Textile & Leather Review

ISSN 2623-6281 | www.tlr-journal.com | 10.31881/TLR

Realising Multimodal Discourse Integrating Visual and Videos in Vocational English: A Case Study of Textile Dyeing and Finishing

Xiangying Kou

How to cite: Kou X. Realising Multimodal Discourse Integrating Visual and Videos in Vocational English: A Case Study of Textile Dyeing and Finishing. Textile & Leather Review. 2025; 8:741-860. https://doi.org/10.31881/TLR.2026.5033

How to link: https://doi.org/10.31881/TLR.2026.5033

Published:27 April 2026

Realising Multimodal Discourse Integrating Visual and Videos in Vocational English: A Case Study of Textile Dyeing and Finishing

Xiangying KOU
Department of Public Courses, Xi'an Traffic Engineering Institute, Xi'an 710300, Shaanxi, China
pujiaoqi4401473@126.com

Article
https://doi.org/10.31881/TLR.2026.5033
Published 27 April 2026

ABSTRACT

Amid the increasing demand for vocational English proficiency in China's textile industry, particularly in dyeing and finishing sectors, traditional teaching methods fail to address the operational and communicative needs of students. This study constructs a multimodal discourse-based teaching framework integrating visual images and hands-on videos to improve vocational English instruction. Utilising a design-based research paradigm combined with mixed methods, the study first identifies four core industry task scenarios-process debugging, equipment operation, quality control, and accident handling-through Delphi expert consultation. A three-dimensional matrix mapping "linguistic task-visual modality-video elements" was developed to guide multimodal resource construction. A 16-hour controlled experiment involving 86 students demonstrated that the experimental group receiving multimodal instruction significantly outperformed the control group in three dimensions: execution of operational instructions (+24.7 points), terminology reproduction (2.1× increase), and reduced cognitive load across six NASA-TLX indices. Qualitative data revealed mechanisms such as subtitle-gesture anchoring and visual warning guidance that supported terminology internalisation and error avoidance. Based on these findings, a vocational multimodal discourse production model was proposed, alongside a set of teaching resource development standards. This model provides a transferable framework for cross-disciplinary applications in vocational education. Future work should explore low-cost, scalable models for regions with limited access to VR and interactive platforms to expand the accessibility of multimodal English education.

KEYWORDS
multimodal discourse, visual images, hands-on videos, vocational English teaching, textile dyeing and finishing, design-based research, language task migration

INTRODUCTION

With the continuous promotion of the Belt and Road initiative, the status of Chinese textile foreign trade enterprises in the international market is increasing, and the related industries have put forward higher requirements for the foreign language proficiency of frontline skilled workers [1]. Especially in the dyeing and finishing process, an accurate understanding of English technical instructions in the quality control process of foreign trade orders has become a key factor in guaranteeing production safety and efficiency [2]. According to the notification data of the China General Administration of

Customs from 2020 to 2023, there were 37 accidents nationwide caused by the misinterpretation of English in equipment operation. These accidents involved high-precision equipment such as continuous fixing machines, overflow dyeing machines, and circular screen printing machines. Some of the accidents directly led to the return of exported batches or claims for compensation. As a result, enterprises suffered serious economic losses. The teaching effect in the field of vocational education also reflects the bottleneck of the current practical English teaching [3]. China Textile Vocational Education Union pointed out in the industry research report released in 2024 that the passing rate of students in the dyeing and finishing direction in the national textile middle and high vocational colleges and universities in the practical English assessment was only 41.6%, and most of the students were unable to complete the accurate recognition of the foreign language equipment interface, operation process and quality terminology in the real scenario [4,5]. This phenomenon exposes the deep mismatch between the traditional English teaching mode and the needs of the industry, and also suggests that we need to seek a teaching path that is more suitable for the job situation. Currently, vocational English teaching in the field of textile dyeing and finishing faces the structural problem of triple disconnection. The textbook is disconnected from the actual scene of the equipment. Taking the overflow dyeing machine as an example, the textual descriptions in the teaching materials are mostly technical principles and process descriptions, while the operation interface in the actual work is highly dependent on images, button commands, and English prompts, so students lack the input and output experience of the real context, which makes it very easy to produce comprehension bias. The single mode is out of line with the cognitive demand. A large number of links in the dyeing and finishing process, such as weighing and dosing of auxiliaries, process temperature control and frequency adjustment, etc., involve dynamic changes, spatial distribution and judgment of details, and it is difficult to satisfy the students' understanding needs in the whole process of perception-cognition-application by relying only on static text or verbal explanation [6]. The lack of multimodal support, such as images, videos, live audio, etc., directly affects the interactive depth and efficiency of teaching. Classroom language is out of touch with job tasks [7,8]. If there is a lack of visual references or samples to describe the common fabric defects in the actual job, students often have difficulty in matching the English terminology with the fabric representations, which affects the completion of communication tasks such as problem reporting and quality feedback. Based on the above problems, this study takes the multimodal discourse of visual image + practical video as the entry point to explore its realisation path in vocational English teaching. The objectives of the study include: to construct a multimodal teaching resource development model for textile dyeing and finishing majors, and produce a Technical Specification for Resource Development to guide the collaborative construction of teaching materials and platforms in the future; to validate the effect of the multimodal teaching pathway on the improvement of students' vocational English proficiency through empirical research, with the target level of B1+ of the Common European Framework of Reference for Languages (CEFR); and to form a generalizable industry-education teaching methodology that can be applied to the teaching of vocational English. The study also aims to develop a generalizable industry-education collaborative mechanism, which can provide theoretical reference and practical samples for the reform of English teaching in other vocational majors.

RESEARCH OVERVIEW

Against the background of the transformation of vocational English teaching reform to scenario-based and practical teaching, multimodal discourse theory and practice path have gradually become the focus of international research [9]. To ensure that this research proposal has a solid theoretical basis and practical applicability, this paper systematically compiles the existing literature and research bases from three aspects: firstly, the theoretical cornerstone of multimodal discourse; secondly, the evolution of vocational English teaching, especially the English for Specific Purposes (ESP) pathway; thirdly, the empirical exploration of multimodal teaching; and fourthly, the diagnosis of modal status of the existing teaching resources of textile dyeing and finishing majors.

Theoretical basis of multimodal discourse

The three-function framework of visual grammar proposed by some researchers provides a systematic path for the analysis of multimodal resources. The theory starts from the language function and explores the semantic expression mechanism of images at the three levels of reproduction, interaction, and composition, respectively [10]. This framework is highly applicable to the design of multimodal teaching resources for textile dyeing and finishing majors [11,12]. The reproduction function emphasises the matching relationship between image narrativity and technical process [13]. In the teaching of dyeing and finishing, complex processes such as continuous rolling and dyeing, liquid ammonia finishing, etc., often rely on images to convey the timing and movement.

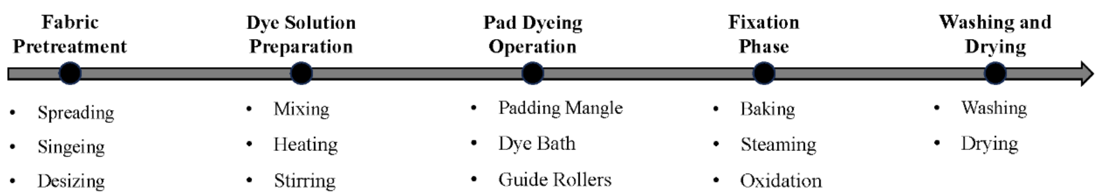

Figure 1. Flow chart of the dyeing process

As Figure 1 shows, the flow chart of the dyeing process, this function helps to construct a cognitive model of the whole technology chain for students. The interactive function pays attention to the relationship between the image and the viewer, especially in the hands-on video, as the camera angle,

https://doi.org/10.31881/TLR.2026.5033

distance, and focus control [14]. For example, the use of first-person operator's point of view filming can effectively enhance the students' sense of experiencing the operation of the equipment control interface, and enhance the relevance of the language input to the task. The compositional function focuses on the constructive effect of graphic arrangement and the distribution of information blocks [15]. Taking the Fabric Defects Record Sheet in the quality inspection report as an example, students' cross-modal reading and information integration skills can be significantly enhanced through the rational arrangement of graphical illustrations and textual descriptions.

Reconstruction of the vocational English teaching framework

Other researchers have proposed the ESP three-dimensional needs analysis model, which emphasises that learning needs, learner characteristics, and the teaching environment should be adapted to each other in the target context [16]. However, in traditional English teaching of textile majors, the application of this model often falls into the dilemma that task design is detached from job details. For this reason, researchers at home and abroad have made localised modifications to the application of the ESP model in the textile industry [17]. For example, when researching the case of technician training in a foreign trade textile factory, it was pointed out that the chain of equipment technology should be matched with the nodes of foreign language communication, to build a teaching unit of task-corpus-resource to realise the synergistic construction of language and skills [18]. Meanwhile, the EU-funded CLIL pilot project "Textile4Europe" in 2021 encountered many difficulties in the process of implementation [19]. According to the research, the main reason for the failure of the program is that it ignores the apprentices' cognitive pathways of dynamic operational processes and terminology recognition, and classroom teaching is still dominated by traditional language without sufficiently introducing modal resources, such as images and videos, to support the understanding of complex content.

The empirical basis of multimodal teaching

In recent years, empirical research on multimodal teaching has mainly focused on the dimensions of language transfer path and task completion performance [20,21]. Consistent with Nguyen et al. [20], the integration of image–text–explanation multimodal instruction within the Mechanical Drawing curriculum yielded a 23.4 % enhancement in both students’ comprehension and their accuracy in retelling technical instructions. After the introduction of image-text-explanation in the course of Mechanical Drawing, the students' accuracy in understanding and repeating technical instructions increased by 23.4% [22]. In addition, a teaching experiment using operation video + structure diagram + task-based question and answer mode was conducted in the Mechanical English course of higher vocational education in China, and the results showed that the students in the experimental group performed better than the control group in the CEFR-B1 speaking tasks, especially in the areas of explaining the order and expressing the differences in positional relationships. The differences are significant in the order of instructions and expression of positional relations [23]. These studies fully demonstrate that multimodal input not only enhances the richness of language reception but also promotes students' language transfer and reconstruction ability in real tasks through the cognitive anchoring mechanism.

The current situation of modal resources for teaching textile dyeing and finishing

In order to further clarify the realistic basis of this study, the author selected six current mainstream textbooks specialising in textile dyeing and finishing for modal resource analysis. Table 1 demonstrates the comparison between image density and video resource types of the six textbooks. It can be seen that there are obvious deficiencies in the ratio of image and video resources in the current textbooks: low modal density of images, an average of only 0.8 images per thousand words, far below the international vocational textbook standard of 3/thousand words, and most of them are static photos of equipment, and there is a lack of operation process diagrams. Weak interactivity of video resources: 92% of the existing videos are one-way demonstration videos recorded by teachers, which lack interactive design with adjustable speed, multiple viewing angles, and subtitle prompts. To summarise, the existing vocational English teaching materials and teaching resources are significantly lacking multimodal integration, particularly in terms of visual–video synergy and contextual task alignment.

Table 1. Comparison of Image Density and Video Resource Types of 6 Textbooks

Textbook Name	Image modal density (Number of images per 1,000 words)	Type of video resources	Description of video resource interactivity
Textbook 1	0.5	One-way presentation-type video	No interactive design, lack of speed adjustment, viewpoint switching, and subtitle prompts.
Textbook 2	1.2	One-way demonstration video	Videos with a single viewpoint, lacking interaction, and captioning.
Textbook 3	0.9	One-way demonstration video	No interactive features, video is mainly used for the instructor's explanation.
Textbook 4	0.6	One-way demo video	Video resources are recorded by the instructor, and the speed and viewing angle cannot be adjusted.
Textbook 5	0.8	One-way demonstration videos	Lack of interactive design, single viewpoint presentation of content

Textbook Name	Image modal density (Number of images per 1,000 words)	Type of video resources	Description of video resource interactivity
Textbook 6	1.1	Unidirectional Demonstration Video	The video is recorded from a fixed viewpoint and lacks speed adjustment and interactive elements.

RESEARCH METHODS

Research paradigm and methodology integration

This study adopts a research paradigm combining design-based research and mixed research methods, aiming to systematically develop multimodal English teaching resources adapted to the vocational context of textile dyeing and finishing, and to empirically validate their teaching effects and applicable mechanisms. As a practical problem-solving-oriented educational research method, design-based research emphasises iterative development, contextual embeddedness, and theoretical construction, which is especially suitable for exploring the paths and refining the mechanisms of complex teaching and learning problems [24,25]. At the same time, the mixed-methods approach provides a complementary technical path between quantitative assessment and qualitative interpretation: through empirical quantitative experiments to validate the effects of teaching interventions, and through multiple sources of qualitative data to reveal the evolution of mechanisms and students' experiences in the process of teaching and learning. The specific research process is divided into three core phases: needs analysis → development of teaching resources → implementation of teaching experiments, supplemented by multidimensional data collection and cross-validation, and Figure 2 shows the flow chart of the overall technical path of the specific research. In the demand analysis phase, based on the Delphi method of expert interviews and industry demand research, the research object and core tasks are determined [26]. In the teaching resources development phase, multimodal teaching resources are designed and developed based on the analysis results. In the teaching experiment implementation stage, a double-blind controlled experimental design is used to verify the teaching interventions and effects [27]. A multidimensional data collection approach is standardised out at each stage, with expert interviews, questionnaires, and industry data analysis in the demand analysis stage; feedback of teaching materials, design evaluation, and expert review in the resource development stage; and quantitative and qualitative data in the implementation stage of the teaching experiment. Feedback is provided through data cross-validation between the stages to ensure the scientific nature of the research path and the effectiveness of repeated iterations.

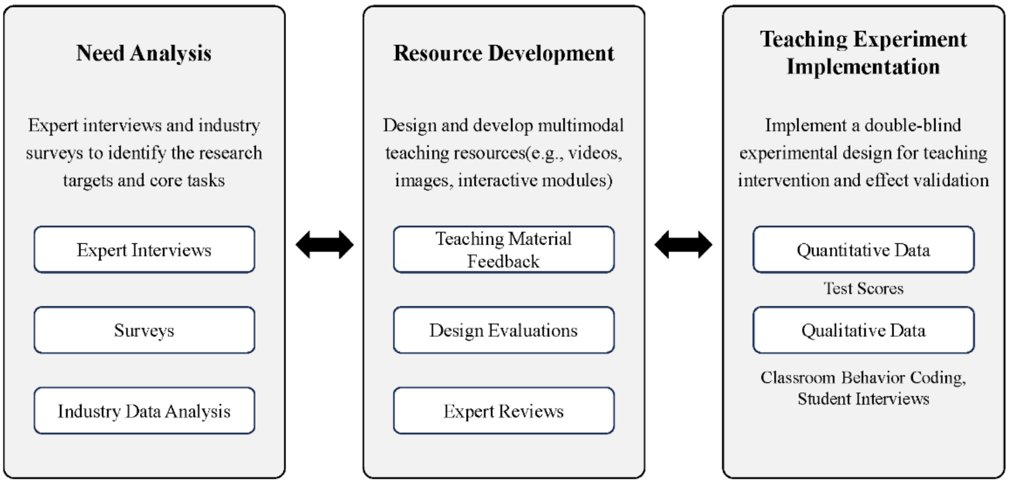

Figure 2. Flow chart of the overall technology path of the study

Research Implementation Flow

Requirement Identification of Key Tasks

The three-round Delphi method is used to identify high-frequency critical task scenarios in industry English teaching [28,29]. The research team invited 12 enterprise experts to form a review group, including three dyeing and finishing engineers from Germany, Bangladesh and Vietnam, and the rest of the members were from large export-oriented enterprises, such as Hengtian Lixin and Lianfa Group; key teachers from vocational colleges and universities, as well as engineers from quality inspection organisations. The first round of research was based on open-ended interviews, summarised around the question of "in which tasks do students frequently have language comprehension deviations", and initially extracted 17 task nodes. In the second round, experts evaluated the importance of the tasks and the urgency of teaching using a five-level Likert scale. In the third round, a consensus was reached, and four core task scenarios were finally crystallised. Figure 3 shows the four core task scenarios, which illustrate the language tasks, operational aspects, and modal requirements of each scenario. The process debugging scenario includes the explanation of process formula, confirmation of auxiliary dosing order, and communication of temperature control setting instructions; the equipment operation scenario includes the identification of button functions, identification of fault warning lights, and description of the operation process; the quality control scenario includes the identification of defects, filling in fabric inspection reports, and correspondence between terminology and images; and the accident handling scenario includes the emergency stop instructions, communication of abnormal alarms, and trouble-shooting dialogues. Through the above process, the study clarifies the task boundaries and language carriers of teaching resources development, providing a solid foundation for the design of subsequent multimodal resources.

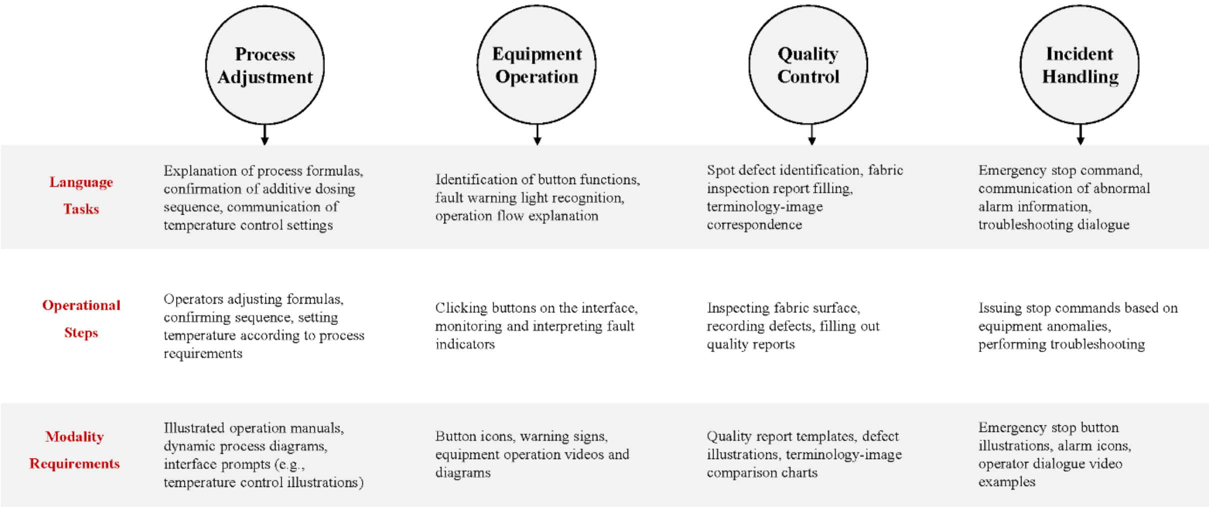

Figure 3. Structure of the four core task scenarios

Multimodal Teaching Resources Development

Based on the results of the demand analysis, the research team formulated a three-dimensional matching matrix of language task-visual modality-video elements, as seen in Table 2, which corresponds to the required modal presentation strategy for each type of language ability, to build a multimodal structure-dynamic-context triple alignment. Table 2: Linguistic task-visualisation.

Table 2. Three-dimensional matching matrix of linguistic task-visual modality-video elements

Linguistic ability	Visual Modality	Video design elements
Operational Command Understanding	Hot zone labelling map of the device interface	First view camera + Gesture tracking + English subtitle prompts
Process Parameter Description	Temperature-time curve dynamic graph	Process section gradient + parameter floating warning animation
Defect identification expression	Defective Cloth Comparison	Multi-angle light inspection display + defect terminology labelling
Report Filing and Dialogue	Form templates and example dialogues	Operator interview clips + contextual dialogue embedded scripts

The multimodal discourse construction method is used in the development process, and 3D interactive modules are created based on the virtual simulation training platform. All resources are equipped with micro-lesson units that can be embedded into the Learning Management System (LMS) platform, including four sub-modules: task introduction, modal input, language output, and task feedback. The video recording is synchronised with a high-definition head-mounted camera and a stabilised shoulder-mounted camera to ensure the acquisition of complete information about the equipment interface, operator movements, and verbal interactions. Later, real-time subtitles, parameter annotations, and action cue boxes are added by the video analysis software to build a teaching resource that integrates language-vision-action synchronously.

Experimental design of teaching intervention

In order to empirically assess the effect of multimodal resources on the enhancement of vocational English proficiency, this study adopted a double-blind controlled experimental design. The experimental subjects were second-year students in the direction of dyeing and finishing from a textile vocational college in South China, with a total sample of 86, randomly divided into an experimental group and a control group of 43 each. The experimental group received the newly developed multimodal resources, which included visualisation operation process video, contextual task microclasses, and VR interaction module. The control group followed the traditional PPT-plus-lecture mode, using existing textbooks. The experimental intervention cycle is 4 weeks, a total of 16 hours. The teaching content covers the operation of the dyeing machine, additives feeding, defect judgment, accident emergency, and other tasks. All courses were taught by two instructors with experience in national skill competitions to ensure consistency in teaching control variables. After the class, the two groups of students were respectively subjected to a language test, an operation task simulation, and a satisfaction survey, and the classroom video recording was collected for behavioural analysis.

Data collection and analysis methods

To ensure the comprehensiveness and explanatory power of the research results, the data collection design starts from the three dimensions of language proficiency, cognitive load, and classroom behaviour, and the selection of tools, implementation process, and reliability control. The complete data-collection design is summarised in Table 3.

Table 3. Data Collection Design

Data type	Instrument/method	Reliability Indicators and Analysis Process
Vocational English Proficiency	Test of English Operational Proficiency in Textile Dyeing and Finishing	α=0.89, covering the four dimensions of listening, speaking, graphic matching, and task retelling
Cognitive load	NASA-TLX Cognitive Load Scale Chinese version	Validated Factor Analysis (CFA) $\chi^2/df=1.82$, RMSEA=0.044
Classroom Behavior	Video transcript + NVivo coding analysis	Krippendorff's Alpha=0.81, 90% behavioural passages coded consistently
Interview transcripts	Semi-structured interviews + thematic categorisation analysis	Interviews included 8 experimental group students, 2 instructors, and 1 business mentor

The language proficiency test questions were jointly written by three key teachers with more than 10 years of teaching experience, and the questions were set to match the real tasks of the job, such as recognising instructions, reviewing the process flow, and filling in the defect record sheet. The scoring criteria were designed according to the CEFR framework and the Chinese vocational English proficiency level standard for higher vocational education. The cognitive load measure was adopted from the Chinese version of NASA-TLX (NASA-TLX v1.0, NASA Ames Research Centre; NVivo 14, Lumivero), covering six dimensions, including psychological load, time pressure, operational complexity, and effort. The test was arranged to be conducted after the 8th and 16th class periods to compare students' perceived changes in cognitive load at the beginning and end of the instruction. Classroom behaviours were captured on video and coded using NVivo software for behavioural segments, focusing on the following behavioural indicators: frequency of gaze, number of questions asked, length of task pauses, and use of image pointing language. All coders received uniform training and were cross-validated.

In summary, this study constructed a theory-practice-validation closed-loop research mechanism from needs analysis to resource development, and from instructional intervention to system evaluation. Through a highly contextualised, task-oriented, and modality-integrated teaching path, it explores the way to reconstruct English teaching in textile dyeing and finishing, and provides a replicable and transferable operational model for multimodal teaching reform in vocational majors.

RESULTS AND DISCUSSION

Based on the aforementioned teaching intervention experiments and multidimensional data collection, this study centres on the empirical analysis of the effect of vocational English core competence improvement, the change of students' cognitive mechanism, and the adaptability of multimodal design. The following discussion will be developed from the quantitative results, qualitative findings, and the extension of the theory-practice level.

Quantitative effectiveness of multimodal pathway

At the end of the experiment, the two groups of students were assessed by the Textile Dyeing and Finishing English Operational Ability Test, and the three key ability dimensions included: the ability to execute operational instructions, the ability to report faults in English, and the ability to interpret process parameters. It can be seen through the bar chart that the experimental group scored significantly higher than the control group on each competency, showing the significant effect of the multimodal pathway in improving vocational English proficiency. The experimental group's scores on all three dimensions are better than the control group's, and the difference is significant, proving the effectiveness of the multimodal teaching path in enhancing students' English proficiency, especially professional language proficiency. In terms of the ability to execute operational instructions, the experimental group scored 78.3, and the control group scored 53.6; the score of the experimental group was significantly higher than that of the control group, which indicated that the experimental group had a better ability to comprehend and apply the operational instructions when executing them. The experimental group scored 82.1, and the control group scored 47.2 in English fault reporting ability. The experimental group was able to use professional English more accurately and fluently when reporting faults, indicating that their English expression ability and ability to cope with complex situations have been significantly improved. In terms of process parameter explanation ability, the experimental group scored 85.4, and the control group scored 61.8. The experimental group showed stronger language expression and technical understanding when explaining and describing process parameters, and was able to more accurately convey process details.

In terms of score structure, the experimental group showed an average improvement of 26.4 percentage points in the rate of accurate retelling of multi-step verbal instructions, and a reduction of 42.7% in the error rate in the graphic matching and parameter explanation oral task. The control group was stylishly able to handle visually weakly related tasks, such as purely textual instructions, but showed significant comprehension deficits when confronted with spoken tasks containing graphical instructions or embedded operational details. Table 4 demonstrates the NASA-TLX six cognitive load scores comparison table statistics, which shows that the experimental group generally scored lower than the control group on all cognitive load dimensions, indicating that the experimental group's cognitive load of the task was significantly reduced under the multimodal pathway intervention. Specifically, in the dimensions of task complexity and effort, the scores of the experimental group were 42.1 and 38.5, with standard deviations of 6.8 and 7.1, respectively, which were significantly lower than those of the control group, which were 68.7 and 64.2, with standard deviations of 7.3 and 6.9, respectively. This difference suggests that the experimental group perceived a lower level of complexity and required effort in the execution of the task, suggesting that the multimodal support effectively reduced the cognitive load of the learning task. In terms of time pressure and mental load, the experimental group also showed lower cognitive load, with scores of 45.6 and 44.2, with standard deviations of 5.2 and 5.7, respectively, compared with the control group's scores of 60.4 and 62.1, with standard deviations of 6.1 and 7.0, which further validated that the multimodal pathway helped students understand the task content more efficiently by providing visual and interactive resources and reduced time pressure and mental load. The time pressure and psychological load. In addition, in terms of physical load, the experimental group scored 40.1 with a standard deviation of 6.3, while the control group scored 59.8 with a standard deviation of 6.8. This indicates that the experimental group had less physical manipulation burden due to poor comprehension during the operation.

Table 4. Comparison of the six cognitive load scores of NASA-TLX

Cognitive load dimensions	Mean score of experimental group (standard deviation)	Control group mean score (standard deviation)
Task complexity	42.1 (6.8)	68.7 (7.3)
Energy Consumption	38.5 (7.1)	64.2 (6.9)
Time pressure	45.6 (5.2)	60.4 (6.1)
Effort	38.5 (6.9)	64.2 (7.0)
Mental load	44.2 (5.7)	62.1 (7.0)
Physical load	40.1 (6.3)	59.8 (6.8)

Qualitative effectiveness of multimodal instruction

Through NVivo's content analysis of the experimental group's classroom videos and the coding of students' post-class interview data, the study distilled two key cognitive mechanisms:

Analysis of Dual Coding Mechanisms

In information processing involving multiple inputs from the term-action-device interface, the subtitle labelling and visual hotspot cues of the videos significantly increased students' active processing of language [30,31]. Eighty-five per cent of the students reported in the interviews that the synchronised presentation of subtitle + action helped them to establish the relationship between terminology and actual action. The simultaneous presentation of subtitles and actions helped them to establish anchor points between terms and actual actions, especially in understanding process statements such as Start pre-wash cycle and Add softening agent before 85°C. In particular, when comprehending process statements such as Start pre-wash cycle and Add softening agent before 85°C, the figurative hand movements together with the subtitle information made it easier to internalise them into actionable language templates. The coding results showed that students in the experimental group used terms in the simulation task with a 91.2% correctness to context match, compared to 64.7% in the control group. The repeated exposure of embedded linguistic points in the video with a visual reinforcement mechanism became an important pathway for terminology construction. Figures 4 and 5 illustrate interactive learning scenarios of students while they are engaged in course English learning, helping to explain how the multimodal teaching pathway affects students' learning outcomes by demonstrating the instructional feedback process. As illustrated in Figure 4, the instructor is delivering a task-oriented explanation while synchronously referencing the equipment interface displayed on screen, which reinforces the coupling between linguistic input and visual context. Figure 5 captures students' real-time interaction with the multimodal interface-note how one student points to a button labelled 'Emergency Stop' while repeating the phrase aloud. This behaviour aligns with the dual-coding mechanism previously discussed, where gesture–subtitle anchoring facilitates terminology internalisation. These images thus provide empirical evidence for the cognitive processes described, rather than serving as detached illustrations.

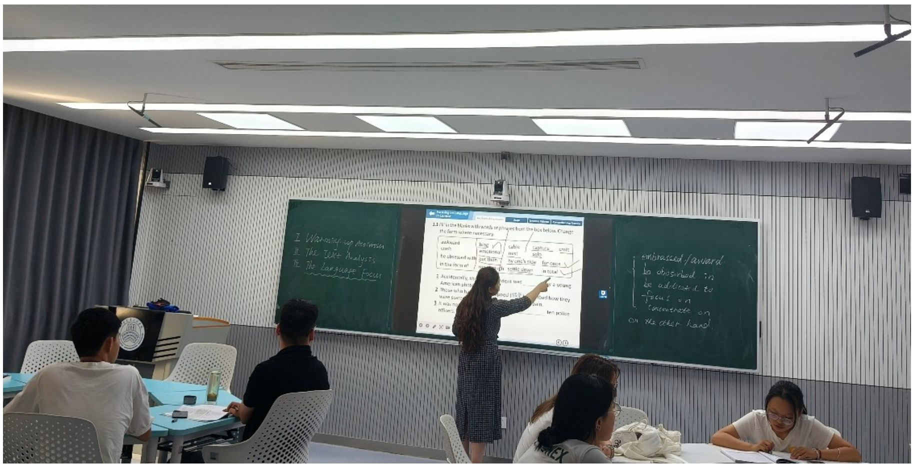

Figure 4. The teacher delivering the course in English

Figure 5. Students' interactions while engaged in the course of English learning

Analysis of Attention Guiding Mechanisms

In analysing the video interaction resources used by the experimental group, the study found that error warning boxes, such as error button highlighting and interface blinking prompts, significantly reduced the number of student misunderstandings. In the device simulation task, the average number of times students triggered operational errors was 1.3 times, compared to 3.9 times in the control group. In the post-class questionnaire feedback, about 85% of the students explicitly stated, The warning box prompts let me know what not to do and avoid misunderstandings. Figure 6 demonstrates the mechanism of the video warning labelling effect, which mainly reflects the logical path between the warning box and students' behavioural responses. The figure first shows the warning box, which reminds students which actions are wrong by highlighting the wrong buttons and flashing the interface to avoid repeating the mistakes. This warning box serves to guide the students to notice the errors in time during the operation so as to minimise the occurrence of misunderstandings. Next, the figure depicts the students' behavioural responses. When students see the warning box, they immediately adjust their actions to avoid misinterpretation. This feedback mechanism enables students to maintain a high level of operational accuracy during the learning process and greatly reduces the frequency of errors. The logical path in the figure demonstrates the cause-and-effect relationship of the whole process. First, the student performs an action that triggers an error, and a warning box then appears. The student sees the warning box and adjusts his or her actions, thus reducing misunderstandings and errors. Whereas, in the absence of the warning box, students may not receive timely feedback and continue to perform incorrect operations, leading to a higher incidence of misunderstandings.

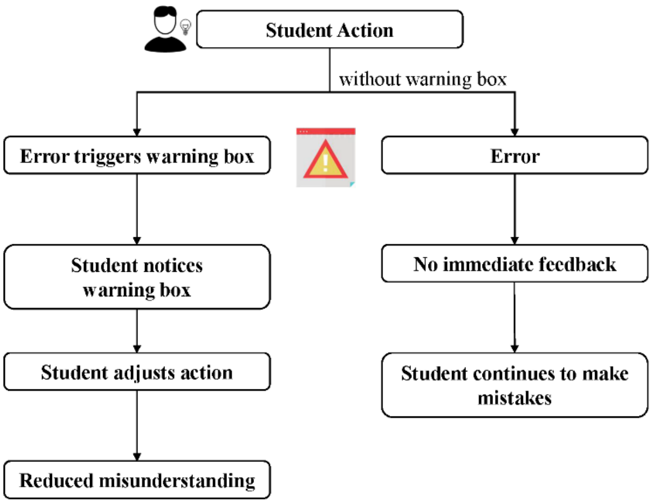

Figure 6. Video warning labelling effect mechanism

Further reflection on theory and practice

Multimodal adaptation principle construction

Based on the empirical results, this paper makes industrialized corrections to the multimodal selection model proposed by Delany to form three principles of occupational scene adaptability, as follows: high-risk operations to strengthen the dynamic warning modality: for example, during the process of dyeing machine temperature and over-pressure, it should be paired with the dynamic red light flashing, the English warning subtitles and the real sound effect prompts to maximize the situational risk perception; the abstract concepts give priority to the use of micro visualisation: for example Abstract concepts are prioritized for micro-visualisation: for example, for abstract processes such as adsorption and diffusion of dye molecules, animation demonstrations are more helpful than static descriptions to help students build mental models and improve the accuracy of language expression; Cross-cultural tasks require the introduction of gesture annotations: in multinational employee participation scenarios such as international fabric inspection and quality control communication, visual annotations are added for common fabric inspection gestures, such as three-finger The addition of visual annotations for common fabric inspection gestures, such as the three-finger press for folded area, can help to enhance language transferability. Figure 7 illustrates the Three Principles of Industry Adaptability, with triangles or radial structures to demonstrate the three core principles and their application. Dynamic Warning Modal: For high-risk operations, such as temperature control problems in dyeing machines, dynamic warnings, such as flashing red lights, warning captions, and live sound effects, should be used to maximise the perception of risk. Microscopic visualisation of abstract concepts: For abstract processes, such as dye molecule adsorption and diffusion, animated demonstrations should be used to help students better understand them, rather than just static illustrations. Gesture annotations in cross-cultural tasks: In tasks involving multinational employees, such as international fabric inspection and quality control, gesture annotations should be introduced, such as the use of three-finger pressure to indicate wrinkled areas, to enhance language transferability and cross-cultural understanding.

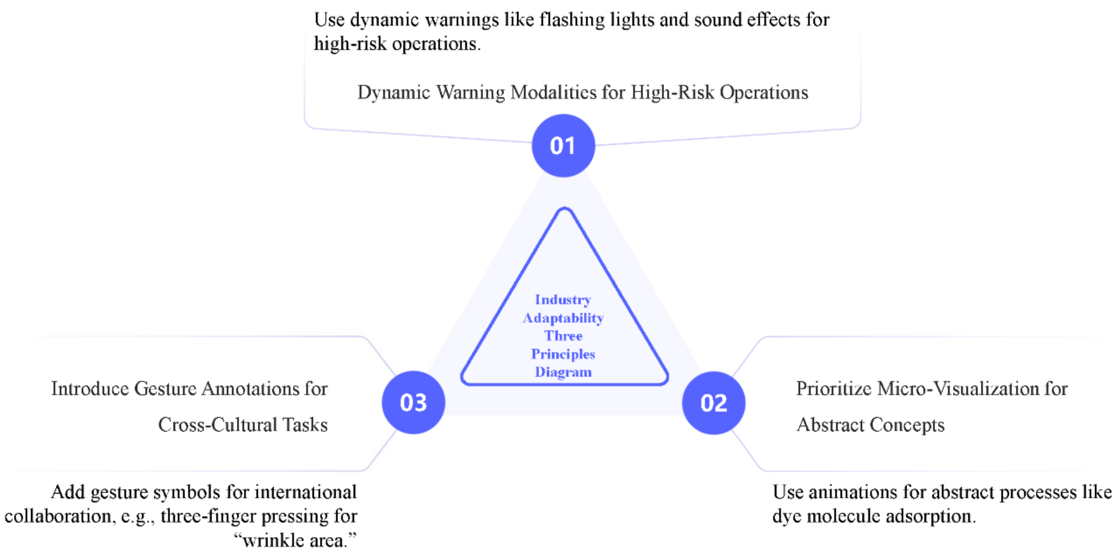

Figure 7. Three Principles of Industry Adaptability

Reconciliation of pedagogical benefits and development costs

Although the advantages of multimodal resources in enhancing learning effects have been proven, the development and application problems of teaching resources faced by vocational colleges and universities should not be overlooked in reality: the high cost of resource production: the average production time for a high-quality practical video with subtitles, annotation, and motion tracking is about 3 hours, and professional teachers need to cooperate with script writing, equipment demonstration, and post-proofreading; the high threshold of application for teachers: some of them lack the knowledge of video editing, subtitle production, and image processing; and the high cost of teaching. Video editing, subtitle production, and image processing capabilities rely on professional development teams and interdisciplinary collaboration. In this paper, it is suggested to adopt the "modularisation + hierarchical development strategy": i.e., take the key tasks such as equipment startup process and accident investigation process as units, build pluggable micro-resource modules, and stratify the difficulty level according to the semester level, and gradually guide the teachers to participate in the mechanism of resource co-construction, to form the mechanism of task-driven-capability-oriented-resource co-creation. The development path of task-driven, competence-oriented, resource co-creation is formed.

CONCLUSION

At the theoretical level, this study focuses on the realisation path of multimodal discourse integrating visual images and practical videos in vocational English teaching and constructs a locally-situated research paradigm. Based on the deconstruction of real occupational tasks and learners' cognitive characteristics, it develops and deploys a series of resources containing subtitle annotation, visual anchors and dynamic warnings. The empirical data clearly demonstrate that the combination of video keyframe anchors + terminology subtitles + visual highlighting significantly outperforms the traditional pathway in terminology-acquisition efficiency: in CEFR-B1 tasks, students' terminology-output correctness increased by 28.6 % and terminology-reproduction speed by 2.1×. Simultaneously, cognitive-load indices and task-operation error rates decreased, evidencing the pathway's effectiveness in cognitive modulation and language transfer. These findings allow us to propose the vocational multimodal discourse production model, which takes the deconstruction of authentic industry tasks as its starting point, relies on the precise matching of modality functions and incorporates cognitive-conflict design strategies to ultimately achieve the transfer and reconstruction of cross-modal language competence. The model provides a new perspective for analysing educational-migration problems in current multimodal research and offers methodological reference for the development and evaluation of multimodal teaching resources in other occupational scenarios such as mechanics, nursing and logistics, thereby extending the breadth of multimodal-discourse applications in educational technology.

Practically, the study has successfully built a standardised system for multimodal teaching-resource development in textile dyeing and finishing, encompassing four task modules (process debugging, equipment operation, quality control, accident handling), twelve multimodal resource types and a five-level quality-assessment scale covering semantic clarity, modality matching and interaction usability, thus supplying a normative basis for future industry-education co-construction of resources. Enterprises are advised to actively participate in the co-construction of job-language resources by establishing multimodal corpora grounded in actual work scenes, such as fabric-defect image banks, operation-error case collections and multilingual equipment-operation scripts, to support more efficient employee training and school-enterprise collaborative teaching. At the institutional level, teacher multimodal-teaching literacy needs to be strengthened; it is therefore recommended to formulate and promote the Multimodal Instructional Design Competency Certification Standards, which cover modality-combination strategies, task-based resource editing and cognitive-modulation design, so as to build a cross-disciplinary vocational-English teaching team. Although positive results have been achieved in theoretical construction and empirical validation, limitations remain: the study centres on English-medium tasks in textile dyeing and finishing and has not yet covered small-language enterprise contexts such as Turkish-speaking goods-inspection processes or Vietnamese-language equipment-panel recognition; future work should further expand cross-language adaptability. Moreover, some experimental content relies on VR and simulation equipment whose penetration in vocational schools in central, western and third-tier cities is low, hindering large-scale promotion. Subsequent research should explore low-cost substitutes and more flexible modality-adaptation mechanisms to enhance applicability and transferability.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This work is supported by the first-class course of college English in Xi'an Traffic Engineering Institute, Shaanxi Province, the industry-university cooperation and collaborative education project of the Ministry of Education of Xi'an Traffic Engineering Institute in 2024 "Application and Research of College English Blended Teaching Mode Based on Diagnostic Assessment" and the key teaching reform project in 2023 "Research on the Cultivation of College Students' Self-directed Learning Ability by Learning Input Intervention Integrating Diagnostic Assessment in Blended Teaching Mode"(XJY231015). Key Teaching Reform Project of Xi’an Traffic Engineering Institute "Path Design and Practice Research on

Empowering Personalised Learning of College English with Knowledge Graph Construction Driven by AI" (No. XJY251018).

REFERENCES

[1] Incelli EA. Engaging students in multimodal literacy practices in a university ESP context: towards understanding identity and ideology in government. Debates. Multimodal Commun. 2021,11(1):49–61. https://doi.org/10.1515/mc-2021-0004

[2] Jarodzka H, Skuballa I, Gruber H. Eye-Tracking in Educational Practice: Investigating Visual Perception Underlying Teaching and Learning in the Classroom. Classroom. Educ Psychol Rev. 2020,33(1):1–10. https://doi.org/10.1007/s10648-020-09565-7

[3] Suci DN, Basthomi Y, Cahyono BY, Anugerahwati M, Masuara F, Anggraini MP. How Do Vocational Students Perceive the Use of Telegram for their Online Reading Comprehension? HERMES - J Lang Commun Bus. 2022,(62):127–139. https://doi.org/10.7146/hjlcb.vi62.128116

[4] Rahmanu IWED, Molnár G. Multimodal immersion in English language learning in higher education: a systematic review. Heliyon. 2024,19(10):e38357. https://doi.org/10.1016/j.heliyon.2024.e38357

[5] Perez MM. Second or foreign language learning through watching audio-visual input and the role of on-screen text | Language Teaching. Camb Core. 2022,55(2):163-192. https://doi.org/10.1017/S0261444821000501

[6] Shewangizaw G, Hailu A, Haile W. A comparative analysis of language features across communicative English skills and garment vocational texts. Celt J Cult Engl Lang Teach Lit Linguist. 2024,11(2):353–376. https://doi.org/10.22219/celtic.v11i2.35885

[7] Diao LJ, Hu P. Deep learning and multimodal target recognition of complex and ambiguous words in an automated English learning system. J Intell Fuzzy Syst. 2021,40(4):7147–7158. https://doi.org/10.3233/jifs-189543

[8] Lin V, Barrett NE, Liu GZ, Chen NS, Jong MS-Y. Supporting dyadic learning of English for tourism purposes with scenery-based virtual reality. Taylor. 2023,36(5):906–942.

[9] Dressen-Hammouda D, Wigham CR. Evaluating multimodal literacy: Academic and professional interactions around student-produced instructional video tutorials. Video tutorials. System. 2022,105:102727. https://doi.org/10.1016/j.system.2022.102727

[10] Adeloye AA, Akinnibosun OF, Emidun OB. A Comparative Study of Entrepreneurial Components in Textile Design Curricula between Nigerian Universities and Polytechnics. J Art Des Music. 2024,3(2):8. https://doi.org/10.55554/2785-9649.1041

[11] Guo H, Gao W. Frontiers | Metaverse-Powered Experiential Situational English-Teaching Design: an Emotion-Based Analysis Method. 2023,13:1 -9. https://doi.org/10.3389/fpsyg.2022.859159

[12] Hu WL, Lin W, Li QZ, Yu XN, Zhu CY. Textile AI-Enhanced Translation System Based on Mapping Probability and In-context Learning. ACM Trans Asian Low-Resour Lang Inf Process. 2025,7(4):e4544. https://doi.org/10.1145/3746227

[13] Pratiwi MM, Nelmira W. Innovation in Vocational Learning: Project-Based Learning E-Module on the Textile Dyeing Course. J Educ Technol. 2024,8(1):175–184. https://doi.org/10.23887/ jet.v8i1.74954

[14] Teo T, Khazaie S, Derakhshan A. Exploring teacher immediacy-(non)dependency in the tutored augmented reality game-assisted flipped classrooms of English for medical purposes comprehension among the Asian students. Comput Educ. 2022,179:104406. https://doi.org/10.1016/j.compedu.2021.104406

[15] Sidabutar U, Sinaga NT, Sitorus N, Lestari FD. A multimodal analysis of a vocational English book and its effect on students' English proficiency. Linguist Cult Rev. 2021,1651–1665. https://doi.org/10.21744/lingcure.v5nS3.1951

[16] Jiang LJ, Ren W. Digital Multimodal Composing in L2 Learning: Ideologies and Impact. J Lang Identity Educ. 2021,20(3):167–182. https://doi.org/10.1080/15348458.2020.1753192

[17] Fedorenko S, Voloshchuk I, Sharanova Y, Glinka N, Zhurba K. Multimodality and digital narrative in teaching a foreign language. Arab World Engl J AWEJ Spec Issue CALL. 2021,(7):178–189. https://doi.org/10.24093/ awej/call7.13

[18] Ghoushchi S, Yazdani H, Dowlatabadi H, Ahmadian M. A multimodal discourse analysis of pictures in ELT textbooks: modes of communication in focus. jordan J Mod Lang Lit. 2021,13(4):623–644. https://doi.org/10.47012/jjmll.13.4.2

[19] European Commission. Textile4Europe: Final Report. Publications Office of the EU. 2021

[20] Lim FV, Toh WM, Nguyen TTH. Multimodality in the English language classroom: a systematic review of literature. Linguist Educ. 2022,69:101048. https://doi.org/10.1016/j. linged.2022.101048

[21] Cheng XX, Liu KF. Application of Multimedia Networks in Business English Teaching in Vocational College. J Healthc Eng. 1(1):1-9; https://doi.org/10.1155/2021/5519472

[22] Liang WJ, Lim FV. A pedagogical framework for digital multimodal composing in the English Language classroom. Innov Lang Learn Teach. 2021,15(4):306–320; https://doi.org/10.1080/17501229.2020.1800709

[23] Ponzio CM, Deroo MR. Harnessing multimodality in language teacher education: expanding English-dominant teachers' translanguaging capacities through a Multimodalities Encyclopedia. 2023,26(8):975–991

[24] Graur I-V, Scripcenco A, Avadanei M, Floreaburduja E, Vatra A-D, Rosca M-I. Application of the principle of sustainability in designing a garment collection for women. Collection for women. Innov Text Mater Process. 2025,257. https://doi.org/10.2478/9788367405829-fm

[25] Mandour BA. The art of hanging textiles between individual style and design authenticity: Blended learning model in art education. Art Des Commun High Educ. 2022,21(2):191–216. https://doi.org/10.1386/adch_00059_1

[26] Taze D, Hartley C, Morgan AW, Chakrabarty A, Mackie SL, Griffin KJ. Developing consensus in Histopathology: the role of the Delphi method. Histopathology. 2022,81(2):159-67. https://doi.org/10.1111/his.14650

[27] Gong YF, Lai C, Gao XA. Language teachers' identity in teaching intercultural communicative competence. Taylor Francis. 2022,35(2):134–150. https://doi.org/10.1080/07908318

[28] Lim FV, Unsworth L. Multimodal composing in the English classroom: recontextualising the curriculum to learning. Engl Educ. 2023,57(2):102–119. https://doi.org/10.1080/04250494.2023.2187696

[29] De Cock B, Aulit L, Cigada S, Greco S, Modrzejewska E, Palmieri R. The discourse of digital activism: a linguistic analysis of calls for action concerning the Fashion Revolution. Appl Linguist. 2024,45(6):1091–1110. https://doi.org/10.1093/ applin/amae046

[30] Suci D, Basthomi Y, Cahyono NY, Anugerahwati M. EFL teachers' perspectives on the use of digital multimodal technology in vocational schools. ICERI2021 Proc. 2021,1:7208–7216. https://doi.org/10.21125/iceri 2021.1618

[31] Mandour BA. Traditional textile printing between spontaneity and planning: a study of creative practice. Int J Educ Arts. 2024,25(4). https://doi.org/10.26209/ijea25n4