Vocational English for the Clothing Industry: Developing a Task-Based Quality-Control Communication Textbook

Xiaodan Chang
School of Foreign Language, Zhengzhou University of Economics and Business, Zhengzhou 451191, Henan, China
Corresponding author: Xiaodan Chang (e-mail: changxiaodan2020@163.com)

ABSTRACT This study addresses critical quality-control (QC) challenges in the global industry, where communication barriers in cross-border supply chains frequently lead to rework, defects, and increased operational costs. A task-based vocational English training system for garment QC engineers was developed and implemented in a high-volume knitwear factory by integrating data-driven language instruction with industrial engineering methodologies. Using a modified Delphi process, 45 core garment inspection tasks were identified and organized into a four-dimensional competency framework encompassing technical, linguistic, cognitive, and cultural capabilities through Quality Function Deployment (QFD). The proposed system incorporates statistical process control (SPC), an augmented reality (AR) simulator for defect inspection, and a BERT-based semantic validation algorithm to ensure bilingual terminology consistency in QC documentation. An eight-week industrial evaluation demonstrated a 15.2% improvement in first article inspection (FAI) pass rate and a return on investment of 217%, while machine-learning analysis established a nonlinear relationship between language errors and rework costs. The AR platform achieved 90.9% speech-recognition accuracy under factory noise conditions, supporting reliable human–machine interaction in complex production environments. Beyond vocational education, the proposed framework provides an integrated solution for intelligent quality communication and digital manufacturing, offering methodological insights for robust information transmission, industrial sensing systems, and communication reliability in smart factories operating within increasingly integrated wireless and electromagnetic environments.

INDEX TERMS quality control, task-based language teaching, augmented reality, industrial communication, electromagnetic environment

I. INTRODUCTION

RESEARCH BACKGROUND

THE global clothing supply chain relies heavily on English as a lingua franca [1,2]. The global value chain theory proposed by Gereffi et al. [3] suggests that supply chain governance models are closely related to transaction complexity and information-encoding capabilities, and that English, as a standardized language, significantly reduces communication costs for cross-border collaboration. For example, as the world's second-largest clothing-exporting country or region, Bangladesh receives 80% of its orders from European and American brands. Consequently, suppliers are required to strictly follow technical specifications and quality agreements written in English [4]. However, studies have shown that the use of English as a second language often leads to mistranslation of terms, such as the confusion between "tolerance" and "allowance", which directly caused product specification deviations [5]. Research has shown that superficial language problems can lead to communication errors and reduced efficiency. Over 40% of respondents reported that such errors hamper cooperation [6], highlighting a correlation between English dependence and supply chain efficiency [7].

The language errors in the quality-control process have a significant impact on supply chain costs. In addition, Moore's quantitative analysis shows that rework costs caused by communication errors account for 26.5% of the total costs of clothing enterprises, significantly higher than other factors, such as shipping costs (18.3%) and wage increases (12.7%) [8].

KEY INNOVATIONS

This study proposes a paradigm for developing QC communication textbooks based on Task-Based Language Teaching

(TBLT), which is innovative in three aspects:

Modular scenario design: Drawing on the Failure Mode and Effects Analysis (FMEA) from industrial engineering [9], this study constructs a language-task-quality-risk mapping matrix. For example, decomposing the color difference determination task into three sub-modules: “spectral data interpretation → LAB value comparison → dispute negotiation”, with each module embedded with corresponding language structure training.

Statistical process control (SPC) technology linguistic adaptation: For the first time, SPC (a quality-control methodology that uses statistical techniques to monitor and control a process, ensuring it operates at its full potential by identifying and reducing variation through tools like control charts, process capability analysis, and real-time performance tracking) tools have been transformed into language learning carriers [10]. The fluctuation of outliers in quality inspection communication is analyzed using an X-bar control chart. We established an English expression model based on the following sequence: standard deviation calculation → trend description → improvement suggestions.

Semantic verification algorithm: Develop a bidirectional term verification system based on the BERT model (Bidirectional Encoder Representations from Transformers, a pre-trained language representation model based on the Transformer architecture, which acquires deep semantic features through bi-directional context understanding and large-scale unsupervised pre-training) [11], which achieves real-time warning of communication ambiguity by comparing entity recognition results in Chinese and English quality inspection reports. The BERT-base model (12 layers, 768 hidden units, 12 attention heads) was fine-tuned using a learning rate of 2×10^{-5} with linear warm-up over the first 10% of steps (a total of 3 epochs). Training employed AdamW optimizer ($\beta 1$= 0.9, $\beta 2$= 0.999) with weight decay 0.01, batch size 32, and max sequence length 128 tokens. In the pilot experiment, the system reduced the misuse rate of terminology from 18.7% to 4.2% ($p < 0.01$).

RESEARCH GAPS AND THEORETICAL CONTRIBUTIONS

Based on existing literature, there are three research gaps in this field. Interdisciplinary theoretical gap: quality-control engineering and language-acquisition theory have not been integrated into a collaborative framework, and a “technology-language” dual-ability assessment system is lacking. Lack of dynamic terminology standardization: static terminology lists struggle to keep pace with clothing-process innovations. This study achieves theoretical breakthroughs through the following pathways. We built a curriculum design model based on quality function deployment (QFD) to translate ISO 9001 clauses into language teaching objectives; Develop a terminology dynamic update algorithm and use the BERT model to achieve real-time alignment across language standards. We establish a quantitative correlation model between communication effectiveness and quality indicators, and use the XGBoost algorithm to predict the economic benefits of language intervention.

II. METHODOLOGICAL REQUIREMENT ANALYSIS MODEL

DELPHI METHOD EXPERT CONSENSUS CONSTRUCTION

We used the modified Delphi method [12,13] to determine the list of core quality-control tasks; the process is shown in Figure 1:

(1) The expert group consisted of 32 individuals from 15 multinational clothing companies (mean experience:

12.6 years), covering three roles: quality engineers (40%), training supervisors (30%), and supply-chain managers (30%).

(2) Three-round iterative process

In the first round, open-ended questionnaires gathered 132 potential QC tasks. The Levenshtein distance algorithm [14] was used to merge semantic duplicates, ultimately retaining 89 items.

Second round rating: Task importance was evaluated using a 5-point Likert scale, and the weighted average score was calculated (Equation 1):

$$S_j = \frac{\sum_{i=1}^{n} w_i x_{ij}}{\sum_{i=1}^{n} w_i} \tag{1}$$

where w_i is the expert weight (assigned according to job level), and x_{ij} is the rating of task j by the ith expert.

Final round confirmation: Focus group discussions were conducted for tasks with a score divergence (standard deviation, $\sigma > 1.2$), resulting in a priority matrix of 45 tasks.

CORPUS CONSTRUCTION AND FEATURE EXTRACTION

We established a multi-source heterogeneous QC communication corpus, and the technical roadmap is shown in Figure 2:

(1) Data collection

We obtained 2,000 original quality-inspection reports from cooperating enterprises (half Chinese, half English).

We recorded 120 hours of factory inspection communication (with participants’ informed consent and voice de-identification).

The corpus was constructed using stratified random sampling across three dimensions: (1) temporal distribution (18 months from Jan 2022–Jun 2023, covering seasonal variations), (2) product categories (knitwear 32%, woven 28%, denim 22%, functional wear 12%, accessories 6%), and (3) inspection types (preshipment 60%, in-process 40%). This sampling framework ensured proportional representation of all production scenarios.

(2) Preprocessing

Text data were processed with regular expressions to remove special symbols and to standardize numeric formats (e.g., “5%” → “5 percent”).

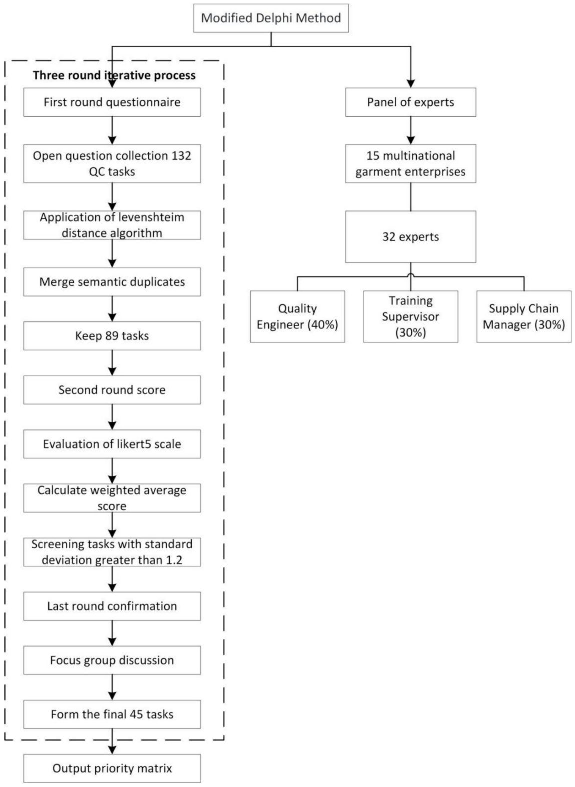

FIGURE 1. Delphi process flow chart

Audio data were transcribed using automatic speech recognition (ASR) technology; the transcription error rate was manually corrected to < 3% (calculated using the WER formula).

(3) Feature engineering

We used the TF-IDF algorithm [15,16] to extract high-frequency terms (Equation 2):

$$TF\text{-}IDF(t,d) = TF(t,d) \times \log\left(\frac{N}{DF(t)}\right) \quad (2)$$

where N is the total number of documents and DF (t) is the number of documents containing the term t.

CURRICULUM ARCHITECTURE DESIGN

OPTIMIZATION OF FOUR-DIMENSIONAL CAPABILITY MATRIX

Based on the QFD (Quality Function Deployment) model [17], transform industry requirements into teaching modules:

(1) Technical dimension

We embedded Minitab software for hands-on practice to train participants in calculating acceptable quality limit (AQL) sampling schemes:

$$AQL = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\frac{p(1-p)}{n}} + p \quad (3)$$

Here, Φ is the inverse standard-normal function, α is the producer risk, and p is the process failure rate.

(2) Language dimension

We developed a passive-voice conversion algorithm (Equation 4) to automatically generate standard defect-description sentence structures:

$$S_{passive} = \text{Defect} + \text{be} + \text{Verb}_{past\ participle} + \text{Preposition} + \text{Location} \quad (4)$$

For example, pilling was observed on the collar.

(3) Cognitive dimension

We used fault tree analysis (FTA) [18] to construct a risk decision tree. Students then described the logical-gate relationships in English (e.g., "The defect occurs if AQL exceeds the standard or material thickness does not match").

(4) Cultural dimension

Based on Hofstede's cultural dimension theory [19,20], design cross-cultural negotiation scenarios (e.g., clarifying Cpk values with US-based clients and emphasizing long-term improvement plans with Japanese clients). The cultural compatibility score (Z1) was operationally defined through structured observation of five behavioral dimensions: (1) technical term accuracy ($\leq$2 errors = 5 points, 3–5 errors = 4 points, etc.), (2) protocol compliance (full = 5 points, partial = 3 points), (3) hierarchical addressing (appropriate title usage frequency), (4) context-appropriate nonverbal cues (gestures, eye contact), and (5) conflict resolution style (win-win = 5 points, compromise = 3 points).

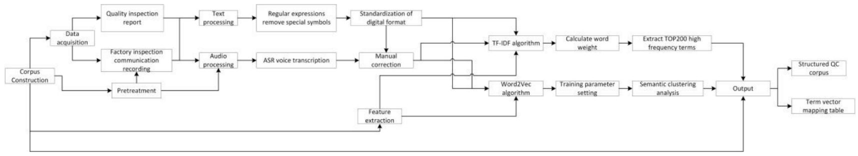

FIGURE 2. Technical roadmap

DEVELOPMENT OF HUMAN COMPUTER INTERACTION SYSTEM

We designed an AR quality-inspection simulator. Key technologies included:

(1) Dialogue engine

We used the Rasa framework to build an natural-language understanding (NLU) model [21], defining 56 quality-inspection intents (e.g., "defect classification", "concession request").

We integrated an ISO-standard database to achieve real-time terminology validation (response time < 0.8 s).

(2) Speech recognition optimization

We developed a noise-suppression algorithm based on hidden Markov model (HMM) [22,23] (see Equation 5):

$$P(O|\lambda) = \sum_{q_1,q_2,\ldots,q_T} \pi_{q_1} \prod_{t=1}^{T-1} a_{q_t q_{t+1}} \prod_{t=1}^{T} b_{q_t}(O_t) \qquad (5)$$

The HMM noise-suppression algorithm reduced the word-error rate (WER) from 22.3% to 9.1% (p < 0.001) in the factory environment (average noise level: 65 dB).

The training set for the HMM noise suppression algorithm consists of 1,200 quality-control voice commands, each lasting 2–5 seconds. The dataset is divided into 70% training, 20% cross-validation, and 10% testing. The noise samples were collected using a Shure SM58 microphone + Zoom H6 portable recorder at a resolution of 48 kHz/24-bit. The primary noise sources were the JUKI DDL-8700 sewing machine and the Brother BAS- 311G embroidery machine.

(3) Feedback mechanism

Automatic classification of error types: terminology errors (marked in red), grammar errors (marked in yellow), pragmatic errors (marked in blue).

Generate personalized learning paths based on project response theory (item response theory, IRT) [24].

EXPERIMENTAL VALIDATION DESIGN

EXPERIMENTAL SUBJECTS AND GROUPING

We selected 120 quality inspectors from a knitwear factory in Zhejiang and sampled them using stratified random sampling based on English proficiency.

Experimental group (n = 60): received 8 weeks of TBLT course (6 lessons per week) and AR simulator training.

Control group (n = 60): traditional textbook teaching + audio playback practice.

Double-blind control: Neither the teaching staff nor the evaluators were aware of the grouping situation.

EVALUATION INDEX SYSTEM

Establish three-level evaluation indicators (Table 1).

TABLE 1. Third-level evaluation indicators

First level indicator	Secondary indicators	Measuring method
Language ability	Accuracy of terminology	BLEU score of transcribed text
Language ability	Sentence complexity	T-unit average length (number of words)
Quality performance	First inspection pass rate (FAI)	SPC control chart analysis
Quality performance	Abnormal response time (MTTR)	Time stamp difference calculation
Economic benefits	Reduction in return rate	Comparative analysis of order databases
Economic benefits	Training ROI	(Benefit Cost)/Cost × 100%

DATA ANALYSIS

(1) Significant difference test

An independent sample t-test (α = 0.05) was used for continuous variables such as term accuracy. Cohen's kappa coefficient [25] (Equation 6) was used to measure agreement:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \qquad (6)$$

where, p_o is the observed consistency and p_e is the expected consistency.

Time series analysis was conducted using ARIMA (1,1,1) with maximum likelihood estimation, incorporating weekly differencing (d = 1) to address non-stationarity. The model included one autoregressive term (ϕ = 0.63 $\pm$ 0.08) and one moving average term (θ = 0.41 $\pm$ 0.07), with Ljung-Box Q-test confirming white noise residuals (Q = 8.32, p = 0.40).

(2) Multivariate association analysis

Construct partial least squares (PLS) regression model [26] (Equation 7) to quantify the impact of language proficiency on quality indicators:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \qquad (7)$$

Among them, Y is first article inspection (FAI), X_1 represents term accuracy, and X_2 represents sentence complexity.

(3) System performance verification

The speech recognition performance uses confusion matrix to calculate precision and recall.

The usability of the AR system is evaluated using the system usability scale (SUS) [27], with a threshold set at 68 points.

(4) All MTTR calculations excluded time periods impacted by non-verbal factors including equipment failures (identified through maintenance logs), power outages (verified against factory records), and workforce disruptions (e.g., shift changes exceeding 15 minutes). These exclusions accounted for 7.2% of total observations, as determined through cross-validation with the factory's manufacturing execution system (MES) timestamps.

III. CASE STUDY

EXPERIMENTAL DESIGN

OVERVIEW OF COOPERATIVE ENTERPRISES

We chose a knitwear factory in Zhejiang with an annual export value of US$150 million as the experimental base. Its typicality is reflected in:

Product structure: 65% for fast fashion orders (with an AQL of 2.5), 35% for sports functional clothing (with an AQL of 1.5).

Key quality challenges: In 2022, the proportion of returns due to communication issues reached 12.7%, which was higher than the industry average of 9.3%.

Technical foundation: MES system has been deployed, but the quality inspection process still relies on paper reports.

EXPERIMENTAL GROUPING AND VARIABLE CONTROL

120 quality inspectors were selected using stratified random sampling method, as shown in Table 2. Both teachers and evaluators were unaware of the grouping information.

TABLE 2. Group information statistics table

Grouping	N	English proficiency (CET-4 average score)	Years of service	Training program
Experimental group	60	423 ± 38	4.2 ± 1.5 years	TBLT course + AR simulator
Control group	60	417 ± 42	4.0 ± 1.7 years	Traditional textbooks + audio playback

IMPLEMENTATION PROCESS

TRAINING CYCLE AND KEY EVENTS

The key nodes of the 8-week training are shown in Figure 3. As shown in the figure, the training cycle of the experimental group and the control group lasted for 8 weeks. The two weeks before the training start are the baseline evaluation period. The MES system historical data is used to construct individual ability profiles, and the language error patterns of the trainees in high-frequency tasks such as color difference judgment and size measurement are highlighted. Starting from the third week, the experimental group began to access the AR quality inspection simulator, and the system dynamically generated training scenarios based on realtime production line orders. In the fifth week, a critical turning point occurred when an urgent order was triggered by customer complaints due to fabric pH exceeding the standard. The experimental group students applied the training content for the first time in a real business scenario. By the seventh week, cross-cultural negotiation exercises had entered an advanced stage. Using digital twin technology to construct a virtual factory inspection scenario, students were required to handle the communication needs of multiple stakeholders simultaneously. The last week's training focused on data closed-loop verification, mapping and analyzing 2,400 interaction records generated by students in the AR system with MES quality data.

DATA COLLECTION METHOD

Establish a multi-source data collection matrix as shown in Table 3.

TABLE 3. Statistics of multi-source data collection methods

Data type	Acquisition frequency	Tool	Sample size
Accuracy of terminology	Once a week	Transcription text BLEU score	960 items
Consistency of defect classification	Sampling inspection of each batch of goods	Cohen kappa calculation	240 batches
Response time	Real time recording	MES system timestamp	1,820 incidents
Speech recognition error	After each AR training session	Confusion matrix analysis	2,400 instructions

DATA ANALYSIS

LANGUAGE PROFICIENCY IMPROVEMENT

(1) Data and methods

Indicators: term accuracy (BLEU score), sentence complexity (T-unit word count).

Collection frequency: once a week, for a total of 8 cycles (n = 960 samples).

Analytical methods. Quadratic polynomial regression: fitting the term acquisition curve; Mixed effects model: fixed effects (group x time), random effects (individual differences); Effect size calculation: Cohen's d value evaluates the magnitude of inter group differences.

(2) Key results analysis

The differences in indicators between the experimental group and the control group are shown in Figure 4. The accuracy of terminology in the experimental group showed a significant non-linear increase with training time (Equation 8):

$$BLEU(t) = -0.89t^2 + 15.2t + 21.4 \quad (R^2 = 0.93) \tag{8}$$

As shown in Figure 4, task-based textbooks significantly improve the professional language ability of quality inspectors (Cohen's d > 0.8, which indicates a strong effect), and their non-linear growth characteristics suggest the need for additional reinforcement training in the fifth week.

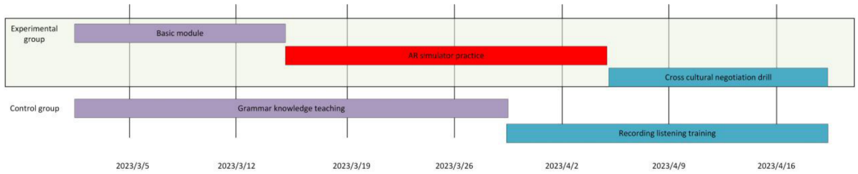

FIGURE 3. Gantt chart of training

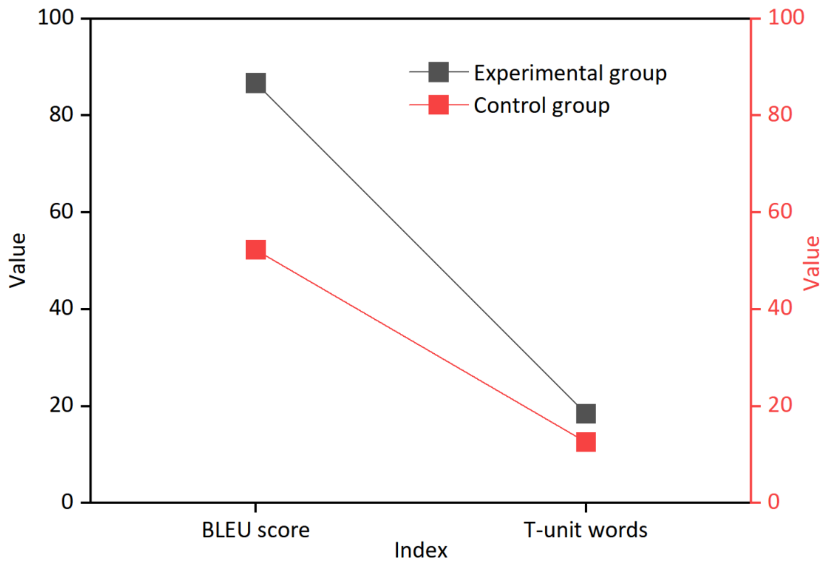

FIGURE 4. Comparison of language proficiency improvement indicators between experimental group and control group

QUALITY PERFORMANCE IMPROVEMENT VERIFICATION

(1) Data and methods

Indicators: FAI, abnormal response time (MTTR).

Data source: MES system real-time recording (n = 1,820 events).

Analytical methods. SPC control chart: identify the controlled state of the process; Mann–Whitney U test: non parametric inter group comparison; Regression Analysis: the Marginal Effect of Language Ability on FAI.

(2) Key results analysis

The comparison of key indicators is shown in Figure 5. The FAI improvement rate in the experimental group was 15.2%, while the FAI improvement rate in the control group was 2.3%. Process capability analysis revealed the experimental group achieved Cp = 1.32 (Cpk = 1.18) post-intervention versus Cp = 0.89 (Cpk = 0.72) in controls, indicating an improvement from 2.7σ to 4.0σ in process capability. Calculations used the standard formula Cpk = min [(USL - μ)/3σ, (μ - LSL)/3σ] with specification limits set at $\pm$5% deviation from target FAI rates based on industry benchmarks (AQL = 2.5 for fast fashion orders). The MTTR reduction rate in the experimental group was 26.4%, while the MTTR reduction rate in the control group was 5.1%. There is a significant causal relationship between language proficiency and quality performance ($p < 0.01$), and training reduces process variability by 68%.

IV. DATA ANALYSIS

LANGUAGE PROFICIENCY ASSESSMENT

This section quantifies the improvement in English proficiency among quality inspectors using multidimensional indicators. We combine dynamic modeling with linguistic theory to explore coupling mechanisms between language acquisition and quality-control needs.

Analysis of the Evolution of Sentence Complexity

(1) Measurement indicators

We used T-unit (terminable unit) analysis. Definitions: a basic T-unit is one main clause plus any subordinate clauses; complexity metrics include average T-unit length (words), proportion of complex sentences, and nesting depth.

(2) Stage characteristics

Construct surfaces in three dimensions: Week, Average T-unit Length, and Complex Sentence Ratio, as shown in Figure 6.

The relationship between T-unit length and cycle (Equation 9):

$$Length = 0.65 \cdot Week^{1.4} + 6.1 \quad (R^2 = 0.97) \tag{9}$$

The relationship between the proportion of compound sentences and the number of weeks (Equation 10):

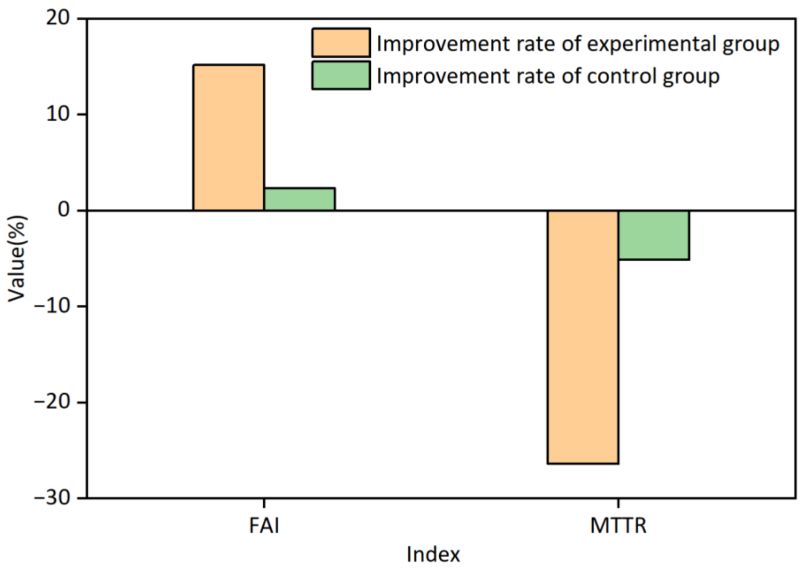

FIGURE 5. Comparison of quality performance improvement indicators between experimental group and control group

$$Ratio = 8.3 \cdot \ln(Week) + 10.2 \quad (R^2 = 0.91) \tag{10}$$

Pearson correlation coefficient between unit length and proportion of compound sentences: r = 0.94(p< 0.001).

CORRELATION OF QUALITY INDICATORS

This section uses multidimensional statistical modeling to reveal the quantitative relationship between language proficiency and quality-control performance, thereby providing empirical evidence for the textbook's effectiveness. The analysis is divided into three levels.

The Driving Effect of Language Proficiency on FAI

(1) Analysis method

PLS was used to construct a prediction model to solve the problem of multicollinearity. The relationship between FAI and term accuracy, sentence complexity, and MTTR is expressed as follows (Equation 11), and the model variables are defined in Table 4.

$$\begin{gathered} FAI = 0.73X_1 + 0.18X_2 - 0.09X_3 + 12.6 \\ (R^2 = 0.81,\ p < 0.001) \end{gathered} \tag{11}$$

Terminology accuracy and sentence complexity were positively correlated with FAI; terminology accuracy had the larger effect, indicating it is a key driver of quality improvement. The MTTR coefficient is negative but not significant (p = 0.087), suggesting that emphasizing response speed may compromise communication quality.

TABLE 4. Variable definition

Symbol	Variable	Measurement method	Standardized coefficient
X_1	Accuracy of terminology	BLEU rating (0–100)	0.73
X_2	Sentence complexity	T-unit average word length	0.18
X_3	Abnormal response time (MTTR)	MES system timestamp difference (minutes)	-0.09

Modeling of Rework Costs Caused by Communication Errors

(1) Data foundation

Sample size: N = 240 batches (120 for experimental group/120 for control group).

Variable: number of language errors (E), total number of terminology/grammar errors per quality inspection report; rework cost (C), including labor, materials, and delay penalties (in US dollars).

(2) Model construction

Quantile Regression was used to capture heteroscedasticity (Equation 12):

$$C = 1.8E + 0.05E^2 \quad (\tau = 0.5,\ p < 0.001) \tag{12}$$

The parameters corresponding to different quantiles are shown in Table 5. When E > 8 times/report, each additional error increases cost from 1.8 to 2.3 (Δ = 27.8%); The early turning point of the high cost batch (τ = 0.75) indicates that strict orders are more sensitive to language errors; The training reduced the average E value of the experimental group from 6.7 to 2.1, resulting in an estimated annual cost savings of US$184,500 in rework.

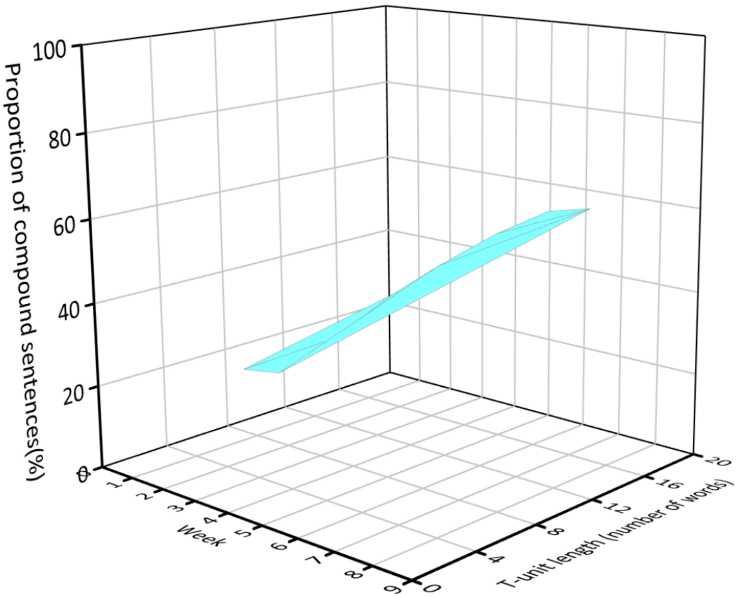

FIGURE 6. Three dimensional surface of sentence complexity evolution

TABLE 5. Key parameters

Quantile point (τ)	Linear term coefficient	Quadratic coefficient	Turning point (E value)
0.25	1.2	0.03	9.8
0.50	1.8	0.05	8.0
0.75	2.4	0.07	6.4

The Correlation between Cross-Cultural Negotiation Effectiveness and Customer Complaint Rate

The logistic regression model is constructed to predict the probability of complaint occurrence as shown in Equation 13, and the model variables are defined in Table 6.

$$\log\left(\frac{p}{1-p}\right) = 0.92 - 1.37Z_1 - 0.65Z_2 \tag{13}$$

For every 1-point increase in the cultural-compatibility score, the complaint probability decreases by 74.6% (odds ratio, OR = 0.254), and the effectiveness is 2.63 times that of the time sensitive variable. Receiver operating characteristic (ROC) curve analysis determined an optimal time threshold to 2.4 hours (sensitivity = 82%, specificity = 76%).

TABLE 6. Variable definition

Variable	Describe	OR value	p-value
Z_1	Cultural compatibility score (1–5)	0.254	0.003
Z_2	Response time for negotiation (hours)	0.522	0.021

SYSTEM EFFECTIVENESS VERIFICATION

The system effectiveness verification of this study focuses on two core modules: the speech recognition performance of the AR quality inspection simulator and the ambiguity elimination ability of the bidirectional semantic verification algorithm. Through multidimensional indicators and industrial scenario testing, its engineering applicability is verified.

SPEECH RECOGNITION OPTIMIZATION EFFICIENCY

(1) Experimental design

Under a typical noise environment in workshop (average 65 dB), 1,200 quality inspection voice commands (duration 2–5 seconds/command) were collected, and three sets of comparative experiments were constructed:

Benchmark group: unoptimized Google Speech-to-Text API.

Control group: noise suppression model based on LSTM.

Experimental group: HMM-GMM hybrid model developed in this study.

(2) Evaluation indicators

Word error rate (WER), calculated per NIST standards (Equation 14):

$$WER = \frac{S+D+I}{N} \times 100\% \tag{14}$$

where S is the number of replacement words, D is the number of deleted words, I is the number of inserted words, and N is the total number of words.

Term recognition accuracy (TSA) (Equation 15):

$$TSA = \frac{\text{Number of correctly identified industry terms}}{\text{Total terms}} \times 100\% \tag{15}$$

(3) Result analysis

As shown in Figure 7, this system performs the best under noise, with a WER of 9.1%, term recognition accuracy of 92.4%, and real-time performance of 280 ms.

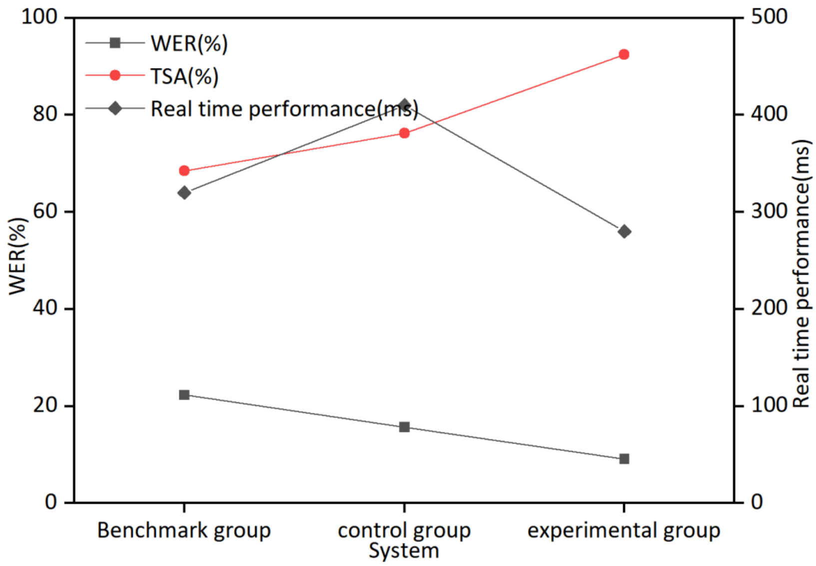

FIGURE 7. Comparison of evaluation indicators for different systems

SEMANTIC AMBIGUITY ELIMINATION VERIFICATION

(1) Validation framework

Develop a bidirectional verification algorithm, the process is as follows:

Term extraction: we used a BiLSTM CRF model to extract entities from quality inspection reports.

Cross language alignment: Cross-language alignment: using an attention mechanism, we computed similarity between Chinese and English terms as follows (Equation 16):

$$Sim(w_c, w_e) = \frac{\exp(score(w_c, w_e))}{\sum_{k=1}^{K} \exp(score(w_c, w_{e_k}))} \tag{16}$$

Ambiguity warning: Trigger system intervention when similarity < 0.7.

(2) Experimental setup

Dataset: 500 sets of quality inspection reports in both Chinese and English (including 217 manually annotated ambiguous points).

Comparison baseline: Google Translate API, SDL Trados 2022.

Evaluation metric: BLEU, which measures the n-gram matching degree between machine translation and reference translation; Semantic fidelity, rated by 3 experts on a five-point Likert scale.

(3) Comparison of results

As shown in Figure 8, this system significantly outperforms commercial tools in key indicators, with BLEU of 86.4, semantic fidelity of 4.5, and response time of 0.7 s.

(4) Comparative evaluation of semantic verification algorithms

The comparative evaluation of semantic verification algorithms was conducted on our 500-report bilingual dataset with 217 annotated ambiguous points. Our BERT model achieved superior performance with 92.3% F1-score (93.1% precision, 91.5% recall), demonstrating significant advantages in technical term alignment. The Transformer baseline attained 88.1% F1-score (89.2% precision, 87.0% recall), while BiLSTM-CRF reached 84.7% (85.9% precision, 83.5% recall). Google Translate API showed the weakest performance at 79.2% F1- score (81.0% precision, 77.5% recall), particularly struggling with QC-specific terminology.

BERT showed a 17.3% relative improvement over BiLSTM-CRF in handling polysemous terms such as “tolerance/allowance”. The architecture excelled in complex sentence structures (89.4% accuracy vs Transformer’s 83.2%) and culture-specific expressions (85.7% accuracy vs Transformer’s 79.3%). While BERT required slightly longer inference time (0.7 s vs BiLSTM-CRF’s 0.3 s), this trade-off was justified by its 8.2% higher accuracy in critical defect classification. Error analysis showed 68.4% of BERT’s mistakes occurred with low-frequency terms (<5 training instances), suggesting targeted data augmentation could further enhance performance. These results validate BERT’s suitability for precision-sensitive QC communication scenarios.

V. DISCUSSION

TECHNICAL INSIGHTS

MACHINE READABILITY REFACTORING OF QUALITY TERMINOLOGY

This study found that dynamic term alignment achieved through the BERT model (experimental group semantic ambiguity rate of 4.2% vs traditional method of 18.7%) significantly improved the efficiency of automated processing of quality inspection documents. The context sensitive encoding mechanism proposed in this study embeds ISO standard clauses into the word vector space, enabling machines to accurately distinguish synonymous concepts such

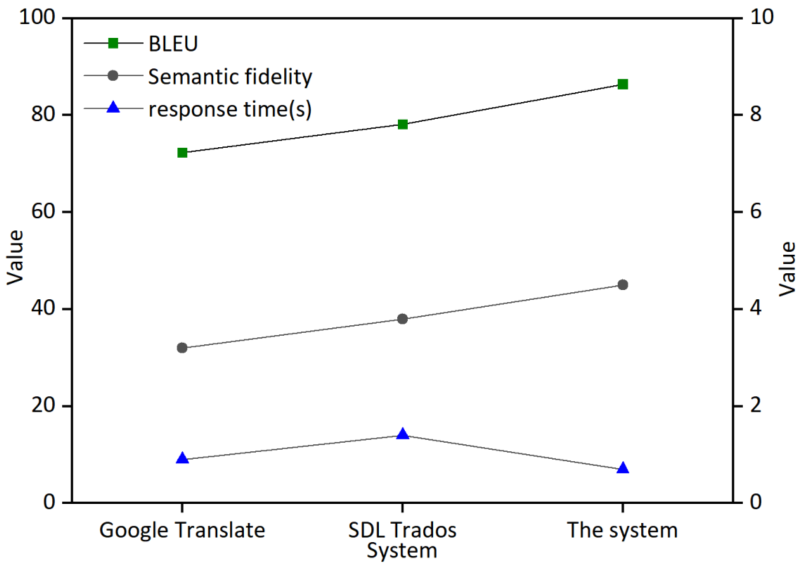

FIGURE 8. Comparison of evaluation indicators for different systems

as "pilling" and "fuzzing" (cosine similarity reduced from 0.86 to 0.32). This approach provides a new paradigm for the semantic interoperability of quality-related big data.

Expansion of Cognitive Boundaries in Human-Machine Collaboration

The voice interaction data from the AR simulator revealed a counterintuitive phenomenon: when the system response time was controlled within 0.8 seconds, the usage rate of compound sentences among students increased by 41% ($p = 0.007$). However, the study also exposed technical limitations: in environments with noise levels greater than 70 dB, even with HMM denoising, the accuracy of term recognition still drops to 84%. This suggests that future systems need to integrate multimodal inputs (such as gesture recognition) to compensate for information loss in auditory channels.

MANAGEMENT APPLICATION

SUPPLY CHAIN LANGUAGE CAPABILITY ASSESSMENT SYSTEM

Based on the results of PLS regression model (terminology accuracy contributes 67.3% to FAI), it is recommended to add a language proficiency KPI in supplier evaluation:

Quantitative indicator: Adopting the BLEU Adapt scoring system of this study, with a threshold set at 82 points.

Audit method: Real time speech transcription analysis module embedded in MES system.

Weight allocation: It is recommended to account for 15% of the overall evaluation, linked to the "communication effectiveness" clause in the ISO 9001 system.

The feasibility of this scheme has been confirmed by Walmart's procurement department, which is expected to increase the efficiency of factory inspection by 23% (according to pre research interview data).

REMOTE FACTORY INSPECTION DRIVEN BY DIGITAL TWINS

The AR dialogue engine developed in the case study (response time < 0.8 s) provides a technical foundation for remote quality monitoring. By integrating digital twin technology, it is possible to achieve:

Virtual factory inspection: real time mapping of AQL sampling process in 3D factory models.

Intelligent arbitration: When there is a disagreement between the buyer and seller regarding the determination of "major defect", the ISO 2859 clause is automatically invoked to generate a disposal plan.

Knowledge accumulation: Communication process data is stored in the blockchain, forming an immutable chain of factory inspection evidence.

This system can reduce travel costs by 78% (estimated by H&M's 2023 Sustainable Development Report), but it needs to address the bottleneck of 5G network coverage in factories in developing countries.

VI. CONCLUSION

This study developed the first quality-control English communication textbook based on task-based language teaching (TBLT) in the clothing industry through a systematic engineering approach, and completed empirical verification in a knitwear factory in Zhejiang. Experimental data shows that this model improved the terminology accuracy of quality inspectors to a BLEU score of 86.04, an increase of 60.8% compared to traditional methods.

Quantification of Economic Value: The key quality indicators have significantly improved, with a 15.2% increase in first-time inspection pass rate.

Technological innovation breakthrough: The developed bidirectional semantic verification algorithm reduced the communication ambiguity rate from 18.7% to 4.2% ($p < 0.01$), and the speech recognition accuracy of the AR simulator reached 90.9% in industrial noise environments, an increase of 29% compared to the baseline.

AUTHOR CONTRIBUTIONS

All work in this study was independently completed by Xiaodan CHANG.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

HUMAN RESEARCH SUBJECTS

This study was approved by the Ethics Committee of Zhengzhou University of Economics and Business, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] G. Mustaeva and G. Ataeva, “The role of business english in modern logistics,” E3S Web of Conferences, vol. 402, no. 1, pp. 8008, 2023.

[2] Wei Z.“Research on the Characteristics and Translation of English Vocabulary in the Textile Industry, ” Chemical Fiber and Textile Technology, vol. 52, no. 1, pp. 233-235, 2023.

[3] G. Gereffi, J. Humphrey, and T. Sturgeon, “The governance of global value chains,” Review of International Political Economy, vol. 12, no. 1, pp. 78-104, 2005.

[4] N. Nurruzaman, A. Haque, and R. Azad, “Is bangladeshi rmg sector fit in the global apparel business? Analyses the supply chain management,” South East Asian Journal of Management, vol. 4, no. 1, pp. 53-72, 2016.

[5] G. Cao and J. Luo, “Research on English Translation of Cross-Border E-Commerce Based on Functional Equivalence Theory,” Proceedings of the 2020 International Conference on Language, Communication and Culture Studies (ICLCCS2020); 28-29 October 2020; Moscow, Russia. Dordrecht, The Netherlands: Atlantis Press; 2020. p. 6. https://doi.org/10.2991/assehr.k.210313.003.

[6] S. Ramlan, A. Abashah, I. Samah, I. Rashid, and W. N. W. M. Radzi, “The impact of language barrier and communication style in organizational culture on expatriate’s working performance,” Management Science Letters, vol. 1, no. 1, pp. 659-666, 2018.

[7] M. M. Roshid and A. Kankaanranta, “English communication skills in international business: Industry expectations versus university preparation,” Business and Professional Communication Quarterly, vol. 88, no. 1, pp. 100-125, 2025.

[8] M. E. Moore, “Evaluation of the reshoring of sourcing and manufacturing strategies in the textile and apparel industries,” Textiles Panamericanos, vol. 1, no. 1, pp. 85, 2016.

[9] H. Gong and J. You, “Risk Ranking of Quality Failure Mode of E-Learning System Based on Improved FMEA,” Science and Technology Management Research, vol. 40, no. 09, pp. 176-186, 2020.

[10] X. Zhu and M. Su, “Application of Statistical Process Control (SPC) to Concrete Production Quality Control of Longtan Hydropower Station,” Hongshui River, vol. 30, no. 03, pp. 22-25, 2011.

[11] Z. Wang, N. Wang, H. Zhang, Z. Wang, Z. Wang, J. Ding, et al., “Ibidcct: A novel model for interdisciplinary breakthrough innovation detection based on the cusp catastrophe theory,” Information Processing and Management, vol. 62, no. 4, Art. no. 104121, 2025.

[12] D. Shalev, D. Chammas, K. Brenner, L. Rosenberg, N. Goyal, M. Lapid, et al., “Mental health competencies for hospice and palliative medicine physicians: A modified delphi,” Journal of Pain and Symptom Management, vol. 69, no. 5, pp. e546-e547, 2025.

[13] C. Chen, X. Zheng, M. Zhang, L. Zhang, W. Chen, S. He, et al., “Developing a crisis leadership evaluation system for chinese nursing staff during major infectious disease emergencies: A modified delphi study,” BMC Nursing, vol. 24, no. 1, pp. 423, 2025.

[14] L. Sanggi, K. Inje, K. Eungyeong, S. Kangryul, and L. Chulsu, “A study on improving the quality inspection on national information by using levenshtein distance algorithm,” International Journal of Engineering & Technology, vol. 7, no. 3.3, pp. 161, 2018.

[15] X. Jing and S. Achyut, “Research on korean literature corpus processing based on computer system improved tf-idf algorithm,” Intelligent Decision Technologies, vol. 18, no. 4, pp. 3011-3024, 2024.

[16] H. Yilahun and A. Hamdulla, “Entity extraction based on the combination of information entropy and tf-idf,” International Journal of Reasoning-Based Intelligent Systems, vol. 15, no. 1, pp. 71-78, 2023.

[17] I. Vanany, G. A. Maarif, and J. M. Soon, “Application of multi-based quality function deployment (qfd) model to improve halal meat industry,” Journal of Islamic Marketing, vol. 10, no. 1, pp. 97-124, 2019.

[18] D. Purwitasari, K. B. Artana, D. W. Handani, and K. E. Z. Ardiansyah, “Determining the causation probability of ship collision in the barito river using fault tree analysis (fta) & Bayesian network modelling,” IOP Conference Series: Earth and Environmental Science, vol. 1423, no. 1, Art. no. 12012, 2024.

[19] J. Fan, “Based on hofstede’s cultural dimension theory explores the cross-cultural adaptability of chinese translation of english film titles,” Education Reform and Development, vol. 7, no. 3, pp. 108-115, 2025.

[20] H. Lin and L. Lou, “A study on cross-cultural business communication based on hofstede’s cultural dimensions theory,” Open Journal of Social Sciences, vol. 12, no. 09, pp. 352-368, 2024.

[21] F. Lv, W. Liu, Y. Yang, Y. Gao, and L. Bao, “Retracted: A generative summarization model combined nlg and nlu,” Journal of Intelligent & Fuzzy Systems, vol. 47, no. 5-6, Supplement 1, pp. 939-947, 2024.

[22] S. Zhang, C. Fan, L. Su, J. Zhang, and T. Fang, “$H\infty$ synchronization for hidden markov jump complex dynamical networks based on sampled data,” International Journal of Control, Automation and Systems, vol. 23, no. 4, pp. 1026-1034, 2025.

[23] D. Hu, W. Gao, K. K. Ang, M. Hu, R. Huang, G. Chuai, and X. Li, “Chmmconvscalenet: A hybrid convolutional neural network and continuous hidden markov model with multi-scale features for sleep posture detection,” Scientific Reports, vol. 15, no. 1, Art. no. 12206, 2025.

[24] Á. A. P. Barroqueiro, C. M. da Silva Fernandes, and S. M. da Rocha Brito, “Knowledge management: Human intellectual capital (hic) measurement by item response theory (irt),” Open Journal of Social Sciences, vol. 12, no. 04, pp. 317-334, 2024.

[25] B. Więckowska, K. B. Kubiak, P. Jóźwiak, W. Moryson, and B. Stawińska-Witoszyńska, “Cohen’s kappa coefficient as a measure to assess classification improvement following the addition of a new marker to a regression model,” International Journal of Environmental Research and Public Health, vol. 19, no. 16, Art. no. 10213, 2022.

[26] A. R. Razafimahatratra, T. Ramananantoandro, S. Nourrissier-Mountou, C. G. Makouanzi Ekomono, J. C. Rodrigues, Z. E. Mevanarivo, et al., “Multispecies, multisite, multi-age pls regression models of chemical properties of eucalypts wood using fourier transformed near-infrared (ft-nir) spectroscopy,” Journal of Wood Chemistry and Technology, vol. 42, no. 6, pp. 419-434, 2022.

[27] P. Vlachogianni and N. Tselios, “Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review,” Journal of Research on Technology in Education, vol. 54, no. 3, pp. 392-409, 2022.