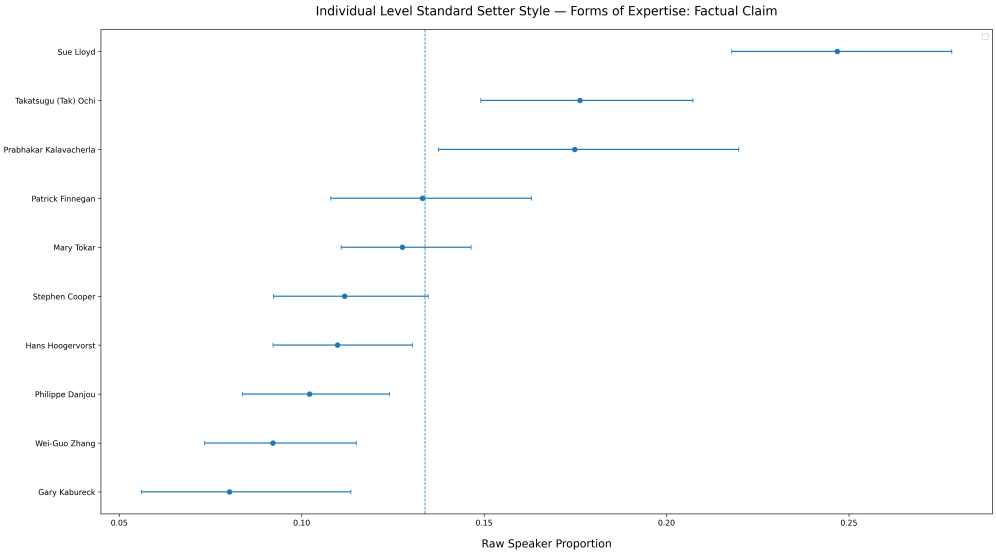
$$\text{HHI}^* = \frac{HHI - \frac{1}{K}}{1 - \frac{1}{K}},$$

where K is the number of categories represented. The balance and polarization index is defined as one minus the normalized HHI. Higher values indicate more evenly distributed deliberations across categories, while lower values indicate stronger polarization around one or a few categories. Time series display both raw meeting date-level values (dots) and rolling averages (lines, window = 5 meeting dates) to highlight trends over time. Analysis excludes the categories "Others" (Stakeholder Orientations) and "Other Claim" (Forms of Expertise).

Figure 7. Individual Level Standard Setter Style — Raw Proportion



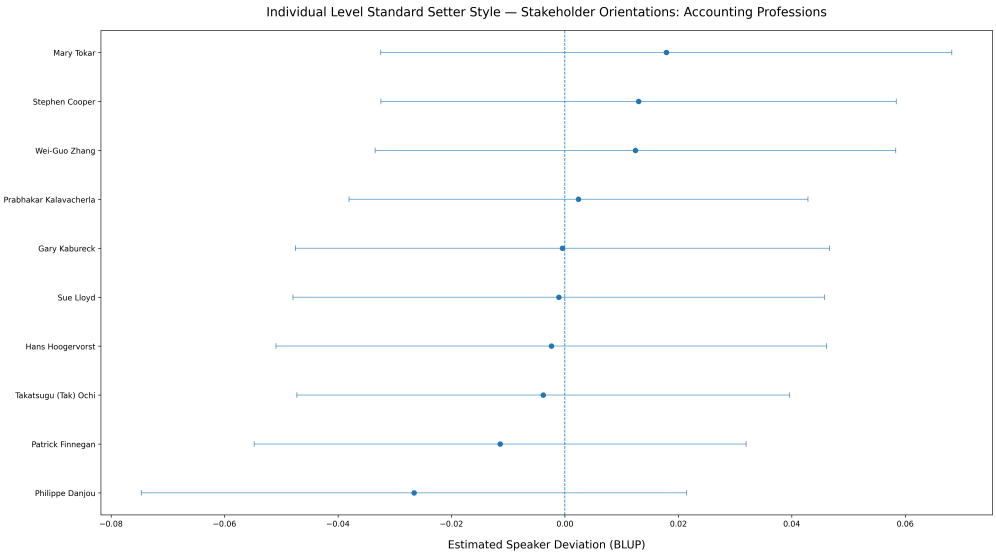
Panel A: Accounting Professions



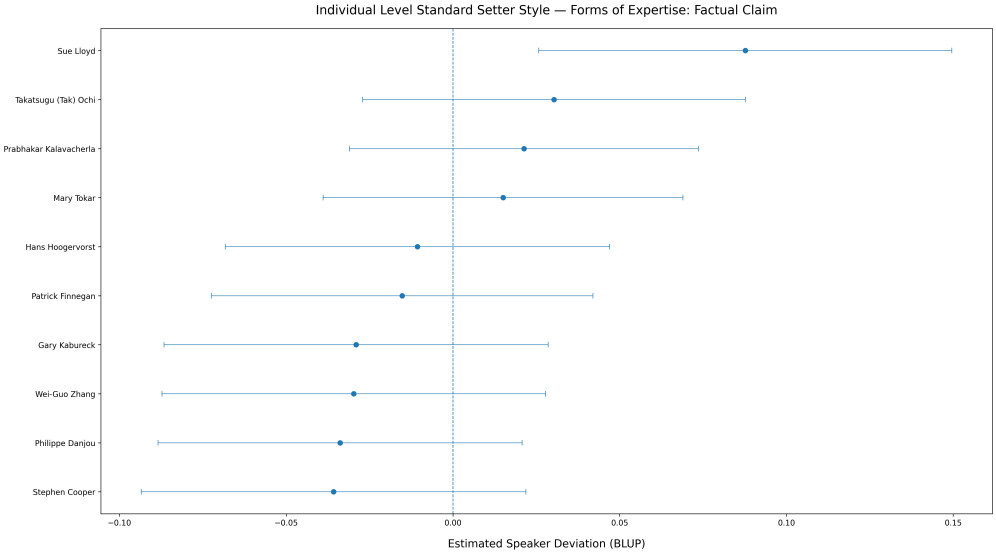
Panel B: Factual Claim

Notes:This figure presents *raw proportion* caterpillars of how much individual IASB board members emphasize a given category. **Panel A** shows results for *Accounting Professions* and **Panel B** for *Factual Claim*. For each speaker, the dot represents the share of that speaker's total speaking segments that fall into the category. The dashed vertical line marks the overall average share across all speakers (grand mean). Error bars are 95% Wilson confidence intervals around each speaker's share, which account for sampling variability. Intervals are wider for speakers with fewer total segments. Points to the right of the dashed line indicate that the speaker places greater emphasis on the category than the overall average, while points to the left indicate less emphasis. This display is unadjusted (no controls for topic or meeting) and unshrunken (no statistical pooling), so it directly reflects observed differences in the raw data. To improve precision, only the top 10 speakers by total segment count are analyzed; these speakers account for 86.2% of all segments.

Figure 8. Individual Level Standard Setter Style — BLUPs



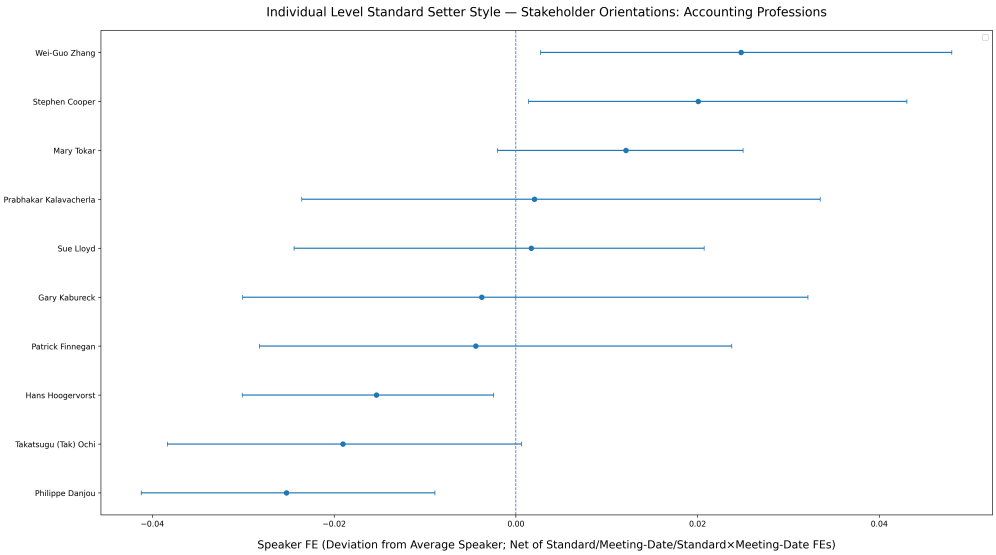
Panel A: Accounting Professions



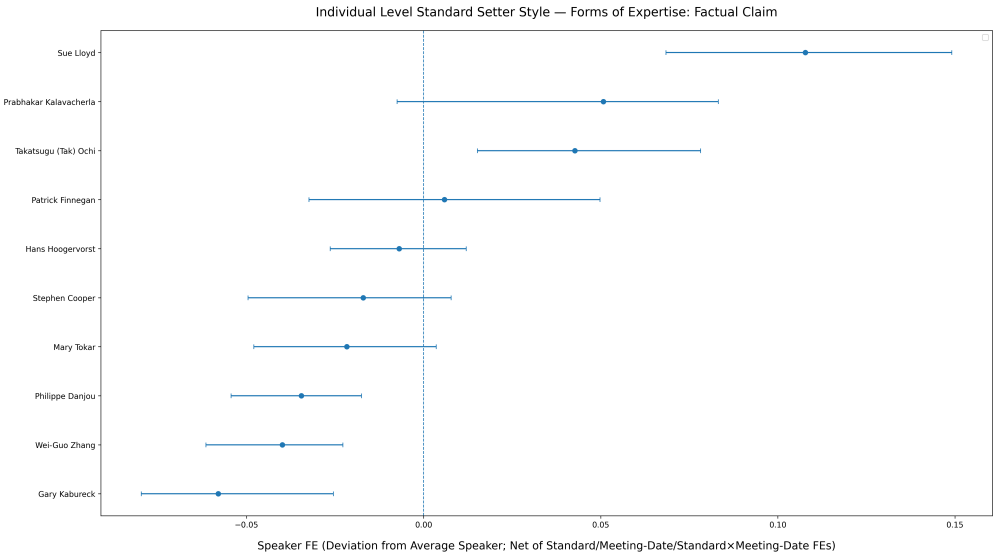
Panel B: Factual Claim

Notes:The figure presents speaker random-effect estimates (BLUPs, Best Linear Unbiased Predictions) from a REML mixed model of category shares at the Speaker×Standard×Meeting-Date unit. The model includes a fixed intercept capturing the overall mean of the category share (the model baseline) and random intercepts for Speaker, Standard, and Meeting-Date. Each point is a speaker's BLUP—an estimated *speaker-specific deviation* from the model baseline. Accordingly, the zero line denotes *no speaker-specific deviation*; values above (below) zero indicate that the speaker tends to emphasize the category more (less) than the average speaker. BLUPs implement *partial pooling*: speakers with fewer observations are pulled more toward the model baseline, while well-observed speakers remain closer to their raw deviations. Error bars are ±1.96× bootstrap standard errors, computed from 1,000 parametric bootstrap replications that simulate data from the fitted model and re-fit the same specification. **Panel A** reports the results for *Accounting Professions* and **Panel B** for *Factual Claim*. To improve precision, only the top 10 speakers by total segment count are analyzed; these speakers account for 86.2% of all segments.

Figure 9. Individual Level Standard Setter Style — Context-adjusted Deviation



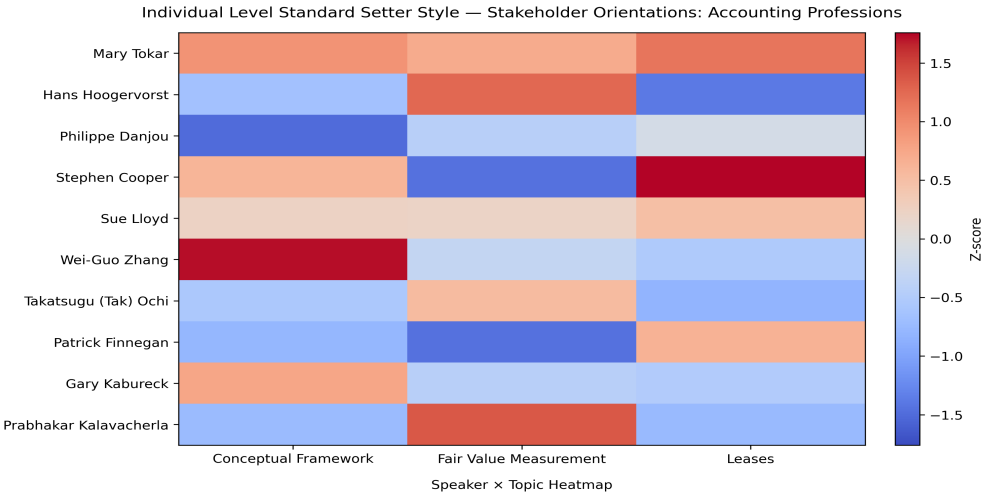
Panel A: Accounting Professions



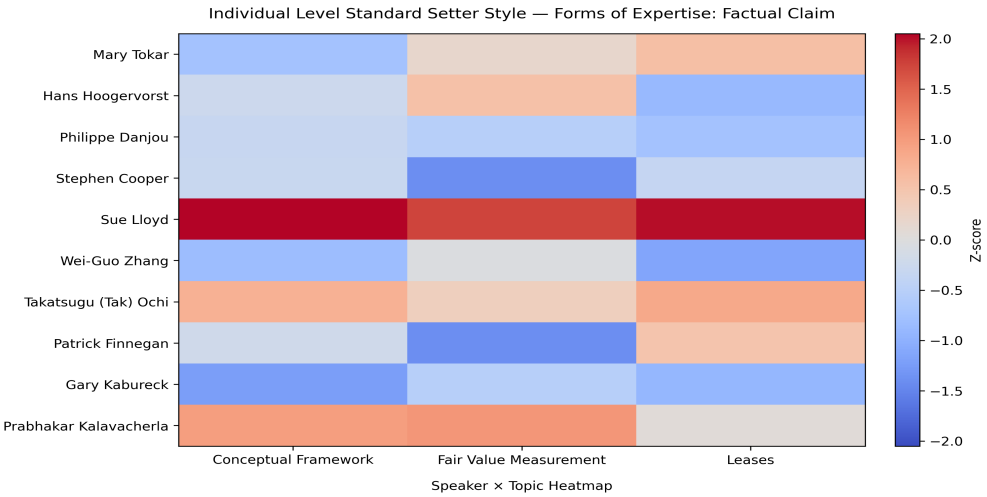
Panel B: Factual Claim

Notes: This figure presents *context-adjusted* speaker fixed effects for emphasis on a given category after removing meeting-context effects. **Panel A** shows *Accounting Professions* and **Panel B** shows *Factual Claim*. For each category, I estimate a WLS model of the category share at the Speaker×Standard×Meeting-Date unit on fixed effects for Standard, Meeting-Date, their interaction (Standard×Meeting-Date), and Speaker; weights equal the unit's segment count. Each dot is the *speaker fixed-effect coefficient*, estimated and re-centered so that the precision-weighted mean across speakers is zero. The dashed vertical line at 0 therefore denotes the average speaker, conditional on meeting context. Points to the right (left) of zero indicate greater (less) emphasis than the average speaker, holding topic and timing fixed. Error bars are 95% cluster-bootstrap confidence intervals, obtained by resampling Standard×Meeting-Date groups, refitting the WLS with the same fixed effects, and re-centering the speaker coefficients in each replicate. To improve precision, only the top 10 speakers by total segment count are analyzed; these speakers account for 86.2% of all segments.

Figure 10. Individual Level Standard Setter Style — Z-Score



Panel A: Accounting Professions



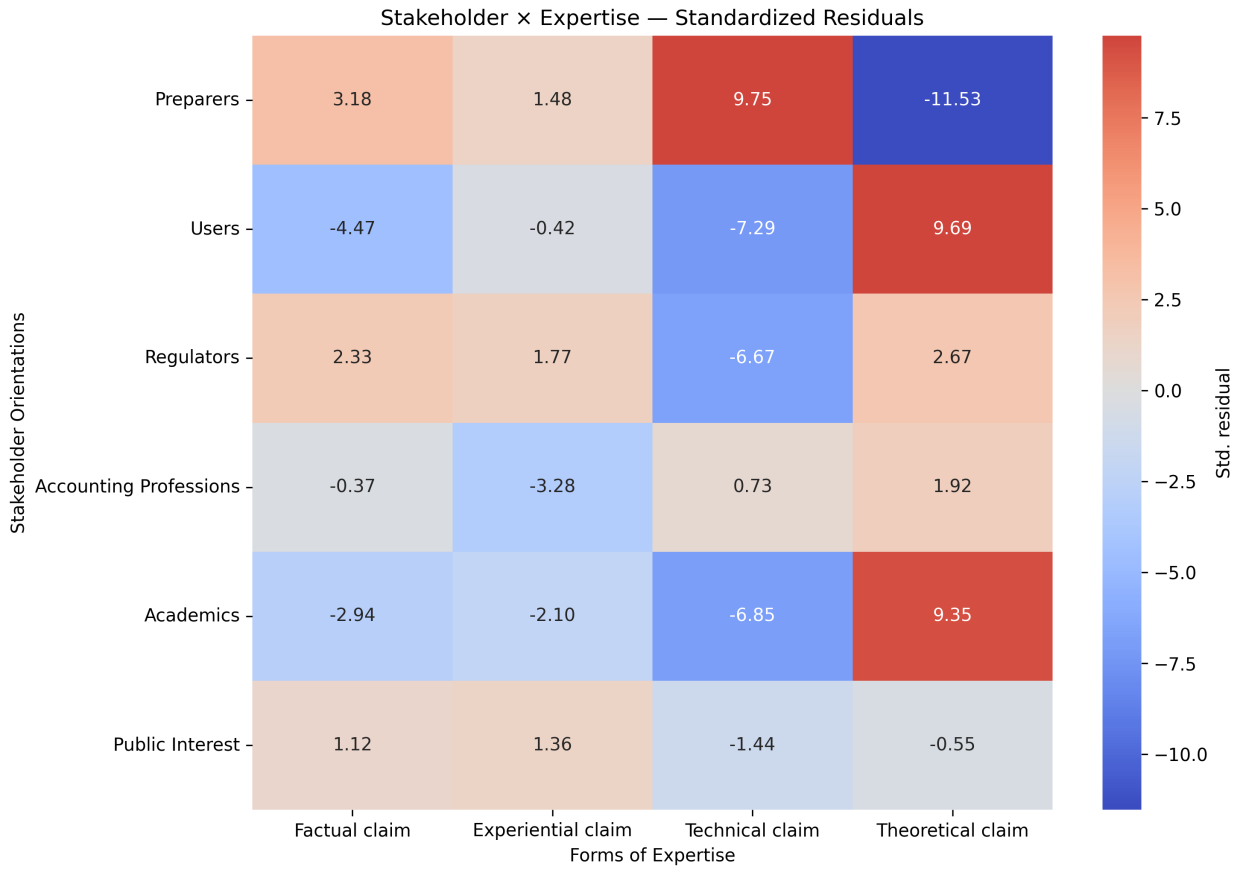
Panel B: Factual Claim

Notes: This figure presents a Speaker×Standard heatmap for the *Accounting Professions* category (**Panel A**) and the *Factual Clam* category (**Panel B**). Rows represent individual IASB board members, while columns correspond to Standards. The color scale is based on column-wise z-scores of category shares, computed as:

$$z_{s,t} = \frac{\text{Share}_{s,t} - \overline{\text{Share}}_{\cdot,t}}{\text{sd}(\text{Share}_{\cdot,t})},$$

where $\text{Share}_{s,t}$ is the weighted average fraction of speaker s's segments that are classified as *Accounting Professions/Factual Clam* within Standard t, and $\overline{\text{Share}}_{\cdot,t}$ and $\text{sd}(\text{Share}_{\cdot,t})$ are the mean and standard deviation across all speakers for Standard t. Weighted averages are taken over meeting-dates, with weights equal to the number of segments in each Speaker×Standard×Meeting-Date unit. Color intensity indicates whether a speaker emphasizes *Accounting Professions/Factual Clam* more or less than peers in a given topic. Positive (red) values indicate above-average emphasis; negative (blue) values indicate below-average emphasis. The symmetric color scale is capped at the 98th percentile of absolute z-scores to prevent extreme outliers from dominating the plot. To improve precision, only the top 10 speakers by total segment count are analyzed; these speakers account for 86.2% of all segments.

Figure 11. Association Between Stakeholder Orientations vs. Forms of Expertise



Notes: This figure presents a heatmap of standardized Pearson residuals from the cross-tabulation (contingency table) of stakeholder orientations (rows) by forms of expertise (columns). The standardized residual for cell (i, j) is

$$R_{ij} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}}},$$

where O_{ij} is the observed count and E_{ij} is the expected count under the null of independence,

$$E_{ij} = \frac{(\text{Row total})_i (\text{Column total})_j}{N},$$

with N the grand total of observations. Positive values (shown in red) indicate cells that are *over-represented* relative to independence; negative values (blue) indicate *under-representation*. The magnitude $|R_{ij}|$ can be read as the approximate number of standard deviations by which the observed cell deviates from its expected value. The sum of squared residuals $\sum_{ij} R_{ij}^2$ equals the Pearson chi-square statistic reported in Table 10. Analysis excludes the categories "Others" (Stakeholder Orientations) and "Other Claim" (Forms of Expertise).

Table 1. Summary Statistics

Panel A: Dataset Composition	
IASB Standards	Conceptual Framework, Fair Value Measurement, Leases
Total Meetings	66
Unique Speakers	26
Segments Count (Substantive)	8758
Time Range	2013 - 2021

Panel B: Taxonomy Composition

Stakeholder Orientations	**Conceptual Framework**	**Fair Value Measurement**	**Leases**	**Full Sample**
Preparers	2092	251	1705	4048
Users	1846	235	347	2428
Regulators	654	74	116	844
Accounting Professions	496	59	176	731
Academics	239	24	9	272
Public Interest	16	1	1	18
Others	316	21	80	417
Total	*5659*	*665*	*2434*	*8758*

Forms of Expertise	**Conceptual Framework**	**Fair Value Measurement**	**Leases**	**Full Sample**
Factual Claim	651	133	410	1194
Experiential claim	586	148	317	1051
Technical Claim	741	129	662	1532
Theoretical Claim	1832	64	294	2190
Other Claim	1849	191	751	2791
Total	*5659*	*665*	*2434*	*8758*

Notes: This table presents an overview of the dataset and the composition of the taxonomy used in the study. **Panel A** summarizes the corpus: standards covered (Conceptual Framework, Fair Value Measurement, Leases), number of meetings (66), unique speakers (26), the count of substantive segments (8,758), and the time range (2013–2021). **Panel B** reports taxonomy composition by standard and in the full sample at three levels: (i) Stakeholder orientations, (ii) Forms of expertise, and (iii) Fine-grained topics (Top-20). Entries are segment counts after filtering out non-substantive speech. The Top-20 topics together account for 6,248 of 8,758 substantive segments (≈71%).

Table 1. Summary Statistics (C'd)

Panel B: Taxonomy Composition (C'd)				
Fined-grained Topics (Top 20)	**Conceptual Framework**	**Fair Value Measurement**	**Leases**	**Full Sample**
Definitional clarity and ambiguity	1670	72	522	2264
Measurement bases selection	475	38	23	536
Information completeness and transparency	274	14	57	345
Process efficiency and simplification	187	30	100	317
Disclosure requirements	109	30	171	310
Framework–standard boundary	291	4	13	308
Informational benefits and costs	76	26	134	236
Drafting precision and readability	180	6	32	218
Reporting consistency and comparability	88	18	100	206
Status quo versus change	113	15	65	193
P&L primacy and OCI role	168	1	1	170
Economic consequences analysis	97	15	53	165
Outreach results and stakeholder feedback	77	47	36	160
User needs and preferences	98	21	32	151
Unit-of-account: aggregation vs disaggregation	65	40	24	129
Application issues across markets	50	14	60	124
Empirical facts and data use	53	31	31	115
Basis for conclusions (BC) adequacy	71	8	25	104
Recognition timing	63	2	36	101
Board voting and consensus	52	9	35	96
Total	*4257*	*441*	*1550*	*6248*

Notes: This table presents an overview of the dataset and the composition of the taxonomy used in the study. **Panel A** summarizes the corpus: standards covered (Conceptual Framework, Fair Value Measurement, Leases), number of meetings (66), unique speakers (26), the count of substantive segments (8,758), and the time range (2013–2021). **Panel B** reports taxonomy composition by standard and in the full sample at three levels: (i) Stakeholder orientations, (ii) Forms of expertise, and (iii) Fine-grained topics (Top-20). Entries are segment counts after filtering out non-substantive speech. The Top-20 topics together account for 6,248 of 8,758 substantive segments (≈71%).

Table 2. Descriptive Statistics — Stakeholder Orientations (Counts)

Panel A: Meeting-Level						
	Mean	**Median**	**St.Dev.**	**Min**	**Max**	**N**
Preparers	61.33	27.50	77.48	1.00	396.00	66
Users	36.79	18.00	45.88	0.00	217.00	66
Regulators	12.79	5.50	14.32	0.00	57.00	66
Accounting Professions	11.08	5.00	14.04	0.00	68.00	66
Academics	4.12	1.00	5.88	0.00	24.00	66
Public Interest	0.27	0.00	0.62	0.00	3.00	66
Others	6.32	3.00	8.20	0.00	36.00	66
Panel B: Speaker-Level						
	Mean	**Median**	**St.Dev.**	**Min**	**Max**	**N**
Preparers	155.69	73.00	178.67	1.00	650.00	26
Users	93.38	50.00	112.76	0.00	361.00	26
Regulators	32.46	15.00	43.93	0.00	175.00	26
Accounting Professions	28.12	13.00	36.44	0.00	140.00	26
Academics	10.46	4.50	13.19	0.00	49.00	26
Public Interest	0.69	0.00	1.46	0.00	5.00	26
Others	16.04	8.50	21.62	0.00	79.00	26
Panel C: Speaker-Meeting-Level						
	Mean	**Median**	**St.Dev.**	**Min**	**Max**	**N**
Preparers	6.04	3.00	9.27	0.00	74.00	670
Users	3.62	2.00	5.67	0.00	49.00	670
Regulators	1.26	0.00	2.08	0.00	13.00	670
Accounting Professions	1.09	0.00	2.02	0.00	15.00	670
Academics	0.41	0.00	0.96	0.00	9.00	670
Public Interest	0.03	0.00	0.19	0.00	3.00	670
Others	0.62	0.00	1.20	0.00	9.00	670

Notes: This table presents summary statistics for the number of segments assigned to each stakeholder orientation, computed across three unit levels. **Panel A** reports Meeting-level counts. **Panel B** reports Speaker-level counts aggregated over all segments per speaker. **Panel C** reports Speaker-Meeting-level counts. Statistics are based on substantive segments only.

Table 3. Descriptive Statistics — Stakeholder Orientations (Proportions)

Panel A: Meeting-Level						
	Mean	**Median**	**St.Dev.**	**Min**	**Max**	**N**
Preparers	0.48	0.44	0.21	0.17	1.00	66
Users	0.25	0.25	0.15	0.00	0.57	66
Regulators	0.11	0.09	0.09	0.00	0.44	66
Accounting Professions	0.08	0.07	0.06	0.00	0.29	66
Academics	0.03	0.01	0.04	0.00	0.27	66
Public Interest	0.00	0.00	0.01	0.00	0.03	66
Others	0.05	0.04	0.05	0.00	0.33	66
Panel B: Speaker-Level						
	Mean	**Median**	**St.Dev.**	**Min**	**Max**	**N**
Preparers	0.52	0.50	0.14	0.22	0.80	26
Users	0.23	0.24	0.12	0.00	0.50	26
Regulators	0.12	0.09	0.11	0.00	0.50	26
Accounting Professions	0.06	0.07	0.04	0.00	0.10	26
Academics	0.03	0.02	0.04	0.00	0.20	26
Public Interest	0.00	0.00	0.00	0.00	0.02	26
Others	0.04	0.04	0.03	0.00	0.10	26
Panel C: Speaker-Meeting-Level						
	Mean	**Median**	**St.Dev.**	**Min**	**Max**	**N**
Preparers	0.47	0.45	0.31	0.00	1.00	670
Users	0.27	0.23	0.26	0.00	1.00	670
Regulators	0.11	0.00	0.19	0.00	1.00	670
Accounting Professions	0.08	0.00	0.14	0.00	1.00	670
Academics	0.03	0.00	0.09	0.00	1.00	670
Public Interest	0.00	0.00	0.01	0.00	0.13	670
Others	0.05	0.00	0.11	0.00	1.00	670

Notes: This table presents summary statistics for the within-unit shares of each stakeholder orientation, computed across three unit levels. **Panel A** reports proportions at the Meeting-level, **Panel B** at the Speaker-level, and **Panel C** at the Speaker-Meeting-level. Shares sum to 1 within each unit. Statistics are based on substantive segments only.

Table 4. Descriptive Statistics — Forms of Expertise (Counts)

Panel A: Meeting-Level						
	Mean	**Median**	**St.Dev.**	**Min**	**Max**	**N**
Factual Claim	18.09	10.00	19.70	0.00	86.00	66
Experiential Claim	15.92	10.50	16.15	0.00	76.00	66
Technical Claim	23.21	10.00	30.31	0.00	123.00	66
Theoretical Claim	33.18	13.00	46.25	0.00	190.00	66
Other Claim	42.29	22.00	46.23	1.00	194.00	66
Panel B: Speaker-Level						
	Mean	**Median**	**St.Dev.**	**Min**	**Max**	**N**
Factual Claim	45.92	21.00	55.48	0.00	193.00	26
Experiential Claim	40.42	18.00	45.91	0.00	149.00	26
Technical Claim	58.92	27.50	70.87	0.00	245.00	26
Theoretical Claim	84.23	40.50	108.84	0.00	359.00	26
Other Claim	107.35	58.00	125.23	0.00	444.00	26
Panel C: Speaker-Meeting-Level						
	Mean	**Median**	**St.Dev.**	**Min**	**Max**	**N**
Factual Claim	1.78	1.00	2.98	0.00	24.00	670
Experiential Claim	1.57	1.00	2.35	0.00	18.00	670
Technical Claim	2.29	1.00	4.03	0.00	36.00	670
Theoretical Claim	3.27	1.00	5.71	0.00	45.00	670
Other Claim	4.17	2.00	5.54	0.00	35.00	670

Notes: This table presents summary statistics for the number of segments assigned to each form of expertise, computed across three unit levels. **Panel A** reports Meeting-level counts. **Panel B** reports Speaker-level counts aggregated over all segments per speaker. **Panel C** reports Speaker-Meeting-level counts. Statistics are based on substantive segments only.

Table 5. Descriptive Statistics — Forms of Expertise (Proportions)

Panel A: Meeting-Level						
	Mean	**Median**	**St.Dev.**	**Min**	**Max**	**N**
Factual Claim	0.15	0.14	0.09	0.00	0.45	66
Experiential Claim	0.14	0.12	0.08	0.00	0.33	66
Technical Claim	0.17	0.16	0.10	0.00	0.44	66
Theoretical Claim	0.19	0.16	0.14	0.00	0.50	66
Other Claim	0.34	0.33	0.12	0.09	1.00	66
Panel B: Speaker-Level						
	Mean	**Median**	**St.Dev.**	**Min**	**Max**	**N**
Factual Claim	0.18	0.13	0.13	0.00	0.50	26
Experiential Claim	0.14	0.13	0.08	0.00	0.43	26
Technical Claim	0.18	0.18	0.09	0.00	0.50	26
Theoretical Claim	0.16	0.18	0.12	0.00	0.32	26
Other Claim	0.35	0.33	0.13	0.00	0.63	26
Panel C: Speaker-Meeting-Level						
	Mean	**Median**	**St.Dev.**	**Min**	**Max**	**N**
Factual Claim	0.13	0.07	0.19	0.00	1.00	670
Experiential Claim	0.14	0.08	0.19	0.00	1.00	670
Technical Claim	0.17	0.12	0.22	0.00	1.00	670
Theoretical Claim	0.20	0.16	0.23	0.00	1.00	670
Other Claim	0.36	0.33	0.28	0.00	1.00	670

Notes: This table presents summary statistics for the within-unit shares of each form of expertise, computed across three unit levels. **Panel A** reports proportions at the Meeting-level, **Panel B** at the Speaker-level, and **Panel C** at the Speaker-Meeting-level. Shares sum to 1 within each unit. Statistics are based on substantive segments only.

Table 6. Standard Setter Style — Variance Components Mixed Model (REML)

Category	Variance Share (%)				N Units	VC Used
	Speaker	Standard	Meeting-Date	Residual		
Panel A: Stakeholder Orientations						
Preparers	1.44	45.18	0.04	53.34	332	Standard+Meeting-Date
Users	0.01	32.98	7.83	59.19	332	Standard+Meeting-Date
Regulators	4.99	13.14	0.00	81.87	332	Standard+Meeting-Date
Accounting Professions	13.36	1.15	11.31	74.19	332	Standard+Meeting-Date
Academics	0.01	12.48	15.30	72.22	332	Standard+Meeting-Date
Public Interest	5.68	2.74	14.74	76.84	332	Standard+Meeting-Date
Panel B: Forms of Expertise						
Factual Claim	13.74	13.85	3.53	68.87	332	Standard+Meeting-Date
Experiential Claim	1.18	26.64	0.26	71.92	332	Standard+Meeting-Date
Technical Claim	0.10	24.01	4.00	71.88	332	Standard+Meeting-Date
Theoretical Claim	0.04	42.91	1.63	55.42	332	Standard+Meeting-Date

Notes: The table reports estimates from restricted maximum likelihood (REML) mixed-effects models of category shares measured at the Speaker×Standard×Meeting-Date unit. For each category, a separate model is estimated; the outcome variable is the category's share within the unit (the category's segment count divided by the unit's total segments). Each model includes (i) a fixed intercept capturing the overall mean of the category share and (ii) random intercepts for the grouping factors—Speaker, Standard, and Meeting-Date. "VC Used" lists the variance components retained in the converged specification after the adaptive fitting procedure. Each component's variance is reported as a share of the total variance; shares sum to 100% within each row. Units with fewer than five segments were excluded. To improve precision, only the top 10 speakers by total segment count are analyzed; these speakers account for 86.2% of all segments.

Table 7. Standard Setter Style — Variance Decomposition with Fixed Effects

Category	ΔR^2 (%)						Residual
	M0	+M1	+M2	+M3	+M4	+M5	
Panel A: Stakeholder Orientations							
Preparers	0.00	50.81	17.11	2.25	5.54	1.22	23.07
Users	0.00	25.32	27.32	4.75	10.61	1.78	30.22
Regulators	0.00	11.23	26.41	3.75	8.56	4.81	45.25
Accounting Professions	0.00	0.82	25.50	10.41	5.14	2.50	55.64
Academics	0.00	11.97	23.20	3.41	2.65	3.90	54.87
Public Interest	0.00	0.76	12.95	12.38	3.96	6.30	63.64
Panel B: Forms of Expertise							
Factual Claim	0.00	8.79	21.04	3.75	17.07	4.08	45.29
Experiential Claim	0.00	12.65	11.37	6.84	14.21	4.94	49.99
Technical Claim	0.00	27.49	21.47	5.47	3.89	3.14	38.55
Theoretical Claim	0.00	42.77	8.14	1.72	10.06	2.58	34.72
Panel C: Included Fixed Effects							
M0: Intercept	YES	YES	YES	YES	YES	YES	YES
M1: Standard FE		YES	YES	YES	YES	YES	YES
M2: Meeting-Date FE			YES	YES	YES	YES	YES
M3: Standard×Meeting-Date FE				YES	YES	YES	YES
M4: Speaker FE					YES	YES	YES
M5: Speaker×Standard FE						YES	YES

Notes: This table reports in-sample fixed-effects R^2 from weighted least squares (WLS) projections of category shares measured at the Speaker×Standard×Meeting-Date unit. Category share is defined as the category's segment count divided by the unit's total segments. The R^2 values are precision-weighted using segment counts as weights, so that units with more segments receive greater influence, reflecting the higher precision of their shares. Columns 2–7 report incremental R^2 gains as successive sets of fixed effects are added in the order listed in Panel C. "Residual" is the unexplained proportion, equal to 100 minus the cumulative explained share. Units with fewer than five segments were excluded. To improve precision, only the top 10 speakers by total segment count are analyzed; these speakers account for 86.2% of all segments.

Table 8. Pairwise Correlations Among Standard Setters — Stakeholder Orientations

Speaker	Gary Kabureck	Hans Hoogervorst	Mary Tokar	Patrick Finnegan	Philippe Danjou	Prabhakar Kalavacherla	Stephen Cooper	Sue Lloyd	Takatsugu (Tak) Ochi	Wei-Guo Zhang
Gary Kabureck	**1.00**	0.72	0.98	0.89	0.97	1.00	0.91	0.89	0.95	0.96
Hans Hoogervorst	0.72	**1.00**	0.83	0.96	0.87	0.76	0.93	0.96	0.89	0.84
Mary Tokar	0.98	0.83	**1.00**	0.95	0.98	0.99	0.96	0.95	0.98	1.00
Patrick Finnegan	0.89	0.96	0.95	**1.00**	0.97	0.91	0.99	1.00	0.98	0.95
Philippe Danjou	0.97	0.87	0.98	0.97	**1.00**	0.98	0.98	0.97	1.00	0.97
Prabhakar Kalavacherla	1.00	0.76	0.99	0.91	0.98	**1.00**	0.93	0.91	0.97	0.98
Stephen Cooper	0.91	0.93	0.96	0.99	0.98	0.93	**1.00**	1.00	0.99	0.96
Sue Lloyd	0.89	0.96	0.95	1.00	0.97	0.91	1.00	**1.00**	0.98	0.95
Takatsugu (Tak) Ochi	0.95	0.89	0.98	0.98	1.00	0.97	0.99	0.98	**1.00**	0.97
Wei-Guo Zhang	0.96	0.84	1.00	0.95	0.97	0.98	0.96	0.95	0.97	**1.00**

Notes: This table presents pairwise correlations between speakers' distributional profiles across stakeholder orientations. The purpose is to characterize how similarly different speakers allocate their discussion across categories; values near +1 indicate very similar mixes, values near 0 indicate little relation, and negative values indicate opposite emphasis patterns. For speakers a and b, with share vectors $\mathbf{x}^{(a)} = (x_1^{(a)}, \ldots, x_K^{(a)})$ and $\mathbf{x}^{(b)}$, the Pearson correlation is

$$\rho(a,b) = \frac{\sum_{k=1}^{K} \left(x_k^{(a)} - \bar{x}^{(a)}\right)\left(x_k^{(b)} - \bar{x}^{(b)}\right)}{\sqrt{\sum_{k=1}^{K} \left(x_k^{(a)} - \bar{x}^{(a)}\right)^2} \sqrt{\sum_{k=1}^{K} \left(x_k^{(b)} - \bar{x}^{(b)}\right)^2}},$$

where each $\mathbf{x}^{(s)}$ contains *row-normalized shares* (row sums equal 1). I compute correlations on the *shares* (not raw counts), so the measure reflects pattern similarity (mix), not speaking volume. If a speaker's profile has zero variance (all mass in one category), the correlation with that speaker is undefined; the implementation replaces such *NaN* with 0 (interpreted as "no measurable similarity"). Analysis excludes the categories "Others" (Stakeholder Orientations) and "Other Claim" (Forms of Expertise); to improve precision, I restrict attention to the top 10 speakers by total segment count.

Table 9. Pairwise Correlations Among Standard Setters — Forms of Expertise

Speaker	Gary Kabureck	Hans Hoogervorst	Mary Tokar	Patrick Finnegan	Philippe Danjou	Prabhakar Kalavacherla	Stephen Cooper	Sue Lloyd	Takatsugu (Tak) Ochi	Wei-Guo Zhang
Gary Kabureck	**1.00**	-0.33	-0.27	-0.17	0.02	0.01	-0.20	-0.87	-0.56	0.16
Hans Hoogervorst	-0.33	**1.00**	0.71	0.71	0.61	-0.44	0.73	0.18	0.65	0.87
Mary Tokar	-0.27	0.71	**1.00**	0.99	0.95	0.32	1.00	0.50	0.94	0.67
Patrick Finnegan	-0.17	0.71	0.99	**1.00**	0.98	0.30	1.00	0.40	0.90	0.73
Philippe Danjou	0.02	0.61	0.95	0.98	**1.00**	0.37	0.97	0.27	0.81	0.73
Prabhakar Kalavacherla	0.01	-0.44	0.32	0.30	0.37	**1.00**	0.29	0.47	0.36	-0.36
Stephen Cooper	-0.20	0.73	1.00	1.00	0.97	0.29	**1.00**	0.43	0.91	0.72
Sue Lloyd	-0.87	0.18	0.50	0.40	0.27	0.47	0.43	**1.00**	0.75	-0.20
Takatsugu (Tak) Ochi	-0.56	0.65	0.94	0.90	0.81	0.36	0.91	0.75	**1.00**	0.46
Wei-Guo Zhang	0.16	0.87	0.67	0.73	0.73	-0.36	0.72	-0.20	0.46	**1.00**

Notes: This table presents pairwise correlations between speakers' distributional profiles across forms of expertise. The purpose is to characterize how similarly different speakers allocate their discussion across categories; values near +1 indicate very similar mixes, values near 0 indicate little relation, and negative values indicate opposite emphasis patterns. For speakers a and b, with share vectors $\mathbf{x}^{(a)} = (x_1^{(a)}, \ldots, x_K^{(a)})$ and $\mathbf{x}^{(b)}$, the Pearson correlation is

$$\rho(a,b) = \frac{\sum_{k=1}^{K} \left(x_k^{(a)} - \bar{x}^{(a)}\right)\left(x_k^{(b)} - \bar{x}^{(b)}\right)}{\sqrt{\sum_{k=1}^{K} \left(x_k^{(a)} - \bar{x}^{(a)}\right)^2} \sqrt{\sum_{k=1}^{K} \left(x_k^{(b)} - \bar{x}^{(b)}\right)^2}},$$

where each $\mathbf{x}^{(s)}$ contains *row-normalized shares* (row sums equal 1). I compute correlations on the *shares* (not raw counts), so the measure reflects pattern similarity (mix), not speaking volume. If a speaker's profile has zero variance (all mass in one category), the correlation with that speaker is undefined; the implementation replaces such *NaN* with 0 (interpreted as "no measurable similarity"). Analysis excludes the categories "Others" (Stakeholder Orientations) and "Other Claim" (Forms of Expertise); to improve precision, I restrict attention to the top 10 speakers by total segment count.

Table 10. Association Between Stakeholder Orientations vs. Forms of Expertise

Panel A: Cross-tabulated Raw Frequency					
	Factual claim	Experiential claim	Technical claim	Theoretical claim	*Total*
Preparers	673	555	1,048	720	2,996
Users	239	272	267	826	1,604
Regulators	129	109	59	232	529
Accounting Professions	106	64	152	231	553
Academics	24	25	5	164	218
Public Interest	3	3	0	2	8
Total	1,174	1,028	1,531	2,175	5,908
Panel B: Pearson Chi-square Test of Independence					
$\chi^2(15)$					635.642
p-value					9.4e-126
N					5,908
Cramér's V					0.189 (bias-corrected=0.187)

Notes: This table reports the association between stakeholder orientations and forms of expertise to examine whether certain stakeholder appeals tend to be associated with certain forms of expertise. **Panel A** shows cross-tabulated raw counts of speech segments (e.g., there are 673 segments in which the speaker appeals to *Preparers* and uses a *Factual claim*). **Panel B** reports the Pearson chi-square test of independence and Cramér's V as an effect size. The chi-square statistic is

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

where O_{ij} is the observed count in row i (stakeholder category) and column j (expertise category), and E_{ij} is the expected count under independence, computed as

$$E_{ij} = \frac{(\text{Row total})_i \, (\text{Column total})_j}{N},$$

where $N = \sum_i \sum_j O_{ij}$ is the grand total of observations. Cramér's V is calculated as

$$V = \sqrt{\frac{\chi^2/N}{\min(r-1,\ c-1)}},$$

where r and c are the numbers of stakeholder and expertise categories, respectively (both uncorrected and small-sample bias-corrected values are presented). For a cell-wise decomposition of the chi-square statistic, see the standardized residual heatmap in Figure 11; note that the sum of squared Pearson residuals satisfies $\sum_{i,j} R_{ij}^2 = \chi^2$. Analysis excludes the categories "Others" (Stakeholder Orientations) and "Other Claim" (Forms of Expertise).

Online Appendix

What Anchors Standard-setting Deliberations?*

Jie Cui†

September 2025

*I gratefully acknowledge funding from the Deutsche Forschungsgemeinschaft Project ID 403041268 – TRR 266.

†**Frankfurt School of Finance & Management**; Adickesallee 32–34, 60322 Frankfurt am Main, Germany; E-mail: `j.cui@fs.de`.

Online Appendix Table 1. Prompt Template

 # Instruction:
- You are an expert annotator of regulatory board meeting transcripts.
- Analyze the following input text data—a speech segment from an International Accounting Standards Board (IASB) meeting—and answer the questions below according to the provided guidelines.

 # Questions:
Q1. Classify the input text into one of the following categories, based on the segment's information content:
1. **Factual claim (verifiable):** The segment contains factual assertions that can be checked or verified as true/false, such as statistics, empirical evidence, references to standards/rules, observable facts.
Examples:
- "The staff report shows that 85% of companies would be affected by the amendment."
- "According to the outreach we have conducted, many small entities struggle with these disclosure requirements."
- "At the IFRS meeting, the representative of New Zealand said that they have a major use of fair value for PPE."
2. **Expert opinion (non-verifiable but relevant):** The segment contains professional judgment, personal experience, interpretation, or opinion, which may be relevant to the standard-setting discussion but cannot be rigorously verified.
Examples:
- "Based on my experience, many preparers struggle with this distinction."
- "Last month or once before when I was talking to people about measurement, I specifically asked this question."
- "Some people would say that the lottery ticket is the asset, and it could be measured."
- "I think the concept of performance is the key and the essence of what we are doing."
3. **Other (non-substantive):** The segment concerns meeting logistics, turn-taking, or is otherwise unrelated to the substantive discussion, such as greetings, thanks, scheduling.
Examples:
- "Good morning, everybody. Today we will begin by discussing the conceptual framework."
- "So, why don't we take a break and come back at 10:30."
- "Thank you. Let's conclude our board meeting."
Guidelines:
- Assign only one category (1, 2, or 3) that best describes the segment's information content.
- If more than one could apply, select the single most relevant category following the priority order: $1 > 2 > 3$.
- Factual content always takes highest priority (assign 1 if the segment contains verifiable facts).
- If no factual claim is present, check for expert opinion (assign 2 if found).
- If neither applies, assign 3.
Q2. Confidence Score: Assign an integer from 0 (lowest) to 100 (highest) reflecting your confidence in the classification.
Q3. Justification: Provide a concise justification (maximum {`justification_length`} words) explaining why you assigned the selected category.
 # Output Format: Provide your answers between the following tags:

```
<category>Your answer to Q1 here (1, 2, or 3).</category>
<confidence_score>Your confidence score (0-100).</confidence_score>
<justification>Your answer to Q3 here.</justification>
```

 # Input Text Data: {`Text`}

Panel A: Filtering for Substantive Content

Online Appendix Table 1. Prompt Template (C'd)

 # Instruction:
- You are an expert annotator of regulatory board meeting transcripts.
- Analyze the following input text data—a speech segment from an International Accounting Standards Board (IASB) meeting—and answer the questions below according to the provided guidelines.

 # Questions:
Q1. Stakeholder Group Classification: Identify the primary stakeholder group whose concerns are addressed in the speech segment. Choose one of the following:
- **Regulators:** Regulatory/standard-setting bodies, legal authorities, government representatives, etc.
- **Accounting Professions:** Accountants, auditors, or their professional bodies (e.g., Big Four firms).
- **Users:** Users of financial statements (analysts, lenders, investors—either individual or institutional).
- **Preparers:** Entities/individuals preparing financial statements (businesses, companies, CEOs, executives).
- **Academics:** Representatives of universities or educational institutions (professors, students, researchers).
- **Public Interest:** NGOs, charities, not-for-profits, consumer advocates.
- **Others:** Only if none of the above apply.

Guidelines:
- Assign only one stakeholder group that is most relevant to the segment's content.
- If more than one could apply, select the single most central group.
- Do not infer the speaker's background—classify based on the concern or perspective evident in the segment.
- Choose category "Others" only if none of the other categories apply.

Examples:
Preparers:
- "The implementation cost of this standard will be significant for many companies."

Users:
- "Investors need comparability across periods to assess management's performance."

Regulators:
- "Enforcement will be challenging for securities regulators in multiple jurisdictions."

Accounting Professions:
- "The Big Four firms have raised concerns about how this would be documented in practice."

Academics:
- "The literature on stewardship suggests a broader definition of performance."

Public Interest:
- "Consumer advocacy groups worry about the lack of transparency in financial reporting."

Others:
- "We need a practical solution that balances all interests." (i.e., the segment is too general to be assigned a single most central group.)

Q2. Confidence Score: Assign an integer from 0 (lowest) to 100 (highest) reflecting your confidence in the classification.
Q3. Justification: Provide a concise justification (maximum {`justification_length`} words) explaining why you assigned the selected stakeholder group.
 # Output Format: Provide your answers between the following tags:

```
<category>Your answer to Q1 here.</category>
<confidence_score>Your confidence score (0-100).</confidence_score>
<justification>Your answer to Q3 here.</justification>
```

 # Input Text Data: {`Text`}

Panel B: Stakeholder Orientations

Online Appendix Table 1. Prompt Template (C'd)

 # Instruction:
- You are an expert annotator of regulatory board meeting transcripts.
- Analyze the following input text data—a speech segment from an International Accounting Standards Board (IASB) meeting—and answer the questions below according to the provided guidelines.

 # Questions:
Q1. Classify the input text into one of the following categories, based on the segment's source of authority or knowledge:
1. **Experiential claim:** The segment reflects personal or collective past practices, events, or experience relevant to the standard-setting discussion.
Examples:
- "Similar practices have caused a lot of problems in the past."
- "We tried to implement this approach in Japan, but it led to confusion."
2. **Technical claim:** The segment reflects technical knowledge or domain-specific know-how related to procedures, rules, or detailed accounting treatments.
Examples:
- "The effective interest method requires calculations not easily understood by most practitioners."
- "Under current fair value measurement techniques, it is difficult to determine the appropriate discount rate."
3. **Theoretical claim:** The segment appeals to accounting theory, conceptual framework, or higher-level principles, reflecting theory-level knowledge rather than practical experience or technical details.
Examples:
- "This proposal does not fit well with the existing conceptual framework."
- "The treatment should be based on the matching principle, which underpins accrual accounting."
4. **Other claim:** The segment expresses professional judgment, interpretation, or opinion relevant to the discussion, but does not fit any of the above categories.
Examples:
- "In my opinion, the proposed approach is too complex."
- "We need to have a rough solution and agree on it."
Guidelines:
- Assign only one category (1, 2, 3, or 4) that best describes the segment's key source of authority or knowledge—what the speaker mainly relies on for their reasoning (experience, expertise, theory, or other).
- If more than one could apply, select the single most relevant.
- Choose category "4" (Other claim) only if none of the other categories apply.
Q2. Confidence Score:
Assign an integer from 0 (lowest) to 100 (highest) reflecting your confidence in the classification.
Q3. Justification:
Provide a concise justification (maximum {`justification_length`} words) explaining why you assigned the selected category.
 # Output Format: Provide your answers between the following tags:

```
<category>Your answer to Q1 here (1, 2, 3, or 4).</category>
<confidence_score>Your confidence score (0-100)</confidence_score>
<justification>Your answer to Q3 here.</justification>
```

 # Input Text Data: {`Text`}

Panel C: Forms of Expertise

Online Appendix Table 1. Prompt Template (C'd)

 # Instruction:
- You are a concise, structured text processor.
- Analyze the following input text data—a speech segment from an International Accounting Standards Board (IASB) meeting—and answer the questions below according to the provided guidelines.

 # Questions:
Q1. Summarize the speech segment in {`summary_length`} words or fewer.

Guidelines:
- **Use case:** Your summary will serve as input for category (taxonomy) generation, labeling the informational content discussed at IASB meetings.
- Extract all key aspects of informational content—what the segment is about, the main factual points, expert opinions, and any supporting rationale.
- Focus on how the speaker presents their views (e.g., by providing empirical evidence, statistics, references to rules, observable facts, or by appealing to: home-country practices, personal experience, investor needs, economic consequences, technical issues or theoretical consistency).
- Make the summary clear, precise, and accurately reflective of the segment's substance.
- Output a single, continuous block of text (no line breaks).

Q2. Explain your approach to generating the summary in {`explanation_length`} words or fewer.

 # Output Format: Provide your answers between the following tags:
`<summary>Your answer to Q1 here.</summary>`
`<explanation>Your answer to Q2 here.</explanation>`

 # Input Text Data: {`Text`}

Panel D: Use-case Summary

Online Appendix Table 1. Prompt Template (C'd)

 # Instruction:
- You are a concise, structured text processor. Analyze the following input text data and answer the questions below according to the provided guidelines.
- The input text data is a markdown table containing summaries of speech segments from International Accounting Standards Board (IASB) meetings, including the following columns:
• **id**: speech segment index.
• **text**: summary of the segment.

 # Questions:
Q1. Cluster the input summaries into meaningful, detailed, and fine-grained categories that best reflect the informational content discussed at IASB meetings.
Guidelines:
- Use case: These categories will form a taxonomy for labeling the informational content that board members contribute to justify or support their views during IASB meetings.
- Output categories should:
• Capture all key aspects of the information content presented in the input summaries (e.g., empirical evidence, statistics, references to rules, observable facts, home-country practices, personal experience, investor needs, economic consequences, technical issues, theoretical consistency).
• Be orthogonal (non-overlapping), specific, clear, and mutually exclusive.
• Match the data closely: do not miss important categories, and do not add categories not directly supported by the data.
• Avoid vague labels ("Other", "General", "Unclear", "Miscellaneous", "Undefined") or overly broad labels.
• Enable consistent classification of new IASB meeting summaries.
- Format your output as a markdown table with the following columns:
• **id**: category index (start from 1, incrementally).
• **name**: concise, clear category name ({`cluster_name_length`} words max), using a noun or verb phrase.
• **description**: brief, specific explanation of the category ({`cluster_description_length`} words max) that distinguishes it from all others.
- The name and description should be consistent with each other.
- The table should be a flat list of mutually exclusive categories, sorted by semantic relatedness.
- Only generate categories supported by the input data; do not hallucinate or include "catch-all" categories.
- Ignore and do not generate categories for low-quality or ambiguous input data points.
- Total number of categories should be no less than {`num_clusters`}.
Q2. Explain, within {`explanation_length`} words, why you grouped the summaries the way you did and how you defined the categories.
 # Output Format: Provide your answers between the following tags:
`<category_table>Your generated markdown category table.</category_table>`
`<explanation>Your explanation of the reasoning process.</explanation>`
 # Input Text Data: {`minibatch_table`}

Panel E: Taxonomy Creation

Online Appendix Table 1. Prompt Template (C'd)

 # Instruction:
- You are a concise, structured text processor. Read the input text data and the reference category table carefully, and answer the questions below according to the provided guidelines.
- The input text data is a markdown table containing summaries of speech segments from International Accounting Standards Board (IASB) meetings, including the following columns:
• **id**: speech segment index.
• **text**: summary of the segment.
- The reference category table is a markdown table containing predefined categories for labeling the informational content discussed at IASB meetings. The table includes the following columns:
• **id**: category index.
• **name**: category name.
• **description**: category description.

 # Questions:
Q1. Review the reference category table and the input text data and provide a rating score of the reference category table (0–100; higher is better).
Guidelines:
- Use case: The category table serves as a taxonomy for labeling the informational content that board members contribute to justify or support their views during IASB meetings.
- Consider both intrinsic and extrinsic quality when rating the category table:
Intrinsic:
• The reference category table should be a flat list of mutually exclusive categories.
• The categories should:
– Capture key aspects of the information content (e.g., empirical evidence, statistics, references to rules).
– Be meaningful, detailed, and fine-grained to best reflect the discussions at IASB meetings.
– Be orthogonal (non-overlapping), specific, clear, and mutually exclusive.
– Avoid vague labels ("Other", "General", "Unclear", "Miscellaneous", "Undefined") or overly broad ones.
Extrinsic:
• Can the reference category table accurately and consistently classify the input data without ambiguity?
• Are there categories missing from the reference category table that appear in the input data?
• Are there unnecessary categories in the reference category table that do not appear in the input data?
Q2. Explain your rating score in Q1 within {`explanation_length`} words.
Q3. How could the reference category table be improved?
- Describe your suggestion within {`suggestion_length`} words.
Q4. Provide an updated category table based on your suggestions in Q3.
Guidelines:
- You can edit category names, descriptions, or remove a category.
- Updated categories should meet the intrinsic and extrinsic quality requirements above.
- Only propose categories supported by the input data; do not hallucinate.
- Ignore and do not generate categories for low-quality or ambiguous input data points.
- Total number of categories should be no less than {`num_clusters`}.
 # Output Format: Provide your answers between the following tags:

```
<rating>Your rating score (0-100).</rating>
<explanation>Your explanation for the rating.</explanation>
<suggestion>Your suggestion for improvement.</suggestion>
<category_table>Your updated category table.</category_table>
```

 # Input Text Data: {`minibatch_table`}
 # Reference Category Table: {`reference_table`}

Panel F: Taxonomy Update

Online Appendix Table 1. Prompt Template (C'd)

 # Instruction
- You are a concise, structured text processor. Read the reference category table carefully and answer the following questions according to the provided guidelines.
- The reference category table is a markdown table containing predefined categories for labeling the informational content discussed at IASB meetings. The table includes the following columns:
• **id**: category index.
• **name**: category name.
• **description**: category description.

 # Questions
Q1. Evaluate the reference category table and provide a rating score (0–100; higher is better).
Guidelines:
- Use case: The category table serves as a taxonomy for labeling the informational content that board members contribute to justify or support their views during IASB meetings.
- The table should be formatted as a markdown table with the following columns:
• **id**: category index (start from 1, incrementally).
• **name**: concise, clear category name ({`cluster_name_length`} words max), using a noun or verb phrase.
• **description**: brief, specific explanation of the category ({`cluster_description_length`} words max) that distinguishes it from all others.
- The name and description should be consistent with each other.
- Total number of categories should be no less than {`num_clusters`}.
- The categories should:
• Capture key aspects of the information content (e.g., empirical evidence, statistics, references to rules).
• Be meaningful, detailed, and fine-grained to best reflect the information content discussed at IASB meetings.
• Be orthogonal (non-overlapping), specific, clear, and mutually exclusive.
• Avoid vague labels ("Other", "General", "Unclear", "Miscellaneous", "Undefined") or overly broad ones.
• Enable consistent classification of new IASB meeting content.
Q2. Explain your rating score in Q1 within {`explanation_length`} words.
Q3. Do you recommend edits to improve the category table (even for a minor improvement such as improving clarity and/or eliminating overlap)?
- If yes, suggest edits within {`suggestion_length`} words.
- If not, output "N/A".
Q4. If you recommended edits in Q3, provide an updated category table.
Guidelines:
- You can edit category names, descriptions, or remove a category.
- You can merge or add new categories if needed.
- Your updated categories should meet the requirements specified in the above guidelines.
 # Output Format: Provide your answers between the following tags:

```
<rating>Your rating score (0-100)</rating>
<explanation>Your explanation for the rating.</explanation>
<suggestion>Your suggested edits.</suggestion>
<category_table>Your updated category table.</category_table>
```

 # Reference Category Table: {`reference_table`}

Panel G: Taxonomy Review

Online Appendix Table 1. Prompt Template (C'd)

 # Instruction
- You are a professional text annotator. Read the input text data and the reference category table carefully, and answer the questions below according to the provided guidelines.
- Input text data: A summary of a speech segment from an International Accounting Standards Board (IASB) meeting.
- Reference category table: A markdown table containing predefined categories, used as a taxonomy for labeling the informational content discussed at IASB meetings. The table includes the following columns:
• **id**: category index.
• **name**: category name.
• **description**: category description.

 # Questions
Q1. Assign the input text to the single most relevant category from the reference category table.
Guidelines:
- For your output, include:
• **category_id**: the id of the most relevant category in the reference table. If you cannot reasonably classify the input text, output "-1".
• **category_name**: the corresponding name from the reference table. If you output "-1" for `category_id`, use "Undefined" here.
• **explanation**: a concise justification (max {`explanation_length`} words) for your classification, or explain why no match was possible.
• **confidence_score**: an integer from 0 (lowest) to 100 (highest) reflecting your confidence in the assignment.
- The `category_id` and `category_name` must exactly match the values in the reference table.
- Assign only one category to the input text; if more than one could apply, select the single most relevant.
- Choose the category that best matches the key aspects of information content presented in the input text.
- Prefer a reference table category unless it is clearly impossible to classify the input text under any provided category.
 # Output Format: Provide your answers between the following tags:

```
<category_id>Your identified category id.</category_id>
<category_name>Your identified category name.</category_name>
<explanation>Your explanation for the classification.</explanation>
<confidence_score>Your confidence score (0-100).</confidence_score>
```

 # Input Text Data: {`text`}
 # Reference Category Table: {`taxonomy_table`}

Panel H: Label Assignment

Notes: This table presents the prompt templates used to construct the study's hierarchical taxonomy, with explicit definitions, edge-case handling, and structured outputs. **Panel A. Filtering for Substantive Content.** Template that screens segments and keeps only substantive content. **Panel B. Stakeholder Orientations.** Template defining the stakeholder groups. **Panel C. Forms of Expertise.** Template defining the subcategories of expertise. **Panel D. Use-case Summary.** Template for producing a short summary for each substantive segment, functioning as a concise and informative feature representation of the original text. **Panel E. Taxonomy Creation**, **Panel F. Taxonomy Update**, and **Panel G. Taxonomy Review** are templates for creating and refining a label taxonomy using the summaries from the previous step, following the TnT-LLM workflow. **Panel H. Label Assignment.** Template for assigning the fine-grained topic to each substantive segment using the taxonomy generated from the previous step. Upon completion of the project, the complete codebook for constructing the taxonomy, along with the full prompts, will be deposited on OSF.

Online Appendix Table 2. Fine-grained Topic List

ID	Label Name	Label Description
1	Definitional clarity and ambiguity	Clarifying or refining meanings, scope, or boundaries of terms or concepts in standards/framework to eliminate ambiguity; distinct from drafting precision (see 2) and terminology consistency (see 3).
2	Drafting precision and readability	Micro-level phrasing or sentence edits that affect interpretation or readability but not the substance of requirements; removing redundancy and unclear sentences; distinct from conceptual changes (see 1) and terminology consistency (see 3).
3	Terminology consistency across standards	Ensuring the same terms are used consistently across standards/framework; avoiding drift or unexplained terminology changes; distinct from conceptual changes (see 1) and drafting precision (see 2).
4	Recognition timing	Determining start/end recognition points (e.g., contract inception, delivery, vesting, enactment, triggering events).
5	Derecognition criteria	Conditions for removing assets/liabilities (transfers, extinguishments, partial derecognition, modifications).
6	Asset control and rights analysis	Assessing control and bundles of enforceable rights to determine asset boundaries and recognition.
7	Economic resources and benefit criteria	Evaluating whether an item has the potential to produce economic benefits and thus qualifies as an economic resource.
8	Economic consequences analysis	Assessing behavioral/market impacts and unintended consequences of proposals.
9	Reporting consistency and comparability	Comparability and consistency of financial statements/reports across entities/periods; addressing unnecessary diversity.
10	Recognition and presentation thresholds	Setting/applying thresholds for recognition or presentation; evaluating whether/how "probable" thresholds affect recognition decisions.
11	Disclosure requirements	What to disclose (topics/content), why it is decision-useful, and where to locate it (notes versus face); distinct from disclosure materiality (see 14) and structuring/order (see 104).
12	Presentation format flexibility	Choosing format (e.g., tables versus narrative) without altering content; balancing prescription and flexibility; distinct from rigid table issue (see 97).
13	Information completeness and transparency	Ensuring the reporting package covers all necessary information for faithful representation; stressing transparency in financial reporting, including explicitness of rationale.
14	Disclosure materiality	Applying materiality to determine the extent of disclosure; avoiding boilerplate and overload; distinct from what to disclose (see 11) and structuring/order (see 104).
15	Empirical facts and data use	Use of empirical facts, data, or observed market behavior to support positions.
16	Performance statement structure rationale	Rationale for structuring P&L, OCI, and cash flows (ordering, subtotals); distinct from P&L primacy (see 17) and performance definition (see 90).
17	P&L primacy and OCI role	Whether P&L is the primary performance measure and how OCI complements it; whether/when OCI is considered an exception; distinct from performance statement structure rationale (see 16) and performance definition (see 90).
18	Recycling between OCI and P&L	Principles/conditions for reclassifying items between OCI and P&L.

Online Appendix Table 2. Fine-grained Topic List (C'd)

ID	Label Name	Label Description
19	Comprehensive income predictive value	The predictive usefulness of comprehensive income versus earnings for future performance.
20	Measurement bases selection	Debating on the choice, application, or consistency of accounting measurement bases (e.g., fair value, current value, amortized cost, historical cost); justifying multiple measurement bases within/across statements; distinct from cash flow–based model selection (see 29).
21	Fair value hierarchy usage	Applying the IFRS 13 Level 1/2/3 input hierarchy and prioritizing observable inputs.
22	Investments and portfolio attributes	Investment-specific considerations (e.g., strategic stakes, funds, block discounts/premiums, stewardship considerations); whether to use price-times-quantity or adjust for control/block premiums in measuring holdings; distinct from minority/NCI rights and protections (see 30).
23	Unit-of-account: aggregation versus disaggregation	Selecting the aggregation level/unit of account (individual item, component, bundle, portfolio) for recognition/measurement/presentation; distinct from lease-specific unit-of-account topics (see 50).
24	Estimation uncertainty and disclosure	Whether measurement input uncertainty constrains recognition or necessitates disclosure; distinct from auditability concerns (see 94).
25	Day-one gains/losses	Recognition of initial gains/losses and justifications (e.g., calibration versus prohibition).
26	Discount rates and changes	Selecting discount rates and updates for measurements; distinct from lease-specific topics (see 52).
27	Inflation effects in measurement	Whether and how to reflect inflation (indexation) in cash flows or discount rates.
28	Amortized versus historical cost	Distinguishing and applying amortized cost versus historical cost, including accretion and updates.
29	Cash flow–based measurement	Evaluating various cash flow-based models (e.g., fulfillment cash flows); distinct from general measurement bases selection (see 20).
30	Non-controlling interests (NCI) issues	Minority (shareholder) protection, other NCI issues and related presentation; distinct from other investment-specific considerations (see 22).
31	Hedge accounting and bridging	Addressing mismatches via hedge accounting, OCI bridges, or related mechanisms.
32	Entry versus exit price distinction	Distinguishing the difference between entry versus exit prices and implications (IFRS 13).
33	Market structure, liquidity, and reliability	The impact of market design/microstructure, trading conventions, liquidity, bid-ask spreads, and market activity on measurement reliability and the availability and quality of observable prices.
34	Outreach policy and response boundaries	Clarity on webinars, emails, and other outreach channels, and limits on responding to entity-specific queries; distinct from outreach results (see 79) and outreach effectiveness (see 99).
35	Business model influence	Role of business activities in measurement/presentation without dictating outcomes.
36	Goodwill and intangibles treatment	Recognition/measurement of goodwill and other intangibles (e.g., impairment-only versus amortization, valuation challenges).
37	Capital maintenance role	Placement and role of capital maintenance (financial versus physical) in the framework and OCI.

Online Appendix Table 2. Fine-grained Topic List (C'd)

ID	Label Name	Label Description
38	Liability recognition and measurement	Recognition timing and measurement aspects specific to obligations and present obligations.
39	Own credit risk in liabilities	Whether/how own credit changes affect liability measurement and OCI/P&L effects.
40	Risk exposure and disclosure	Disclosing risk exposures, sensitivities, and asymmetric outcomes relevant to users.
41	Contractual options and uncertainty	Role of prepayment/renewal/cancellation options in contracts and their impact on recognition/measurement/classification.
42	Due process legitimacy	References to comment periods, procedural legitimacy, timelines, and other due process issues.
43	Substance over form	Prioritizing economic substance over legal form in recognition and presentation.
44	Combined financial statements and carve-outs	Use and conditions for combined financial statements and carve-outs.
45	Mezzanine and secondary equity	Considering intermediate equity categories and their conceptual justification.
46	Control versus risks-and-rewards	Whether control or risks/rewards drive recognition and derecognition decisions.
47	User needs and preferences	Prioritizing user (investor/analyst) needs in guiding recognition, measurement, and disclosure, including information needs specific to creditors (e.g., liquidity, maturities, covenants).
48	Bundled contracts and leases	Distinguishing leases from services; separating components within bundled contracts for accounting.
49	Foreign currency translation and recycling	IAS 21 translation issues, OCI versus P&L effects, and recycling criteria.
50	Lease unit-of-account	Unit-of-account considerations specific to lease assets and liabilities; distinct from general unit-of-account topics (see 23).
51	Lease term and options	Determining lease term, renewal/cancellation assessments, and option incentives.
52	Lease discount-rate determination	Deciding implicit versus incremental borrowing rates and updates for leases; distinct from general discount-rate topics (see 26).
53	Lessor versus lessee models	Differences in economics and accounting models for lessors versus lessees.
54	Lease price changes	Accounting for scope or price changes and substantiality assessments in leases.
55	Short-term and small-ticket leases	Treatment and disclosure of short-duration or low-value leases; exemptions.
56	Right-of-use (ROU) asset measurement	Initial/subsequent measurement and labeling/presentation of ROU assets.
57	Hypotheticals and scenarios	Using scenarios to anticipate implications.
58	Reputation and credibility	Managing perceptions of IFRS credibility and public trust in outputs.
59	Going concern assumption	Treatment/primacy of going concern versus liquidation perspectives.
60	Lease presentation in statements	Effects of leases on P&L and cash flow classification (e.g., interest versus principal, netting).
61	Revenue guidance at portfolio level	Applying revenue guidance at portfolio level and materiality expedients.

Online Appendix Table 2. Fine-grained Topic List (C'd)

ID	Label Name	Label Description
62	Contract combination and allocation	Combining contracts and allocating consideration/discounts among performance obligations.
63	Sales and buybacks	Accounting for sales with repurchase options, buybacks, and related features.
64	Educational versus authoritative guidance	Distinguishing educational materials/webinars from authoritative texts.
65	Basis for conclusions (BC) adequacy	Adequacy of BC explanations and rationales for decisions.
66	Rate-regulated activities and entities	Recognition of rights/obligations in rate-regulated entities.
67	Financial institution-specific topics	Bank/insurer-specific topics (liquidity, mixed models, regulatory overlays).
68	Continuous recognition needs	Requirements for ongoing reassessment/recognition (e.g., provisions, variable consideration, option re-measurement).
69	Liability transfers and settlement	Measuring transferred liabilities and expectations about settlement (timing/form).
70	Estimate changes and decommissioning	Retrospective updates to estimates (e.g., decommissioning, provisions).
71	Digital (XBRL) and layered reporting	Effects of digital reporting (e.g., XBRL) and layered disclosures on communication quality.
72	Application issues across markets	Practical application challenges and feasibility across markets/jurisdictions (execution).
73	Transition and change management	Adoption timing, transition methods, and comparability during change.
74	Informational benefits and costs	Balancing preparer/user costs with informational benefits when choosing requirements; evaluating compliance-related workload and other costs for preparers/users.
75	Process efficiency and simplification	Calling for prioritization, narrowing scope, sequencing work, and avoiding distractions; arguing for simplification and against undue complexity.
76	Enforcement and governance environment	Role of enforcement, governance, and controls in consistent application and comparability.
77	Stewardship versus valuation objective	Positioning stewardship versus valuation/decision-usefulness as reporting objectives.
78	Reporting gap and adjustment	Users have to adjust/reconstruct reported numbers due to reporting gaps.
79	Outreach results and stakeholder feedback	Evaluating outreach results (comment letters, stakeholder meetings) and post-implementation review (PIR) findings to inform decision-making; distinct from outreach policy (see 34) and outreach effectiveness (see 99).
80	Cash flow classification and noncash	Classifying cash flows (operating/investing/financing), vendor financing, and noncash transactions.
81	Board voting and consensus	References to vote results or consensus-building.
82	Staff paper analysis and critique	Reliance on or critique of staff papers, analyses, and recommendations.
83	Non-exchange transactions and grants	Recognition and presentation of gifts, grants, lawsuit proceeds, and other non-exchange events.

Online Appendix Table 2. Fine-grained Topic List (C'd)

ID	Label Name	Label Description
84	Framework–standard boundary	What belongs in the conceptual framework versus standards-level guidance; authority of the framework versus standards; when to lean on/depart from the framework.
85	Convergence with US GAAP	Harmonization or managed divergence with US GAAP and implications.
86	Regulatory influence and independence	Considering regulators while maintaining IASB independence and due process.
87	National practices and differences	Home-country or regional practices and jurisdictional differences.
88	Translation burden and issues	Translation issues in IFRS standards/Framework and their impact on clarity and faithful interpretation.
89	Sector/industry applicability	Sector-specific issues and suitability of general rules in industries.
90	Performance definition and boundaries	Defining "performance," its components, and period attribution; distinct from performance statement structure rationale (see 16) and P&L primacy (see 17).
91	Primary users and inclusivity	Ensuring inclusion of primary-user perspectives.
92	Measurement robustness and bias	Risks of manipulation, optimism bias, and verifiability of measures.
93	Reliability versus relevance trade-off	Explicit trade-offs or prioritization between reliability and relevance.
94	Auditability and assurance	Whether information is capable of being audited and implications for assurance; distinct from estimation uncertainty (see 24).
95	Rebuttable presumptions	Burden of proof when deviating from default treatments or presumptions.
96	Examples and illustrations	Relying on examples for illustration, or request additional/clarification examples.
97	Table-driven misinterpretation	Risks that rigid, table-heavy formats cause misunderstanding or boilerplate; distinct from format flexibility choices (see 12).
98	Objective articulation	Clarifying and positively framing objectives to avoid redundancy or confusion.
99	Outreach effectiveness and actionable input	Improving methods/materials to obtain actionable input from users (investors), preparers, or other stakeholders; distinct from outreach results (see 79) and outreach policy (see 34).
100	Historical precedent and legacy	Using past practices/decisions to justify choices; legacy effects.
101	Status quo versus change	Arguments to retain or alter current rules/practices and their justifications.
102	KPI interpretive cautions	Highlighting risks of key performance indicator (KPI) misinterpretation and communicating necessary caveats/context.
103	Interdependencies across standards	Linkages among standards; consequential amendments and conflict resolution.
104	Disclosure structure and order	Ordering of disclosure objectives/items for clearer, more digestible communication; distinct from what to disclose (see 11) and disclosure materiality (see 14).

Notes: This table presents the study's detailed-level taxonomy ("fine-grained topics"). It lists the complete set of mutually exclusive topic labels with concise, operational descriptions produced via the TnT-LLM workflow, followed by conservative human-in-the-loop edits.