



Assessing Value-Oriented Education in Tertiary EFL: A Formative Feedback Perspective

Liu Tianwei^{1,2}, Hema Vanita Kesevan^{1*}

¹ Faculty of Languages and Communication, Universiti Pendidikan Sultan Idris,
35900 Tanjong Malim, Perak, Malaysia

² College of Foreign Languages, Inner Mongolia University of Technology, 010051 Inner Mongolia, China

*Corresponding Author

p20212002830@siswa.upsi.edu.my

hemakesevan@fbk.upsi.edu.my

Abstract: Against the national policy mandate of integrating value-oriented education (curriculum-based ideological and political education) into tertiary EFL teaching in China, assessment practices remain fragmented and over-reliant on summative language tests, with few empirical qualitative studies exploring frontline college English lecturers' authentic perceptions of value evaluation mechanisms. This qualitative study adopted semi-structured interviews with 20 college English teachers from four types of public universities in Inner Mongolia, positioning Feedback Theory (Nicol & Macfarlane-Dick, 2006) as the central analytical lens, supplemented by Teacher Cognition Theory (Borg, 2003) and Sociocultural Theory (Vygotsky, 1978). Through six-phase thematic analysis via NVivo software, three core thematic clusters emerged regarding assessment practice: (1) core dilemmas of value evaluation, including subjectivity of value growth, language-value assessment incompatibility, and the problem of performative compliance; (2) multi-source formative assessment pathways favored by teachers, demonstrating how diversified feedback from self-assessment, peer evaluation, and teacher process observation can overcome single-source scoring bias; (3) structural constraints restricting the implementation of multidimensional value assessment. The integrated "Cognition–Practice–Assessment" model constructed in this study begins to address the gap in holistic assessment frameworks for value-oriented EFL education. Findings suggest practical pedagogical and institutional implications: universities should build formative evaluation systems centered on multi-source feedback, provide targeted teacher training on value assessment design, and balance linguistic proficiency measurement with long-term value development tracking. Keywords: value-oriented education; college English; formative assessment; multi-source feedback; teacher cognition; qualitative thematic analysis.

Keywords: value-oriented education; college English; formative assessment; multi-source feedback; teacher cognition

1. Introduction

1.1 Research Background

China's national higher education policy issued in 2020 stipulated that all undergraduate courses must integrate value cultivation alongside knowledge transmission, naming this educational paradigm value-oriented education (ideological and political curriculum education) (MOE, 2020). As a compulsory two-year foundational course for all university undergraduates, College English bears dual missions: improving students' intercultural linguistic competence and shaping their patriotism, cultural confidence, critical thinking, and social responsibility (Liu & Yue, 2020). Nevertheless, long-standing EFL evaluation mechanisms in Chinese universities prioritize summative language examinations (CET, final written tests) while neglecting measurable, operable assessment tools for students' value development (Zhang & Wang, 2020).

Current classroom assessment for value growth faces prominent contradictions: teachers recognize the necessity of evaluating value outcomes but lack standardized formative evaluation routines; value formation is a slow, internalized psychological process difficult to quantify via traditional scoring rubrics; institutional assessment systems do not incorporate value growth indicators into regular teaching evaluation. Existing domestic literature mostly discusses value education assessment from theoretical or macro policy perspectives, lacking micro-level qualitative evidence from frontline English lecturers' real classroom experience (Zhang & Wang, 2020). Internationally, comparative moral education research (e.g., LePage et al., 2011) has primarily examined Western K-12 contexts, with limited empirical data on Chinese tertiary EFL value evaluation. This study responds to this research gap by exploring how college English teachers perceive, implement, and struggle with value-oriented assessment practices.

1.2 Statement of the Problem

Three core research problems frame this inquiry:

Most Chinese college English assessment systems separate linguistic testing from value cultivation evaluation, with no unified multidimensional evaluation framework covering process and outcome of value growth.

Frontline EFL lecturers lack specific assessment literacy for value-oriented evaluation—including rubric design, feedback provision, and formative tracking—leading to mechanical or absent value assessment in daily instruction.



The performative nature of students’ expressed values, combined with institutional reliance on summative language tests, creates persistent evaluation dilemmas that cannot be resolved by single-teacher scoring without systematic multi-source feedback mechanisms.

1.3 Research Questions

RQ1: What perceptions do Chinese college English lecturers hold regarding the core challenges of assessing value-oriented education? RQ2: What multi-source formative assessment practices do teachers regard as effective for tracking students’ value development in EFL classrooms? RQ3: What institutional and pedagogical barriers hinder the full implementation of integrated value-and-language assessment systems?

1.4 Significance of the Study

Theoretically, this study positions Feedback Theory (Nicol & Macfarlane-Dick, 2006) as the central analytical lens, supplemented by Teacher Cognition Theory (Borg, 2003) and Sociocultural Theory (Vygotsky, 1978), to construct an integrated “Cognition–Practice–Assessment” analytical framework for value-oriented EFL teaching, expanding the application of formative assessment theory into Chinese curriculum value education contexts. Practically, the thematic findings offer actionable formative assessment strategies for university English instructors and institutional curriculum administrators, providing evidence to reform over-summative evaluation mechanisms in tertiary EFL courses.

1.5 Scope and Limitations

Participants are 20 college English lecturers from four public universities in Inner Mongolia, limiting cross-regional generalizability; data rely only on semi-structured interview narratives without classroom observation or student evaluation artifacts for triangulation. Future mixed-method research combining classroom recordings and student assessment portfolios can supplement this study’s limitations.

2. Literature Review

2.1 Value-Oriented Education in Chinese Tertiary EFL

Value-oriented education (VOE) in Chinese College English is defined as embedding patriotism, cultural identity, critical cross-cultural awareness, and civic responsibility into language input, output, and classroom interaction, rather than delivering isolated political content (Pan & Yuan, 2021). Unlike Western moral education, VOE is embedded within all disciplinary curricula as a national educational requirement, with College English acting as a critical medium to resolve “cultural aphasia” (students’ inability to express Chinese culture in English) (Chen, 2022).

Throughout this paper, the term “value-oriented education” (VOE) is used consistently to refer to China’s national policy of integrating moral and civic education into all university curricula. Alternative labels such as “ideological and political education” (课程思政) refer to the same policy initiative; VOE is preferred here because it more accurately captures the educational purpose of the policy as implemented in EFL contexts.

Existing domestic scholarship focuses on textbook value element excavation and VOE teaching methods, while assessment research remains underdeveloped. Most studies only propose theoretical evaluation indexes without empirical verification from teacher practice (Li, 2022). Overseas moral assessment literature prioritizes standardized questionnaires for adolescent moral development, which cannot adapt to the localized policy context of Chinese tertiary EFL (Asif et al., 2020).

2.2 Core Theoretical Foundations

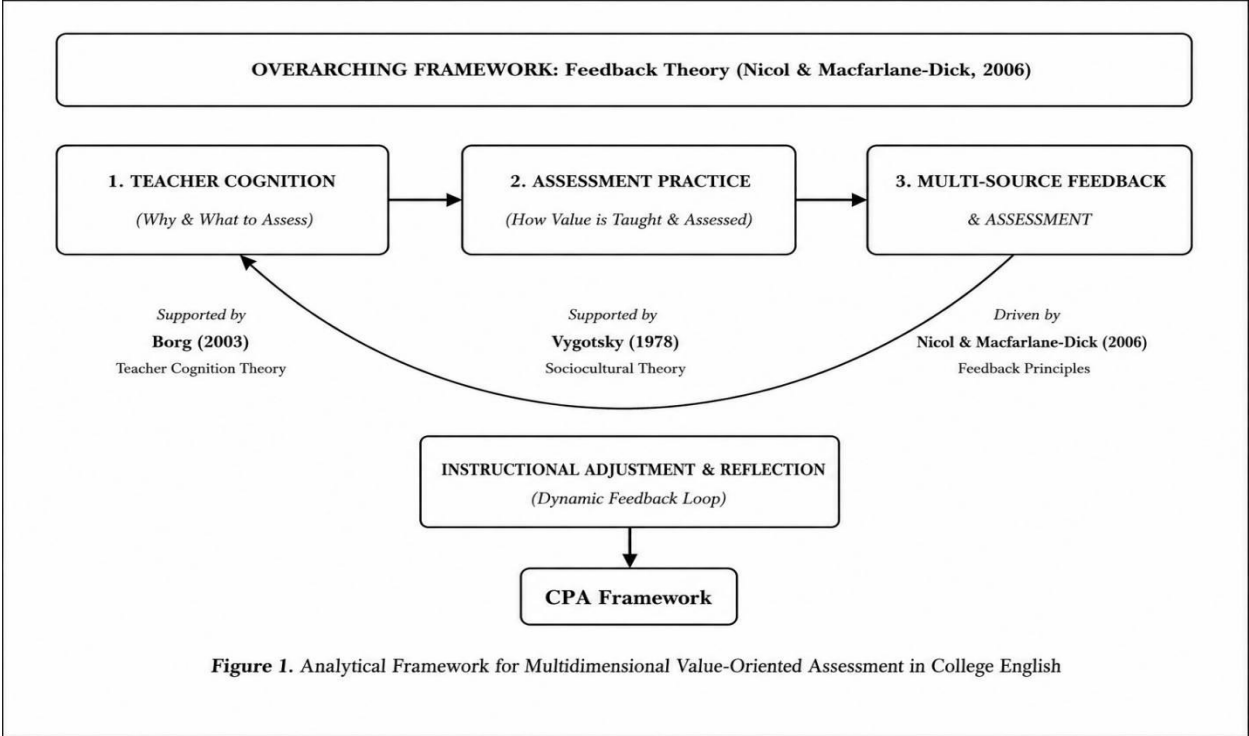


Figure 1. Analytical Framework for Multidimensional Value-Oriented Assessment in College English

Figure 1. Analytical Framework for Multidimensional Value-Oriented Assessment in College English

Figure 1 illustrates the theoretical framework integrating Teacher Cognition Theory (Borg, 2003), Sociocultural Theory (Vygotsky, 1978), and Feedback Theory (Nicol Figure 1 presents the conceptual analytical framework derived from the integration of three theoretical perspectives; it should be distinguished from Figure 2, which operationalizes this conceptual framework into the empirically constructed CPA model incorporating specific dimensions and elements identified from the interview data.& Macfarlane-Dick, 2006) for value-oriented assessment in tertiary EFL. Positioned as the central analytical lens, Feedback Theory provides the overarching logic that connects teacher cognition and assessment practice through an iterative evaluation loop. Specifically, Teacher Cognition Theory accounts for how instructors' policy understanding and value beliefs shape their assessment design (Cognition). Sociocultural Theory explains how interactive, task-based, and simulation activities mediate value learning among students (Practice). Finally, Feedback Theory binds these dimensions together by demonstrating how multi-source evaluation data (Assessment) generate actionable feedback to continually adjust teacher cognition and pedagogical practice.

2.2.1 Feedback Theory as the Central Framework

Nicol and Macfarlane-Dick's (2006) formative assessment model serves as the core analytical framework for this study. Feedback Theory emphasizes that assessment is not merely measurement but a formative process that promotes learning through systematic evaluation cycles. In the VOE context, feedback theory explains why single-teacher assessment cannot adequately capture value development—a multi-source feedback loop integrating self-assessment, peer feedback, and teacher observation is required to construct a balanced evaluation system. The CPA framework (Cognition–Practice–Assessment) constructed in this study is derived directly from Feedback Theory: cognition corresponds to teachers' assessment beliefs and cognition, practice corresponds to assessment practices, and assessment corresponds to feedback cycles that close the loop between evaluation data and instructional adjustment, forming a continuous “assessment–adjustment–reassessment” mechanism.

Specifically, Nicol and Macfarlane-Dick's (2006) seven principles of good feedback practice stipulate that good feedback practice: (1) helps clarify what good performance is (goals, criteria, expected standards); (2) facilitates the development of self-assessment (reflection) in learning; (3) delivers high quality information to students about their learning; (4) encourages teacher and peer dialogue around learning; (5) encourages positive motivational beliefs and self-esteem; (6) provides opportunities to close the gap between current and desired performance; and (7) provides information to teachers that can be used to help shape teaching. In the VOE assessment context, principles 2 (self-assessment), 4 (teacher–peer dialogue), and 6 (closing the gap) are particularly salient—teachers must help students self-assess their value development, provide clear criteria to close the gap between current and desired value expression, and leverage peer evaluation as a complementary source of assessment evidence.

2.2.2 Teacher Cognition Theory as Supporting Framework

Borg's (2003) teacher cognition theory explains how teachers form assessment decisions based on their beliefs and professional knowledge. Within the Feedback Theory framework, teacher cognition explains why different instructors hold varying perceptions and practices regarding VOE assessment. Lecturers' understanding of national value education policies determines whether they allocate assessment weight to students' value growth beyond language scores. Teachers with fragmented VOE cognition tend to ignore formative value tracking and rely only on final exam grades, whereas those with integrated cognition actively design multi-source feedback mechanisms.

2.2.3 Sociocultural Theory as Supporting Framework

Vygotsky's (1978) zone of proximal development (ZPD) and internalization theory explain value formation as a social interactive process: students absorb core values through scaffolded classroom discussions, cross-cultural text analysis, and group collaborative tasks. Within the Feedback Theory framework, sociocultural theory explains why peer feedback and collaborative assessment are effective—value internalization occurs through social interaction, and assessment cannot be limited to one-time testing; it requires continuous process evaluation capturing interactive value internalization over a semester. Scaffolding assessment (e.g., reflective journals, group presentation value rubrics) aligns with sociocultural learning mechanisms and reinforces the feedback loops central to the overarching theoretical framework.

2.3 Research Gap Synthesis

Lack of qualitative empirical research recording frontline EFL teachers' authentic assessment dilemmas and practical strategies for VOE;

No integrated theoretical framework linking teacher cognition, sociocultural value learning, and multi-source formative assessment for Chinese College English;

Insufficient discussion of institutional barriers restricting the popularization of multidimensional value evaluation.

3. Methodology

3.1 Qualitative Research Design

This study adopted an interpretivist qualitative design using semi-structured one-on-one interviews as primary data collection tools, with thematic analysis (Braun & Clarke, 2006) supported by NVivo 12 for data coding and theme extraction. Qualitative methods are appropriate here because the research targets teachers' subjective perceptions, unquantifiable classroom practices, and contextual institutional constraints that cannot be captured by quantitative surveys. This study draws on interview data from a larger qualitative project examining College English lecturers' perceptions of ideological and political education across multiple dimensions. The present paper focuses specifically on the assessment dimension, extracting and analysing assessment-related themes from the complete dataset.

3.1.1 Data Filtering Process

To ensure that the present analysis focuses exclusively on assessment-related dimensions, we applied a systematic data filtering procedure. Using NVivo 12, we first created a parent node labeled ‘Assessment & Evaluation’ with child nodes including ‘Assessment Challenges,’ ‘Formative Assessment Strategies,’ ‘Feedback Mechanisms,’ ‘Assessment Tools and Rubrics,’ and ‘Institutional Assessment Barriers.’ Two researchers independently coded all 20 interview transcripts against the pre-established node structure. Inter-coder agreement was calculated as the percentage of coding decisions in which both coders assigned the same node to a given text segment, computed as follows: $\text{Agreement} = (\text{Number of agreements} / \text{Total number of coding decisions}) \times 100\%$. The initial coding round yielded an agreement rate of 87%, which exceeds the commonly accepted threshold of 80% for qualitative research (Campbell et al., 2013). Discrepancies (13%) were resolved through discussion until consensus was reached. To illustrate the coding process, an initial code such as “no clear standards” extracted from a statement like “There is no clear standard for evaluating students’ value growth” was first assigned to the parent node ‘Assessment Challenges.’ After discussion, this code was refined and subsumed under the sub-theme ‘Lack of standardized assessment criteria’ within Theme 1. Another example: the code “use peer feedback” from “I combine self-assessment, peer feedback and my observation” was initially placed under ‘Formative Assessment Strategies’ and later consolidated into the sub-theme ‘Multi-source evaluation strategies’ within Theme 2. Interview segments that addressed general VOE teaching methods, resource development, or technology integration without direct reference to assessment, evaluation, or feedback were excluded from the present analysis. This filtering yielded 142 assessment-relevant coded segments across the three RQs, which formed the data corpus for the thematic analysis reported below.

3.2 Participants

Purposive sampling recruited 20 full-time college English lecturers from four Inner Mongolian public universities (comprehensive, tech, agricultural, ethnic university). Sampling criteria: minimum 1-year College English teaching experience, participation in institutional VOE training, balanced gender (10 male / 10 female), mixed master’s and doctoral qualifications (10 each), teaching experience ranging from 2–26 years. All participants used pseudonyms (Na, Gao, Hu, etc.) to guarantee anonymity for double-blind compliance.

Table 1

Participant Characteristics (N = 20)

Characteristic	Category	n
Gender	Male	10
	Female	10
Academic qualification	Master’s degree	10
	Doctoral degree	10
Institution	Four public universities in Inner Mongolia	20
Teaching experience	2–26 years	—
Total participants		20

Note. Data summarized from 20 full-time College English lecturers recruited through purposive sampling.

Table 1: Participant Characteristics

3.3 Interview Protocol & Procedures

The semi-structured interview guide centered on three RQs, covering:

Perceived difficulties of evaluating students’ value growth;

Usable formative assessment activities teachers adopted;

Institutional and classroom barriers to integrated value-language evaluation.

Each interview lasted 40–50 minutes, conducted offline in teachers’ offices or via WeChat video, fully audio-recorded with written informed consent. All recordings were transcribed verbatim into English transcripts for coding. A pilot study with 5 teachers refined ambiguous interview questions before formal data collection.

All interviews were conducted in Mandarin Chinese, which is the primary language of instruction and daily

communication for all participants. Following each interview, the audio recordings were transcribed verbatim in Chinese. The Chinese transcripts were then translated into English by the first author, who is bilingual in both languages. To ensure translation accuracy, a bilingual colleague with expertise in applied linguistics independently back-translated three randomly selected transcripts into Chinese. Comparison of the back-translations with the originals revealed no substantive discrepancies. The English translations were used for all subsequent coding and analysis.

This study received ethical approval from the Ethics Committee of Universiti Pendidikan Sultan Idris (Approval No. DPPK/2023/003). All participants provided written informed consent prior to the interviews. Participants were assured of their right to withdraw at any time and of the confidential treatment of their data.

The pilot study involved five additional college English lecturers who were not included in the final sample of 20 participants. The pilot data were used solely to refine the interview protocol and were not incorporated into the analysis.

3.4 Data Analysis Steps

Thematic analysis was conducted following Braun and Clarke's (2006) six-phase procedure.

Phase 1 (Data familiarization): The first author read and re-read all 20 interview transcripts multiple times to achieve immersive familiarity with the data.

Phase 2 (Generating initial codes): Using NVivo 12, the first author conducted open inductive coding of the 142 assessment-relevant segments, generating initial codes directly from the data without imposing pre-existing theoretical categories. For example, the statement "Value development is invisible and cannot be measured like language scores" was coded as "value invisibility," and "I combine self-assessment, peer feedback and my observation" was coded as "multi-source integration."

Phase 3 (Searching for themes): Initial codes were collated and grouped into potential sub-themes based on conceptual overlap. This stage involved an iterative process of moving between the coded extracts and the overall dataset.

Phase 4 (Reviewing themes): The preliminary sub-themes were reviewed against the full dataset to ensure internal coherence and distinctiveness. Overlapping sub-themes were merged; sub-themes lacking sufficient data support were discarded.

Phase 5 (Defining and naming themes): Three overarching themes were finalized, each capturing a distinct dimension of the assessment phenomenon and corresponding to one research question.

Phase 6 (Producing the report): Selected vivid extracts were chosen to illustrate each theme, and the narrative was constructed to present a coherent analytical argument.

The relationship between the deductive filtering and inductive coding deserves clarification. The deductive phase (creating the 'Assessment & Evaluation' parent node and extracting assessment-relevant segments) served as a data scoping procedure to focus the analysis on the assessment dimension. Within this scoped dataset, the six-phase thematic analysis proceeded entirely inductively: initial codes emerged from the data, and themes were constructed bottom-up rather than being derived from the theoretical framework or predetermined by the research questions. The three final themes thus represent genuine data-driven patterns that align with the three RQs, rather than categories imposed by the RQs.

3.5 Trustworthiness Strategies

Following Lincoln & Guba (1985):

Credibility: Triangulation of multi-institutional teacher perspectives + member checking;

Transferability: Thick description of participant demographics and university VOE institutional contexts;

Dependability: Complete audit trail of all coding memos and interview transcripts;

Confirmability: Reflexive researcher journal to reduce interpretive bias.

4. Findings

Three interconnected thematic clusters corresponding to the three research questions are presented below, supported by direct interview quotations (block quotes formatted per APA 7).

4.1 Theme 1: Core Dilemmas in Assessing Value-Oriented Education

All 20 lecturers highlighted the inherent subjectivity of value development as the primary assessment barrier. Unlike vocabulary, grammar, or writing skills with objective scoring standards, patriotism, cultural confidence, and critical thinking are internal affective-cognitive traits without uniform measurement benchmarks. Participants acknowledged that the fundamental challenge lies in the difficulty of quantifying ideological outcomes:

Among the 20 participants, 14 specifically mentioned the difficulty of distinguishing genuine value internalization from performative compliance, while 8 raised concerns about the attribution problem—that is, uncertainty about whether observed value changes resulted from deliberate instruction or broader social maturation. Three participants (Lecturers 5, 9, and 16) expressed relatively more confidence in their ability to assess values through longitudinal behavioral observation, representing a minority view that contrasts with the predominant uncertainty expressed by other participants.

"The effectiveness of ideological and political teaching is difficult to quantify. Unless using high-tech methods to observe students' brain activity, which isn't practical, we rely on behavioral indicators. The ultimate goal should be behavioral change, but this is long-term and difficult to measure directly in classroom contexts. We need better assessment approaches for value development." (Interviewee 9, Zhang)

Table 1
Thematic Map of Formative Assessment of Value-Oriented Education in EFL Classrooms

Research Question	Theme (Level 1)	Subthemes (Level 2)	Brief Description	Illustrative Example (Participant)
RQ1: What perceptions do Chinese college English lecturers hold regarding the core challenges of assessing value-oriented education?	Theme 1: Core dilemmas in assessing value-oriented education	1.1 Subjectivity and attribution uncertainty of value development	Value growth is internal, long-term, and influenced by many factors, making it difficult to measure objectively. Assessment instruments must distinguish genuine internalization from performative compliance, requiring multi-source verification and longitudinal observation.	"Value development is invisible and cannot be measured like language scores." (Interviewee 7, Wang 1)
		1.2 Lack of standardized assessment criteria	Universities lack unified indicators and rubrics for value assessment, creating inconsistency across instructors and institutions.	"There is no clear standard for evaluating students' value growth." (Interviewee 12, Guo)
		1.3 Insufficient assessment training	Lecturers have limited training and experience in designing formative value assessment tools, particularly rubrics and multi-source feedback protocols.	"We haven't received systematic training on how to assess values." (Interviewee 3, Hu)
RQ2: What multi-source formative assessment practices do teachers regard as effective for tracking students' value development in EFL classrooms?	Theme 2: Effective multi-source formative assessment practices	2.1 Multi-source evaluation strategies	Combining self-, peer-, teacher feedback and longitudinal tracking to gain a holistic picture of value development, addressing single-teacher scoring bias.	"I combine self-assessment, peer feedback and my observation to evaluate students." (Interviewee 5, Ding)
		2.2 Creative task-based assessment	Using authentic and creative tasks to evaluate students' value understanding and expression, making abstract values visible and assessable.	"Projects, posters and presentations help students express their values authentically." (Interviewee 9, Zhang)
		2.3 Simulation-based and process-oriented assessment	Employing real-life simulations and continuous process assessment to capture growth in values, with specific scoring rubrics for intercultural competence negotiation.	"Simulations and journals show how students think and grow over time." (Interviewee 16, Wu)
RQ3: What institutional and pedagogical barriers hinder the full implementation of integrated value-and-language assessment systems?	Theme 3: Institutional and pedagogical barriers to implementation	3.1 Assessment system misalignment	Current systems prioritize language scores; value indicators are not included in official evaluation, creating a structural gap between policy mandates and assessment practice.	"University evaluation still focuses on CET and final exam scores." (Interviewee 2, Gao)
		3.2 Professional development gaps	Lack of targeted training and ongoing professional support for value assessment design, particularly in rubric development and formative feedback techniques.	"We need more practical training and examples." (Interviewee 11, Chen)
		3.3 Practical operational constraints	Large class size, heavy workload and time pressure limit formative assessment. Teachers lack standardized evaluation templates and rubric frameworks to manage assessment at scale.	"Large classes and heavy workload make it hard to track values." (Interviewee 17, Niu)
		3.4 Fragmented institutional support	Insufficient policy, resources and collaboration to support systematic value assessment. Current support tends to be episodic rather than sustained.	"There is limited institutional support and no clear guidelines." (Interviewee 18, Zhou)

Note. The table summarizes the three overarching themes and their subthemes derived from the qualitative analysis of 20 college English lecturers, with updated descriptions reflecting attribution uncertainty, performative compliance, and language-value uncoupling dimensions.

Table 2 Summary of Themes and Subthemes

Beyond measurement ambiguity, teachers confronted the challenge of attribution—observing positive behavioral changes in students while being uncertain whether these changes resulted from deliberate value-oriented instruction or broader social maturation. Several participants expressed that this attribution problem made them reluctant to assign formal value assessment scores, fearing that evaluation might reflect students' general development rather than the specific impact of their teaching. This attribution uncertainty compounds the assessment difficulty: even when teachers observe desirable value-related behaviors—such as students voluntarily engaging in community service or expressing cultural pride in class discussions—they cannot confidently attribute these behaviors to their instructional interventions. As a result, multiple participants argued that formal summative scoring of value development risks penalizing or rewarding students for factors beyond the teacher's influence, thereby undermining the validity and fairness of value assessment.

This attribution problem is further complicated by the absence of clear causal indicators linking specific instructional activities to observed value changes. Unlike language skill development, where pre-tests and post-tests can provide tangible evidence of learning gains, value development unfolds across multiple contexts simultaneously, making it nearly impossible to isolate the teacher's contribution from the influence of family, peers, media, and broader societal forces.

Third, teachers reported insufficient professional training targeting value assessment design. Institutional VO workshops mostly focus textbook value mining, with zero sessions on formative evaluation tools such as value rubrics or peer evaluation scales.

4.2 Theme 2: Effective Multi-Source Formative Assessment Practices Identified by Teachers

Participants collectively recognized three mutually supportive multi-source assessment modes aligned with feedback theory:

While the majority of participants (17 out of 20) expressed enthusiasm for multi-source assessment approaches, three lecturers (Lecturers 2, 11, and 18) voiced skepticism about the feasibility of peer evaluation in the Chinese educational context. These participants argued that students may lack the critical distance to provide honest peer feedback on values due to relational harmony norms (*guanxi*), potentially compromising the reliability of peer assessment data. This contrasting perspective highlights an important boundary condition for the multi-source assessment model and suggests the need for cultural adaptation of peer feedback protocols.

4.2.1 Multi-Source Evaluation Strategies

Na described a triangulation approach that integrates multiple assessment sources to build a comprehensive picture of VOE effectiveness, reflecting Feedback Theory's multi-source feedback loop:

"I use a combination of teaching observations, student evaluations, and self-evaluations for a broad assessment of VOE effectiveness. Teaching observations allow colleagues to see VOE in practice, student evaluations capture student perceptions, and self-evaluation provides reflection on my own teaching. The combination of multiple perspectives provides a more complete picture than any single method." (Interviewee 1, Na)

4.2.2 Creative Task-Based Assessment

Teachers described using creative task-based assessment activities that make abstract values visible and assessable, embodying Feedback Theory's concept of assessment as learning. These included poster-making and fashion presentation tasks:

"I incorporate ideological elements through poster-making activities. The key is providing clear instructions and requiring value-based content so that students can visualize their ideas and beliefs. When students create posters about topics like cultural confidence or social responsibility, the posters themselves become assessable evidence of their understanding and expression of values." (Interviewee 7, Wang 1)

"I use fashion presentation tasks featuring traditional clothing such as Mongolian, Tibetan, Hanfu, and Qipao to strengthen national identity through language practice. Students research, prepare, and present about traditional clothing, which requires engaging with cultural content while developing language skills. The presentations become assessable evidence of both language learning and cultural understanding." (Interviewee 4, Si)

4.2.3 Simulation-Based and Process-Oriented Assessment

Beyond creative peer tasks, teachers employed simulation-based assessment and shared resource repositories to evaluate students' practical application of intercultural competence and ensure assessment consistency:

"I organize cross-cultural communication simulations featuring value conflicts between cultures, training students to seek consensus amid differences. These simulations require students to apply intercultural competence and values negotiation skills in practical contexts. The quality of their engagement in simulations provides evidence of their developing intercultural competence and value negotiation abilities." (Interviewee 16, Wu)

"When I design cross-cultural simulation tasks, I focus on observing whether students can negotiate value differences respectfully during role-play. I evaluate their ability to articulate their own cultural stance while showing openness to other perspectives. The key is not whether they give the 'right' answer, but how they navigate value conflicts in real-time communication." (Interviewee 8, Xu)

These diverse practices—multi-source evaluation, creative task-based peer assessment, and simulation-based assessment—all exemplify the core principles of formative assessment as conceptualized in Feedback Theory: shifting evaluation focus from one-time summative exams to ongoing process-oriented evaluation, with feedback loops that continuously inform and improve both teaching and learning.

4.3 Theme 3: Institutional and Pedagogical Barriers to Integrated Assessment

Four layers of constraints emerged from interview data:

Assessment system misalignment: University teaching evaluation systems prioritize language test performance, with no

formal assessment indicators for value-oriented education. Teaching evaluation forms primarily record CET pass rates, leaving VO assessment outside institutional quality assurance mechanisms. Zhou highlighted the need for clear evaluation frameworks:

“We need clear evaluation frameworks for VOE implementation. Currently, there’s uncertainty about what good VOE looks like and how it should be evaluated. Clear frameworks would give us standards to aim for and criteria for evaluation. Without such frameworks, implementation is uneven and evaluation is subjective.” (Interviewee 18, Zhou)

Professional development gaps: Institutional VOE training primarily focuses on textbook value mining, with insufficient attention to assessment tool design. Wang 1 observed this disconnect:

“The biggest challenge in assessing students’ ideological development is that there are no clear-cut criteria like in language testing. How do you score ‘cultural confidence’ or ‘critical thinking’? I often find myself relying on subjective impressions rather than systematic rubrics, which makes me uncomfortable.” (Interviewee 7, Wang 1)

“Sometimes students write exactly what they think I want to hear in their reflection journals, but I cannot tell if they truly internalize these values or are just performing for a grade. This authenticity problem makes me question the validity of any value assessment I design.” (Interviewee 11, Chen)

Practical operational constraints: Large class sizes (40–60 undergraduates per English class) make grading reflection journals and reviewing peer evaluation feedback extremely time-consuming for individual instructors. Na expressed the need for specific, practical support:

“With 50 students in one class, it is nearly impossible to give individualized feedback on their value development. I lack a standardized rubric to efficiently assess reflective journals, so I end up giving everyone the same participation grade, which defeats the purpose of formative assessment.” (Interviewee 1, Na)

Fragmented institutional support: Current institutional support for VOE tends to be episodic rather than sustained. Ji emphasized the need for systematic, long-term commitment:

“The current support for VOE tends to be episodic—training sessions, occasional meetings, one-time initiatives. What we need is sustained, systematic support that provides ongoing resources, continuous professional development, and long-term institutional commitment. Episodic support is insufficient for developing the expertise needed for effective VOE implementation.” (Interviewee 14, Ji)

5. Discussion

5.1 Theoretical Contributions

The core contribution of this study is the extension of Feedback Theory (Nicol & Macfarlane-Dick, 2006) into the Chinese VOE context. By providing qualitative evidence suggesting that multi-source formative assessment (self + peer + teacher) may offer a promising approach to the subjectivity dilemma of value evaluation, this study supplements Nicol and Macfarlane-Dick’s (2006) student feedback model with evidence from EFL teacher practice, demonstrating that feedback loops in value education require institutional scaffolding beyond individual classroom implementation.

The integrated “Cognition–Practice–Assessment” framework, derived from Feedback Theory, links Teacher Cognition Theory and Sociocultural Theory as supporting frameworks. It explains how lecturers’ policy understanding (cognition) shapes their classroom value assessment operations (practice), while multi-source feedback (assessment) reversely optimizes their VOE instructional decision-making. This three-way cyclic model, grounded in Feedback Theory’s regulatory cycle, begins to address the gap in holistic assessment frameworks for value-oriented EFL education.

The CPA model integrates both theoretically derived and empirically constructed components. The Cognition dimension was theoretically derived from Teacher Cognition Theory (Borg, 2003); however, the specific content of this dimension in the VOE context—teachers’ policy understanding of value education and their perceived self-efficacy in value assessment—emerged inductively from the interview data. The Practice dimension was theoretically informed by Sociocultural Theory (Vygotsky, 1978); the specific assessment practices identified—creative task-based assessment, simulation-based assessment, and multi-source evaluation—were empirically constructed from the interview data. The Assessment (Feedback) dimension was theoretically grounded in Feedback Theory (Nicol & Macfarlane-Dick, 2006); the multi-source feedback mechanisms and the specific institutional constraints that shape feedback loops were empirically identified from the interview data. The Instructional Adjustment dimension emerged entirely from the empirical data as participants described iterative cycles of assessment-informed teaching modification, representing the study’s unique empirical contribution.

The CPA framework differs from Nicol and Macfarlane-Dick’s (2006) original model in three important respects. First, whereas the original model focuses on individual student self-regulation through feedback, the CPA model operates at the collective level of teacher cognition and institutional practice, integrating multiple theoretical perspectives (Teacher Cognition Theory and Sociocultural Theory) absent from the original feedback model. Second, the CPA model incorporates the specific challenge of assessing value development—an affective-cognitive construct with no parallel in the original model’s focus on academic performance. Third, the CPA model explicitly addresses institutional contextual factors (assessment policies, class size, professional development structures) as mediating variables that shape the feedback cycle, whereas the original model treats institutional context as largely external to the feedback process.

Empirical qualitative data confirm the gap between national VO policy requirements and university assessment institutional design, providing micro-teacher-level evidence for prior macro theoretical arguments about the disconnect between policy mandates and assessment practice in Chinese higher education.

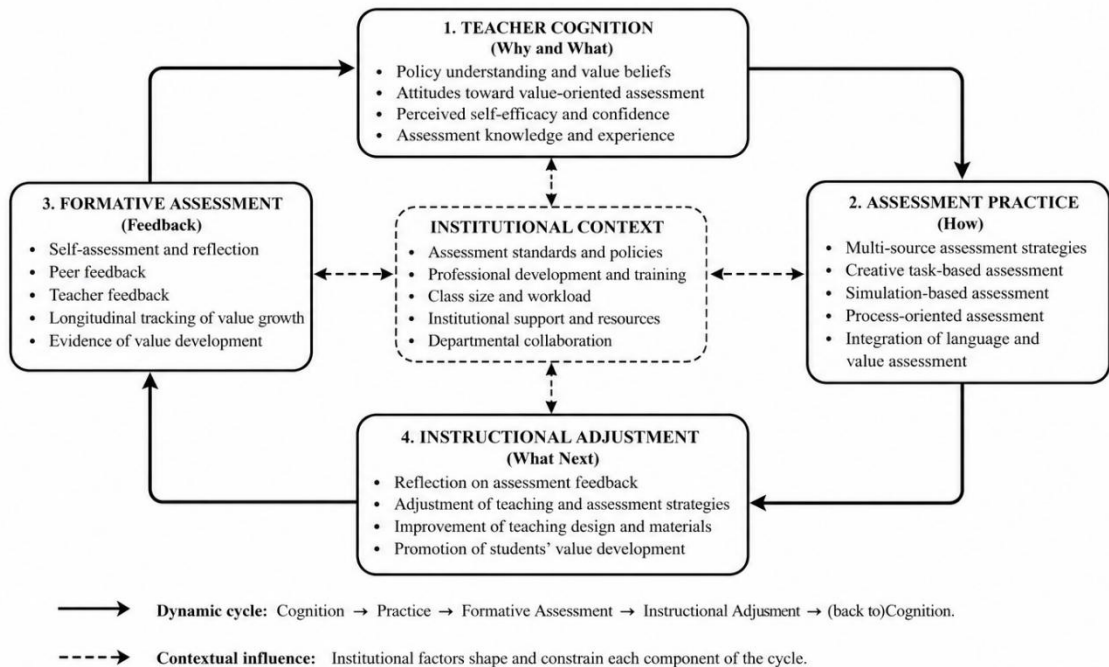
5.1.1 Addressing Core Assessment Challenges in VOE

The findings reveal two underexplored dimensions of value assessment that merit deeper theoretical attention. First, the

problem of performative compliance: several participants reported that students sometimes produce formulaic expressions of correct values to achieve higher assessment scores, raising questions about the authenticity of assessed value development. This phenomenon parallels social desirability bias in psychological measurement and suggests that value assessment instruments must incorporate mechanisms to distinguish genuine internalization from performative conformity. Potential strategies include open-ended reflective tasks, longitudinal observation, and peer verification, all of which align with Feedback Theory’s emphasis on multi-source evidence.

Second, the uncoupling of language proficiency and value expression presents a unique challenge in EFL contexts. Students with strong English skills may produce linguistically sophisticated but substantively empty value statements, while students with weaker language skills may demonstrate deeper value understanding through simpler expressions. Teachers reported difficulty in designing rubrics that fairly assess value development without confounding it with language ability. This tension suggests the need for dual-track assessment instruments that separately evaluate linguistic competence and value engagement, a design principle that has not been adequately addressed in existing VOE assessment literature.

Figure 2
Integrated Cognition–Practice–Assessment Framework for Value-Oriented College English Assessment



Note. This framework illustrates the reciprocal relationship among teachers’ cognition, assessment practice, formative assessment, and instructional adjustment within institutional contexts in college English classrooms.

Figure 2. Integrated Cognition–Practice–Assessment Framework for Value-Oriented College English Assessment

Grounded in the empirical narratives of 20 frontline college English lecturers, Figure 2 operationalizes Figure 1’s conceptual framework into an integrated “Cognition–Practice–Assessment” model embedded within concrete institutional contexts. While Figure 1 outlined the initial theoretical assumptions, Figure 2 incorporates empirical dimensions identified in the findings—such as institutional evaluation constraints, class size challenges, and multi-source feedback mechanisms—demonstrating how value assessment operates in real-world tertiary EFL classrooms.

5.2 Practical Pedagogical Implications

For College English Teachers

Adopt combined multi-source formative assessment: Weekly value reflection journals + group peer rubric evaluation + semester classroom observation portfolios, decouple value evaluation from single final language exam scores.

Develop simplified 4-level value assessment rubrics (Emerging → Developing → Proficient → Exemplary) with observable behavioral indicators for each dimension (cultural identity, critical thinking, social responsibility, intercultural awareness).

For University Curriculum Administrators

Provide a standardized Value Assessment Rubric Template for all college English instructors to reduce individual workload and ensure assessment consistency.

Establish annual review mechanisms that include value cultivation assessment portfolios alongside traditional language test performance metrics.

5.3 Comparison with Prior Research

Domestic existing VOE assessment research stays at theoretical index construction without teacher practice data (Li, 2022); this paper uses interview narratives to identify implementable formative assessment routines. International comparative studies on moral education have examined teachers’ perceptions across different cultural contexts (Asif et al., 2020), yet remain incompatible with China’s specific curriculum value education policy background. This study’s multi-

source formative system offers contextualized teacher perspectives that may inform formative assessment design in similar Chinese tertiary EFL settings.

5.4 Study Limitations & Future Research Directions

Limitations: Single regional sample (Inner Mongolia); only interview data without classroom observation or student assessment artifact triangulation. Future research:

Mixed-method design combining teacher interviews, classroom video observation, and student assessment portfolio analysis;

Cross-provincial multi-university sampling to improve result generalizability;

Develop quantitative value assessment scales based on this study's thematic findings for large-scale questionnaire verification.

6. Conclusion

Against the policy background of mandatory value-oriented education in Chinese College English, summative language-dominated evaluation systems cannot adequately capture students' long-term value growth. This qualitative interview study of 20 tertiary EFL lecturers yields three core conclusions:

The biggest assessment obstacles stem from the inherent subjectivity of value development and the lack of institutional unified formative evaluation standards.

Three complementary multi-source formative assessment modes (student self-reflection journals, peer rubric evaluation, teacher longitudinal classroom portfolios) are recognized by frontline teachers as feasible strategies to balance value tracking and linguistic assessment.

Institutional appraisal bias toward language test outcomes and large-class teaching pressure are the primary structural barriers preventing widespread implementation of integrated value-language evaluation.

The integrated "Cognition–Practice–Assessment" theoretical framework proposed in this study systematically connects teachers' VOE cognition, sociocultural value teaching activities, and multi-source feedback assessment cycles. Practically, the formative assessment routines summarized from teacher narratives offer potentially transferable assessment strategies that may inform practice in comparable tertiary EFL contexts. Universities should revise curriculum evaluation mechanisms to prioritize process-oriented value tracking rather than overemphasizing one-time language examinations, realizing the dual teaching goals of linguistic competence and sound value cultivation in College English courses.

References

- [1]Asif, T., Ouyang, G., Haider, M. A., Colomer, J., Kayani, S., & Amin, N. u. (2020). Moral education for sustainable development: Comparison of university teachers' perceptions in China and Pakistan. *Sustainability*, 12(7), 3014. <https://doi.org/10.3390/su12073014>
- [2]Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36(2), 81–109. <https://doi.org/10.1017/S0261444803001903>
- [3]Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- [4]Chen, L. (2022). Cultural aphasia in Chinese college English teaching (in Chinese). *Journal of Chinese Language and Culture*, 11(1), 45–58.
- [5]LePage, P., Akar, H., Temli, Y., Hasser, N., & Ivins, I. (2011). Comparing teachers' views on morality and moral education: A comparative study in Turkey and the United States. *Teaching and Teacher Education*, 27(2), 366–375. <https://doi.org/10.1016/j.tate.2010.09.003>
- [6]Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.
- [7]Liu, Z. G., & Yue, M. M. (2020). Resetting the educational philosophy and restructuring learning content to implement morality cultivation in foreign language course instruction (in Chinese). *Journal of Foreign Languages*, 43(5), 21–28.
- [8]Ministry of Education of China. (2020). *Guiding outline for curriculum-based ideological and political education in higher education (Doc. No. [2020]2) (in Chinese)*. http://www.moe.gov.cn/srcsite/A08/s7056/202006/t20200603_462437.html
- [9]Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- [10]Pan, H., & Yuan, Y. (2021). Practical exploration of ideological and political education in college foreign language courses: Problems and solutions (in Chinese). *Shandong Foreign Language Teaching*, 43(3), 53–62. <https://doi.org/10.16482/j.sdwy37-1026.2021-03-005>
- [11]Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- [12]Zhang, L., & Wang, Q. (2020). Evaluation index system for ideological English teaching. *Journal of Foreign Languages and Literature*, 42(6), 88–94.
- [13]Li, M. (2022). Research on the evaluation system of ideological and political education in college English courses (in Chinese). *Foreign Language World*, 2022(3), 45–56.
- [14]Campbell, M., Campbell, J., & Fullarton, S. (2013). Using inter-coder agreement to validate content analysis in

