# Assessing Supply Chain Management Practice Disclosures in Annual Reports of Listed Home Appliance Manufacturers

**Yu Yin, Jinru Li*, Yurong Zhao**
*Beijing Union University, Beijing, China*
**Corresponding Author*

**Abstract: The annual report provides the public with information on the organization of procurement and production logistics. Identifying specific management practices from the scattered annual report text is the key to evaluating these disclosures. This study examines 60 annual reports released by 20 listed home appliance manufacturers between 2023 and 2025, and builds an eight-evaluation framework covering the four dimensions of procurement and supply security, production and delivery coordination, inventory and logistics management, risk monitoring and emergency response. The rule-based text analysis method is used to identify the management object, operational actions and implementation status, and test the output results through manual review, expected coverage under common text length, and practical fragment density. Among the 13,373 qualified fragments, 187 practice fragments were analyzed and identified, and the three-year average disclosure coverage scores were 24.38, 26.88 and 25.63 respectively. The identified practices were mainly concentrated in the two dimensions of production and delivery and inventory and logistics. The increase in the number of practical fragments is not accompanied by the continuous expansion of project coverage. Under the common length of 68 fragments, the expected coverage scores are 11.21, 13.36 and 11.40 respectively, indicating that the evaluation results are sensitive to the disclosure length. The micro-F1 value of the prototype rule is 25.93%, which is lower than 32.43% of the keyword baseline. The independent evaluation of the revised rules has yet to be completed. This study mainly provides a traceable framework for organizing disclosed supply chain practices and shows how identification rules and text length affect disclosure evaluation.**

**Keywords: Supply Chain Management; Corporate Disclosure; Annual Report Text Analysis; Home Appliance Manufacturing; Rule-Based Identification**

## 1. Introduction

Supply chain management involves coordination within and between enterprises [1]. The process perspective is to connect procurement, production, inventory and delivery through cross-functional relationships [2]. The coordination of these activities directly affects the availability of materials, production continuity and order fulfillment. For home appliance manufacturers, multi-product categories, multi-part demand and multi-sales channels make cross-operation coordination very important. Income, profit and asset scale describe the business results of the enterprise, but direct information on how the enterprise organises procurement, adjusts production plans or manages delivery is limited. Identify specific management practices to explain the reasons for these organisational arrangements and business results.

Corporate disclosure is a channel for releasing information about the company to outside investors, and at the same time, management's reasons for reporting are also revealed [3]. The public source of information for operational arrangements of listed companies is their annual reports. Supplier Qualification Audit, Demand and Production Alignment, Inventory Control and Distribution Practices are frequently found in the Management Discussion and Analysis section. Although these narratives add some operational details to the financial figures, they also pose problems of comparability. For example, one enterprise may describe the manufacturing execution system as supporting production scheduling, while another enterprise discusses the role of the system in product quality testing. Therefore, only statistical system names or supply chain keywords may mix activities with different operational meanings. Li [4]'s research

 Copyright @ STEMM Institute Press

shows that the length and readability of the report are related to the surplus characteristics, suggesting that the text form can be considered when interpreting the disclosed information. Longer reports will mention more assessment items, so the difference in scores may partially reflect the difference in the length of disclosure.

Supply chain research provides a conceptual basis for classification management activities. Zheng et al. [5] summarized the meaning and dimensions of supply chain-related capabilities, and emphasized the consistency between concepts and measurements. Chen and Paul Laji [6] and Li et al. [7] developed ideas and measurement tools for organizational relations, information sharing and internal management practices. Enterprise text research promotes the quantification of qualitative data via domain dictionaries, text features and external validation [8-10]. In short, these studies suggest specifying the contents for observation in operational terms and determining whether the text evidence supports each project. When applying this method to China's annual report, specific actions need to be distinguished from general statements, as well as their implementation status, business scope and text length.

This paper will investigate 60 annual reports of 20 listed home appliance companies in 2023-2025 to answer the following questions: Which management practices are disclosed? How are these identified practices distributed in supply chain activities, and how do the measured indicators obtained vary according to the identification rules and text length.An eight-item framework covering four dimensions is implemented through local operation mode and context filtering rules. The process retains source text paragraphs, page number positions and project mappings, and examines annual changes, dimensional structure and measurement sensitivity. Disclosure coverage is defined as the proportion of items supported by rule-identified evidence within the specified text range, measuring the observable breadth of practical disclosure.

The contribution of this research is mainly reflected in two aspects. First, transform procurement, production, inventory and logistics and risk response practices into observable annual report projects, so that disclosure can be associated with specific operational activities. Second, integrate the management object, implementation status and source location into a traceable workflow, and report the identifiable performance and text length sensitivity. Thus, it can be seen how the evaluation indicators are formed: Enterprises verify the disclosed information based on a specific text paragraph, and researchers observe how project definitions and identification rules affect output results.

## 2. Literature and Assessment Framework

### 2.1 Definition of Supply Chain Management Pratices

Supply chain capabilities, management practices and operational performance refer to resource integration capabilities, specific organisational actions and results, respectively. Zheng et al. [5] have shown in their review that different research subjects require different measurement dimensions. This paper focuses on disclosed management practices and uses specific actions such as supplier evaluation, order tracking and replenishment arrangements as text evidence. The above definition assigns an actual working meaning to each project and thus provides a basis for verifying whether the supporting materials are consistent with the source report.

Chen and Paul Raji [6] developed the measurement concept around the supply chain relationship and coordination. Li et al. [7] distinguished strategic supplier partnership, customer relationship, information sharing, information quality, internal lean practice and delay strategy. These studies provide reference for the conceptual scope of procurement relationship, process coordination and information exchange considered in this article. Since questionnaire responses and public disclosure provide different forms of evidence, the items in this article are defined as management objects and actions that can be observed in the annual report. Flinn et al. [11] distinguished internal integration, supplier integration and customer integration, and examined their relationship with performance. This distinction separates the possible results of operational coordination and coordination.

Pegel [12]'s research on operation, procurement and logistics integration emphasizes cross-functional coordination. Accordingly, the management meaning of information technology is determined by the business process embedded in it. This study considers the system name as an identification clue and examines its relationship with procurement, scheduling, warehousing or

Copyright @ STEMM Institute Press 

delivery. Tang [13] summarized supply chain risk management from the perspective of supply, demand, product and information management, and believed that both daily operations and responses to interruptions should be paid attention to.

### 2.2 Measuring and Validating Corporate Text

Enterprise text analysis needs to pay attention to the meaning of the domain and the measurement verification at the same time. Lauran and McDonald [8] showed the possible misalignment between the general dictionary and the corporate language. Grimmer and Stuart [9] emphasized that verification should be aimed at specific analytical tasks. Following these principles, this study uses the supply chain dictionary to generate candidate fragments, applies filter rules about the operation object, implementation status and context, and compares the rule output with manual judgment. The review of Lauren and McDonald [14] further emphasizes the choice of methods in financial text analysis and reinforces the need to record dictionaries, text scope and verification procedures.

Chinese research also shows how text indicators are related to external information. Chen and Chen [10] construct a risk information index from text intonation and complexity, and examine its relationship with credit rating and other results. Liu Heshi [15] found that the inquiry letter of the stock exchange will affect the management language in the annual report, indicating that the external governance conditions will affect the disclosure expression. Wu [16] measured the risk exposure of the supply chain from the surplus conference call and verified the indicator using risk events and enterprise results. These studies show that the interpretation of text indicators depends on their objectives to build and verify evidence. This study focuses on the structure of disclosed management practices, and the correlation with business results is the subject of subsequent external verification.

Therefore, literature provides a complementary basis for supply chain measurement and enterprise text analysis. This study connects the two by mapping from the concept of operation to the evidence of the annual report. The framework focuses on organizing practical evidence, identifying implementation status, and establishing comparable measurement definitions. Its application is tested through project distribution, manual review and sensitivity analysis.

### 2.3 Four Dimensions and Eight Disclosure Items

The framework follows supply chain activities and comprises procurement and supply assurance, production and delivery coordination, inventory and logistics management, and risk monitoring and contingency response. The first two dimensions are supply-side arrangements and the transition from demand to production and delivery. Inventory and logistics management handle material and product flows, risk monitoring and contingency response address abnormal circumstances and continuity plans, etc. Each dimension has the two items shown in Table 1.

**Table 1. Framework for Assessing Disclosed Supply Chain Management Practices**

| Dimension | Items | Evidence considered |
|---|---|---|
| Procurement and supply assurance | Supplier management; supply source assurance | Supplier qualification, audits, evaluation, and classification; alternative sources and supply assurance arrangements |
| Production and delivery coordination | Production coordination; delivery management | Demand–production alignment; order tracking, delivery lead times, and fulfillment management |
| Inventory and logistics management | Inventory management; logistics coordination | Replenishment and inventory tracking; warehousing, distribution, and logistics information coordination |
| Risk monitoring and contingency response | Risk monitoring; contingency adjustment | Identification of supply risks; contingency plans, substitution, and continuity arrangements |

The consequences of a practice are affected by its operating conditions. Tomlin [17] shows that the value of inventory, reliable sourcing and contingent responses depend on disruption characteristics, capacity and supplier flexibility. Extra inventory can cover shortages but requires more funds. There are also trade-offs among reliability, cost and flexibility in the selection of sources [13,17]. Therefore, the scores for these management arrangements are based on whether supporting evidence has been disclosed; no overall performance ranking is assigned to specific strategies. A segment may support several items, but repeated disclosure of the same item is counted only once.

## 3. Data and Text Processing

### 3.1 Sample Selection and Reporting Years

A Purposeful Sample of 20 A-share listed companies engaged in home appliance

 Copyright @ STEMM Institute Press

production will be selected. The enterprises are classified into the following categories: Diversified and major appliance manufacturers, kitchen appliance producers, and small-appliance-related manufacturers, as shown in Table 2. A fixed set of firms is observed over three years to obtain 60 firm-year observations. Groups have certain primary business characteristics, and the observation unit is the consolidated listed entity.

**Table 2. Sample Composition**

| Business group | Firms | Count |
|---|---|---|
| Diversified and major appliances | Midea Group; Haier Smart Home; Gree Electric Appliances; Hisense Home Appliances; Changhong Meiling; Whirlpool China; Aucma; TCL Smart Home | 8 |
| Kitchen appliances | Robam Appliances; Vatti; Vanward; Zhejiang Meida; Marssenger; Shuaifeng Electric Appliances | 6 |
| Small appliances and related businesses | Supor; Joyoung; Bear Electric Appliance; Xinbao Electrical Appliances; Flyco Electrical Appliance; Kingclean Electric | 6 |

The materials are annual reports for fiscal years 2023, 2024, and 2025 disclosed through CNINFO. Fiscal years and announcement dates are stored separately. One candidate report record is kept per firm-year, and the download log has no errors. The source records have announcement titles, dates and links. Stock codes will be used to unify changes in company names, such as TCL Smart Home and its previous name, Homa Appliances. A-share and H-share listings are not counted separately, and brands are not considered independent listed entities.

### 3.2 Text Extraction and Section Alignment

The program will extract the text of each page in a PDF using pypdf and maintain the association with that page. The average number of Chinese characters per page and low-text pages are saved to check for scanned content and encoding errors. Then, it will find the start of the management discussion and analysis or discussion and analysis of operations section, and use the following sections, such as corporate governance, as the end boundary. Content after the heading "Future Development Outlook" is not included in the main analysis of implemented practices within this range. Retain the section divisions for inspection.

Cleaning removes the obvious headers and page numbers, merges layout-induced line breaks, and splits text at Chinese sentence-ending punctuation and semicolons. Text across pages is combined, and each section is assigned a start PDF page. Eligible segments have 15-1,500 characters and contain at least 10 Chinese characters. Exact duplicate segments in a report are not included. The resulting corpus has 13,373 eligible segments. A segment is an analysis unit generated by these processing rules.

### 3.3 Manual Review and Rule Versions

Manual review includes 44 first development records, 120 prototype-review sections and 20 boundary-case development segments, all reviewed by a single annotator. The 20 boundary cases are selected from the 120 prototype-review sections, and the 44 initial records also contain item-based selections. Therefore, these figures will not be added to arrive at a total number of independent texts. The initial frozen rules are denoted as V1, and the rules revised based on the review results are denoted as V2.

V1 was frozen, and 40 segments were sampled from each of the three mutually exclusive strata: rule-accepted, candidate-rejected, and unmatched. The original sizes of the strata were 153, 112 and 13,080. A single segment can have multiple items, and these items may be undecided. The eight items had a total of 960 segment-item judgments, and among them, 914 were certain and 46 were uncertain.

The above 20 boundary cases are for reference in implementation status and evidence specificity. Finally, there were eight implemented-practice segments, six general descriptions, and six irrelevant or uncertain segments. The two segments that were first labelled "supply chain description" were found to be implemented practices after verification by the annotator. Segment-level classifications were given to these cases without re-annotating all eight items.

The Study presents the post-freeze review of V1 independently of the development comparison for V2.Once the 120 segments had informed V2 revision, they could no longer provide an independent test of V2. Original labels, clarification records, and rule versions provide the record of this development process.

## 4. Practice Identification and Measure Construction

### 4.1 Identifying Operational Objects and Actions

V1 requires the common appearance of project terms, action terms and mechanism terms.

Although this method is easy to implement, it may combine weakly related words in the same fragment and omit the specific practice of not using words such as "system" or "mechanism". V2 is changed to match the local operation mode in comma-separated sentences, such as supplier audit and evaluation, alignment of demand with production plan, order tracking, and replenishment or inventory control. The rules are implemented in regular expressions, and each match produces a local source text paragraph and a project mapping.

The information system is evaluated by its operational purpose. The manufacturing execution system (MES) used for quality inspection will not be automatically included in production coordination. The supplier relationship management system (SRM) associated with online supplier operations, or the warehousing management system (WMS) associated with warehousing activities, can provide evidence for the corresponding project. The connection mode also captures the system integration description across comma boundaries and preserves the relationship between the system and the action.

### 4.2 Filtering Implementation Status and Business Context

Time filtering expresses the state of implementation between sentences. After the expression of the plan (such as "will be implemented" or "plan to be established"), the follow-up action will remain in the planned state until a clear completed or current operational status appears. If the sentence only summarizes an implemented project to lay the foundation for future development, the above-mentioned implemented practices should be retained. This rule is used to deal with paragraphs that mix current implementation with future goals.

The business background excludes clearly identifiable descriptions of logistics equipment or solutions sold to customers, foreign exchange hedging and other financial transactions and general industry trends. The practice of diversified enterprises is recorded at the main level of the consolidated report, and the cost management involving multiple activities has been implemented. It is kept separately when it cannot be attributed to eight projects. Such evidence is included in the number of practice fragments, but the eight coverage scores are not obtained.

### 4.3 Disclosure Coverage and Practice Density

Let $x_{itj}$ denote item j for firm i in year t. The item equals one if at least one segment supports it and zero otherwise. The overall disclosure coverage score $C_{it}$ gives equal weight to all eight items; each dimension score gives equal weight to its two items. A missing report is recorded as missing. An item value of zero means that no supporting evidence was identified within the specified sections and rules.

$$C_{it} = \frac{100}{8}\sum_{j=1}^{8} x_{itj}\,;\, C_{itd} = \frac{100}{2}\sum_{j\in Jd} x_{itj} \quad (1)$$

Here, $J_d$ contains the two items belonging to dimension d. Both $C_{it}$ and $C_{itd}$ range from 0 to 100. Each item is counted once per firm-year, so the index describes the breadth of disclosed items. Equal weighting gives each item the same contribution to the overall score.

Let $M_{it}$ be the number of distinct identified practice segments and $N_{it}$ the number of eligible segments. Practice density per 1,000 segments is defined as follows.

$$D_{it} = 1000 \times \frac{M_{it}}{N_{it}} \quad (2)$$

Annual coverage and density statistics are equally weighted means across the 20 firms. The mean of firm-level densities differs from the pooled ratio of total practice segments to total eligible segments. Evidence concerning governance across activities is included in $M_{it}$ but excluded from the eight-item coverage score.

### 4.4 Text Length Sensitivity and Rule Output Comparisons

The common-length calculation uses the same without-replacement occurrence probability as analytical rarefaction [18], applied here separately to each disclosure item. Unlike mutually exclusive species categories, a segment may support several items; linearity of expectation still permits summing their marginal occurrence probabilities. To compare coverage at a common text length, consider a report with N eligible segments, of which $m_j$ support item j. If L segments are sampled with equal probability and without replacement, the probability $p_j$ of observing item j at least once and the expected coverage score $C^{(L)}$ are:

$$p_j = 1 - \frac{C(N - m_j, L)}{C(N, L)}$$

$$C^{(L)} = \frac{100}{8}\sum_{j=1}^{8} p_j \quad (3)$$

The binomial coefficient $C(a, b)$ denotes the

 Copyright @ STEMM Institute Press

number of ways to select b objects from a; it is set to zero when $a < b$. The common length L is 68 segments, equal to the shortest report in the sample. Probabilities are calculated directly. This measure describes expected coverage at a common sampling length and is used to examine length sensitivity.

The analysis also computes the Spearman correlation between coverage and segment count, and recalculates coverage after removing the two risk monitoring and contingency items. Annual V1 and V2 outputs describe differences between versions, while identification performance is evaluated separately through manual review. The difference between the prototype sampling frame and final corpus is discussed in Section 6; the difference in corpus definitions prevents attributing output differences entirely to rule revision. Because the observations follow 20 firms repeatedly, the analysis is descriptive and does not treat the 60 firm-years as independent observations in significance tests.

## 5. Results

### 5.1 Annual Disclosure Coverage and Practice Density

The 60 annual reports contain 13,373 qualified fragments. V2 identified 187 different practice fragments and generated 241 evidence mapping records, including separately retained governance records across multiple activities. Since a fragment can support multiple projects, the number of mappings is different from the number of fragments of different practice. Table 3 summarizes the scale and coverage of the annual corpus. These are the descriptive outputs of the V2 development stage.

**Table 3. Annual Corpus Size and Disclosure Coverage**

| Year | Eligible N | Practice N | V1 | V2 | D/1k |
|---|---|---|---|---|---|
| 2023 | 4,225 | 55 | 25.00 | 24.38 | 10.63 |
| 2024 | 4,574 | 65 | 28.75 | 26.88 | 12.92 |
| 2025 | 4,574 | 67 | 28.75 | 25.63 | 12.99 |

Note: Coverage is measured in index points. Density is practices per 1,000 segments, averaged equally across firms. Each year includes 20 firms.

The average disclosure coverage rose from 24.38 to 26.88, and then fell to 25.63. During the same period, the number of practical fragments increased from 55 to 65 and 67, and the average density of enterprises increased from 10.63 to 12.92 and 12.99 (per 1,000 fragments). Therefore, more practical description fragments are not accompanied by the continuous expansion of project coverage. Enterprises may provide more details to existing projects while maintaining the same breadth of disclosure.

### 5.2 Distribution across Dimensions and Items

Table 4 summarizes mean disclosure coverage by dimension. The average coverage of production and delivery coordination is 47.50, 52.50 and 47.50 respectively in three years, and inventory and logistics management is 37.50, 37.50 and 40.00 respectively. Both are higher than the other two dimensions. The procurement and supply security scores are 12.50, 15.00 and 12.50, and the risk monitoring and emergency response scores are 0, 2.50 and 2.50. The evidence identified is mainly concentrated in the practice of production, delivery and material flow. Therefore, the differences between dimensions reflect both the disclosure and the coverage of the identification rules.

**Table 4. Mean Disclosure Coverage by Dimension**

| Dimension | 2023 | 2024 | 2025 |
|---|---|---|---|
| Procurement and supply assurance | 12.50 | 15.00 | 12.50 |
| Production and delivery coordination | 47.50 | 52.50 | 47.50 |
| Inventory and logistics management | 37.50 | 37.50 | 40.00 |
| Risk monitoring and contingency response | 0.00 | 2.50 | 2.50 |

Note: Scores are in index points. Each dimension equally weights its two items; annual scores equally weight firms.

Table 5 reports the number of firms with identified disclosure for each item. At the project level, the production coordination was identified by 11, 13 and 12 enterprises respectively in the three years, and the logistics coordination was recognized by 9, 10 and 7 respectively. The supply source guarantee is only identified by 1, 1 and 2 enterprises respectively. No evidence of risk monitoring was identified in all 60 reports, so the project does not currently provide a distinction in the sample. This result may reflect limited disclosure, incomplete dictionary coverage or strict evidence standards, rather than a lack of risk monitoring practices.

**Table 5. Number of Firms with Identified Disclosure for Each Item**

| Item | 2023 | 2024 | 2025 |
|---|---|---|---|
| Supplier management | 4 | 5 | 3 |
| Supply source assurance | 1 | 1 | 2 |
| Production coordination | 11 | 13 | 12 |
| Delivery management | 8 | 8 | 7 |
| Inventory management | 6 | 5 | 9 |
| Logistics coordination | 9 | 10 | 7 |
| Risk monitoring | 0 | 0 | 0 |
| Contingency adjustment | 0 | 1 | 1 |

Note: Each year includes 20 firms. A firm can cover multiple items.

Copyright @ STEMM Institute Press 

### 5.3 Sensitivity to Text Length and Item Composition

Under the common length of 68 fragments, the average expected coverage is 11.21, 13.36 and 11.40 respectively, which is lower than the full-text length coverage every year. Both definitions will achieve the highest value in 2024, and the difference between 2025 and 2023 will be smaller under the common length. Figure 1 shows this comparison. The full-text length coverage measurement reports the number of project categories identified at any location in the qualified text. The expected coverage under common length is a measure of the average number of categories that can be observed in the same number of sample texts. The two indicators provide different perspectives for disclosure comparison.

The common sampling length is 68 segments, and the index ranges from 0 to 100.

The Spielman correlation coefficient between V2 coverage and the number of qualified fragments is 0.413. Although the binary score limits the impact of repeated mentions of the same item, the coverage is still directly related to the length of the report. Enterprise size, business scope and disclosure content may jointly affect the length of the text; therefore, this correlation is used to describe the dependency of measurement.

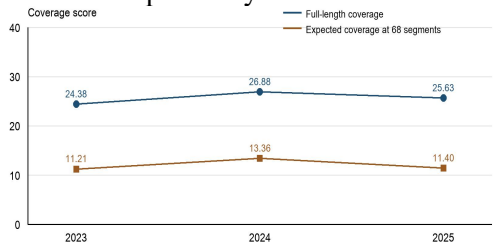



**Figure 1. Annual Full-Length and Expected Common-Length Coverage**

Note: Values are V2 development outputs.

After excluding risk monitoring and emergency adjustment and weighting the remaining six projects, the average annual coverage is 32.50, 35.00 and 33.33 respectively, and the year ranking remains unchanged. Table 6 summarizes the alternative definition. The six-item and eight-item scores have different denominators, and the level differences reflect the changes in the composition of the project rather than the improvement of the management level.

**Table 6. Mean Coverage under Alternative Measurement Definitions**

| Definition | 2023 | 2024 | 2025 |
|---|---|---|---|
| Eight-item full-length coverage | 24.38 | 26.88 | 25.63 |
| Eight-item expected common-length coverage | 11.21 | 13.36 | 11.40 |
| Six-item full-length coverage | 32.50 | 35.00 | 33.33 |

Note: Scores are in index points. The six-item definition excludes risk monitoring and contingency adjustment. The common length is 68 segments.

### 5.4 Identification Performance and Development Comparison

When relevant cases are rare, the accuracy rate and recall rate have high information value [19]. Both and micro-F1 values are reported here at the same time to clarify the trade-off between false positives and missing evidence. For the deterministic fragment-project judgment in the 120 fragment layered review, the accuracy rate of V1 is 33.33%, which is higher than 29.27% of the keyword baseline, but the recall rate is 21.21%, which is lower than 36.36% of the baseline. Its micro-F1 values are 25.93% and 32.43% respectively. Therefore, the additional filtering rules of V1 do not improve the overall recognition performance: the moderate improvement of the accuracy rate means a large loss of relevant evidence. Table 7 presents the precision, recall, micro-F1, and weighted micro-F1 results for the keyword baseline and V1 rules.

**Table 7. Review Results for the Keyword Baseline and V1 Rules**

| Method | Precision | Recall | Micro-F1 | Weighted micro-F1 |
|---|---|---|---|---|
| Keyword baseline | 29.27% | 36.36% | 32.43% | 6.69% |
| V1 rules | 33.33% | 21.21% | 25.93% | 4.60% |

Note: The first three metrics are unweighted review-sample results. The final column uses inverse within-stratum inclusion probabilities from the prototype sampling frame. The 46 uncertain judgments are excluded from evaluation of determinate combinations.

The sampling ratio of the three layers is significantly different. After the inverse probability weighted average, the micro-F1 of the baseline is 6.69% and the V1 is 4.60%. Due to the large scale of unmatched layering, a small number of positive labels in this layer have a greater impact on the weighted results. These estimates reveal the gap between the performance of unweighted samples and the overall projection, indicating that expanding the review of unmatched texts is of great value for evaluating the evidence of omission.

In 20 boundary case development fragments, V2 identified all eight manually confirmed implementation practices. Of the remaining 12 clips, 10 were not recognized, 2 were retained, and 18 clips were agreed upon. The disagreement involves a platform that integrates overseas production, delivery and order information, as

 Copyright @ STEMM Institute Press

well as end-to-end visualization of cross-retail logistics and installation. The annotator classifies it as irrelevant or uncertain and a general description, respectively, and the rules can be retained because they contain expressions of fulfillment or logistics actions. These cases reveal the classification boundaries between channel visualization and specific supply chain operations. Since these fragments are used for development and have obtained binary fragment-level labels, and 18 uniforms constitute a development comparison, they cannot be directly compared with the eight-label F1 of V1.

### 5.5 Examples of Identified Management Practices

Page 22 (PDF) of the annual report of Zhejiang University of Art in 2025 describes demand-driven production, safety inventory and MES platform. The production organization and inventory arrangement provide evidence for two independent projects, and the system name provides the business background. Page 12 of its 2023 annual report (PDF) describes the procurement supply chain platform and material quality, price, delivery and service management, and links the procurement process with clear management objects.

Page 12 of Jiuyang's 2023 Annual Report (PDF) relates SRM with online supplier management, MES and manufacturing process visualization, and WMS with warehousing management.Manual review classifies the paragraph as an implemented practice, and the rules generate project allocation according to the operation object. Therefore, to determine whether the system is for management or not, we need to look at what actions are taken; merely noting their names will not be sufficient.

Page 12 of the 2025 Annual Report of Boss Electric Appliances (PDF) shows a complete closed-loop cost management system for procurement, production, logistics and delivery. The evidence is kept separately as governance across many activities. Multiple operational stages are listed above; therefore, no single score will be assigned to each project. Conversely, statements that highlight supply chain efficiency, inventory reduction or risk prevention but do not specify the implementation of arrangements are considered general descriptions. Distinguish specific practices from intentional statements in evidence mapping.

## 6. Discussion and Limitations

### 6.1 Research and Practical Implications

Connect the operating significance of supply chain practice with specific sections of the annual report to provide a framework that adds to the business-results-based method. Keep local evidence and source pages to split the summary score into the relevant text of the project and its support. The annual results also show an increase in the quantity of practical fragments, an expansion of project coverage, and an improvement in text density. Report these contents separately to determine whether the company is adding new disclosure items or providing more details for the existing projects.

For enterprise reports, the description of the supply chain should clarify the management object, implementation status and applicable business scope. Discussion of digital systems can show how they are used in procurement, scheduling, warehousing or delivery. The platform is for different parts of the company, such as home appliance manufacturing, other types of production and various services offered to consumers. Distinguish among existing arrangements, projects that have been implemented in the reporting year, and future plans to help readers understand the current situation of disclosed practices. Researchers can preserve the sensitivity of the original text paragraphs and report text length to verify and extend subsequent indicators.

### 6.2 Limitations and Further Research

The main deficiencies are the scope of samples and measurement verification. Purposive selection, a three-year period and consolidated reports of 20 enterprises limit the extension to the whole industry, long-term trends, and individual home appliance business divisions. PDF extraction and fixed chapter boundaries may omit images, tables, other chapters, and existing practices discussed in the outlook section. Verification of the different versions of the report has not been recorded in the current records. In addition, the three V1 samples have a total of 13,345 fragments, which is 28 fewer than the 13,373 fragments in the final corpus. The reasons for this difference are not available in the current data, so a direct comparison of the prototype and final output cannot be made.

All reference labels are from a single annotator; therefore, the confidence between annotators

Copyright @ STEMM Institute Press 

cannot be assessed. Atstein and Poesio [20] showed how consistency measures and their assumptions can provide information for evaluating the labelling corpus. V1 has a lower micro-F1 score than the keyword baseline, and V2 lacks independent tests with non-development text. Zero coverage of risk monitoring and the inability to classify some practices across multiple activities also show that there is still room for improvement in project definitions and dictionaries. Independent multi-labeler evaluation and project-level performance estimation based on fixed coding rules will help ensure the reliability and generalisability of the indicator. At that time, the disclosed evidence can be matched with the delivery, inventory and supply interruption information - examine the external relationship. The current findings involve the identifiable disclosure structure in the sample report, and do not establish the ranking of operational capabilities or performance.

## 7. Conclusion

This study uses 60 annual reports of 20 listed home appliance manufacturers from 2023 to 2025 to construct an eight framework covering four dimensions to evaluate the disclosure of supply chain management practices. The workflow connects chapter extraction, operational identification and source tracing. 187 practice fragments were identified out of 13,373 qualified fragments, resulting in structured evidence that public reports are conducive to management arrangements.

The research has found three. First, the average disclosure coverage rose and then fell, with annual scores of 24.38, 26.88 and 25.63 respectively; the identified practices are mainly concentrated in the dimensions of production and delivery as well as inventory and logistics. Second, the number of practical descriptions and the breadth of project coverage do not change in parallel. The number of practical fragments continued to increase throughout the period, while the coverage decreased in 2025, indicating that more descriptive details do not necessarily expand the scope of the disclosed projects. Third, the length of the text and the rule design will affect the interpretation. Under 68 fragments, the expected average coverage is 11.21, 13.36 and 11.40 respectively, while the micro-F1 of V1 is still lower than the keyword baseline. Therefore, more filtering rules need to be added for the specific task.

The study links operational definitions, implementation status and traceable paragraphs to explain the disclosure score with specific evidence. The company will set a limit on how much power it can exercise in this area to improve the transparency of information disclosure. At the same time, for research, we can report the project's scope, density of practice, and text-length sensitivity to differentiate between quantity and structure. Independently label and externally verify the next applications of the framework on large-scale data.

## References

[1] MENTZER J T, DEWITT W, KEEBLER J S, et al. Defining supply chain management. Journal of Business Logistics, 2001, 22(2): 1-25. https://doi.org/10.1002/j.2158-1592.2001.tb00001.x.

[2] LAMBERT D M, COOPER M C. Issues in supply chain management. Industrial Marketing Management, 2000, 29(1): 65-83. https://doi.org/10.1016/S0019-8501 (99)00113-3.

[3] HEALY P M, PALEPU K G. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. Journal of Accounting and Economics, 2001, 31(1-3): 405-440. https://doi.org/10.1016/S0165-4101 (01)00018-0.

[4] LI F. Annual report readability, current earnings, and earnings persistence. Journal of Accounting and Economics, 2008, 45(2-3): 221-247. https://doi.org/10.1016/j.jacceco.2008.02.003.

[5] ZHENG X L, MA H J, WU J. Research on Capabilities Based on the Supply Chain Perspective: Review and Prospects. Foreign Economics & Management, 2018, 40(7): 59-72. https://doi.org/10.16538/j.cnki.fem.2018.07.005. (In Chinese).

[6] CHEN I J, PAULRAJ A. Towards a theory of supply chain management: the constructs and measurements. Journal of Operations Management, 2004, 22(2): 119-150. https://doi.org/10.1016/j.jom.2003.12.007.

[7] LI S, RAO S S, RAGU-NATHAN T S, et al. Development and validation of a measurement instrument for studying supply chain management practices. Journal of Operations Management, 2005, 23(6): 618-

 Copyright @ STEMM Institute Press

641. https://doi.org/10.1016/j.jom.2005.01.002.

[8] LOUGHRAN T, MCDONALD B. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. The Journal of Finance, 2011, 66(1): 35-65. https://doi.org/10.1111/j.1540-6261.2010.01625.x.

[9] GRIMMER J, STEWART B M. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis, 2013, 21(3): 267-297. https://doi.org/10.1093/pan/mps028.

[10] CHEN Y Y, CHEN M L. Qualitative Textual Information and Credit Rating: A Study Based on Risk Information in the Annual Report Text Content. Chinese Journal of Management Science, 2023, 31(9): 94-104. https://doi.org/10.16381/j.cnki.issn1003-207x.2021.0833. (In Chinese).

[11] FLYNN B B, HUO B, ZHAO X. The impact of supply chain integration on performance: A contingency and configuration approach. Journal of Operations Management, 2010, 28(1): 58-71. https://doi.org/10.1016/j.jom.2009.06.001.

[12] PAGELL M. Understanding the factors that enable and inhibit the integration of operations, purchasing and logistics. Journal of Operations Management, 2004, 22(5): 459-487. https://doi.org/10.1016/j.jom.2004.05.008.

[13] TANG C S. Perspectives in supply chain risk management. International Journal of Production Economics, 2006, 103(2): 451-488. https://doi.org/10.1016/j.ijpe.2005.12.006.

[14] LOUGHRAN T, MCDONALD B. Textual analysis in accounting and finance: A survey. Journal of Accounting Research, 2016, 54(4): 1187-1230. https://doi.org/10.1111/1475-679X.12123.

[15] LIU C L, SHI R. Do the stock exchange comment letters affect management tone? Evidence from textual analysis of listed companies' annual reports. Journal of Management Sciences in China, 2024(11): 119-135. https://jmsc.tju.edu.cn/jmsc/article/abstract/20241108. (In Chinese).

[16] WU D. Text-Based Measure of Supply Chain Risk Exposure. Management Science, 2024, 70(7): 4781-4801. https://doi.org/10.1287/mnsc.2023.4927.

[17] TOMLIN B. On the value of mitigation and contingency strategies for managing supply chain disruption risks. Management Science, 2006, 52(5): 639-657. https://doi.org/10.1287/mnsc.1060.0515.

[18] HURLBERT S H. The nonconcept of species diversity: A critique and alternative parameters. Ecology, 1971, 52(4): 577-586. https://doi.org/10.2307/1934145.

[19] SAITO T, REHMSMEIER M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 2015, 10(3): e0118432. https://doi.org/10.1371/journal.pone.0118432.

[20] ARTSTEIN R, POESIO M. Inter-coder agreement for computational linguistics. Computational Linguistics, 2008, 34(4): 555-596. https://doi.org/10.1162/coli.07-034-R2.

Copyright @ STEMM Institute Press 