

Operationalizing Learner Well-Being in ELT Textbooks: A Qualitative Study of Expert Reasoning and Methodological Constraints

Musa Nushi (موسی نوشی)

Shahid Beheshti University, Tehran, Iran (دانشگاه شهید بهشتی، تهران، ایران)

m_nushi@sbu.ac.ir

<https://orcid.org/0000-0003-1917-5372>

Corresponding author: Mehran Hosseinkhani (مهران حسین خانی)

Shahid Beheshti University, Tehran, Iran (دانشگاه شهید بهشتی، تهران، ایران)

mehranhxani@gmail.com

<https://orcid.org/0009-0002-4018-2235>

Abstract

This study investigated how experts in applied linguistics and positive psychology (PP) reason about the feasibility and limits of operationalizing complex psychological learner well-being constructs for English language teaching (ELT) textbook evaluation. While PP models such as EMPATHICS have substantially enriched theoretical understandings of learner well-being, their application to textbook evaluation raises unresolved methodological questions. Drawing on a qualitative, interpretive research design, the study foregrounded expert reasoning itself as the primary object of inquiry rather than treating expert consultation as a procedural step toward consensus or instrument validation. Data were collected through a recursive expert consultation process involving specialists in PP, English language education, and ELT materials development, and were analyzed using reflexive thematic analysis. The findings showed that experts differ systematically in their judgments of construct observability, the role of contextual mediation, and the legitimacy of inferential claims based on textbook content. These disagreements were not attributable to technical deficiencies or lack of expertise, but to deeper epistemological

orientations concerning evidence, pedagogy, and the relationship between texts and lived learning experience. By making visible the interpretive work underlying feasibility judgments, the study contributes to methodological debates on discourse- and materials-based analysis and argues for a more reflexive approach to applying holistic psychological models in ELT materials research.

Keywords: EMPATHICS; English language teaching textbooks; expert judgment; positive psychology; textbook evaluation

Introduction

Over the past two decades, positive psychology (PP) has gained increasing prominence in the field of second language acquisition (SLA). This shift reflects a growing recognition that language learning is not only a cognitive process but also an affective, motivational, and identity-shaping experience (MacIntyre et al. 2019). Research informed by PP has examined a wide range of constructs including enjoyment, resilience, empathy, motivation, and learner well-being, demonstrating their relevance for sustained engagement and successful language development (Wang et al. 2021). This shift has been further reinforced by the emergence of second-wave and third-wave PP, which emphasize contextual embeddedness, the coexistence of positive and negative emotions, and the complex, non-linear nature of human functioning (Lomas et al. 2020). Within this landscape, comprehensive theoretical models such as Oxford's EMPATHICS model (2016) have been proposed to capture the multifaceted psychological dimensions that shape language learning experiences.

Alongside these theoretical developments, English language teaching (ELT) textbooks remain a central component of instructional practice worldwide. For many teachers and learners,

textbooks function as de facto curricula, shaping not only linguistic content but also pedagogical priorities, classroom routines, and representations of what counts as legitimate learning (Richards 2014). As a result, textbook evaluation has long been a key methodological concern in applied linguistics, often grounded in theoretical frameworks that assess the effectiveness of ELT textbooks in promoting target language proficiency. Accordingly, studies have examined ELT textbooks from different perspectives such as grammatical progression, communicative authenticity, intercultural representation, and learner autonomy (Vitta 2021).

Despite the growing recognition of learner well-being as a central concern in language education, its implications for ELT materials evaluation remain insufficiently theorized and methodologically underexplored. Given that textbooks often function as de facto curricula—particularly in contexts where teachers rely heavily on published materials—what textbooks prioritize, omit, or normalize carries practical consequences for classroom experience and learner engagement (Richards 2014). Similarly, in the absence of explicit attention to psychological constructs in evaluation frameworks, materials research may remain largely confined to linguistic and methodological criteria, thereby offering a partial account of how materials shape learners' classroom experience. At the same time, existing PP models such as EMPATHICS (Oxford 2016) were developed as holistic theoretical visions rather than as tools for materials analysis, raising unresolved questions about how, and to what extent, well-being constructs can be meaningfully inferred from textbook content. Addressing this gap, the present study contributes to both theory and practice by examining how experts reason about the feasibility and limits of operationalizing learner well-being constructs for ELT textbook evaluation. By foregrounding expert judgment as an object of inquiry, the study provides a reflexive foundation for future

materials research that seeks to engage with learner well-being without oversimplifying its complexity.

Literature Review

Positive Psychology in SLA: From Constructs to Models

The incorporation of PP into SLA and language education emerged largely as a response to the field's long-standing emphasis on deficit-oriented constructs such as anxiety, demotivation, and boredom (Dewaele et al. 2019). Early PP-oriented work in SLA foregrounded constructs such as enjoyment, resilience, hope, and motivation, arguing that positive emotional and psychological experiences play a central role in sustained engagement and long-term language development (MacIntyre et al. 2019). Over time, this line of research moved beyond the examination of isolated variables toward more integrative perspectives that conceptualize learner well-being as dynamic, contextual, and socially embedded (Lomas et al. 2020).

This conceptual shift parallels developments in PP, particularly the movement from first-wave PP, which emphasized individual strengths and positive affect, to later waves that highlight complexity, the coexistence of positive and negative experiences, and the role of sociocultural context (Lomas et al. 2020). Within this broader trajectory, Oxford (2016) proposed the EMPATHICS model as a comprehensive model intended to capture multiple psychological dimensions relevant to language learning, including emotion and empathy, meaning and motivation, perseverance, agency and autonomy, time perspective, hardiness and habits of mind, intelligences, character strengths, and self-factors. As emphasized in the model's original formulation, EMPATHICS was designed as a holistic theoretical model rather than as a measurement instrument (Oxford 2016).

One of the most widely cited models in PP is Seligman's (2011) PERMA model, which conceptualizes individual well-being through five core elements: positive emotion, engagement, relationships, meaning, and accomplishment. While PERMA has been influential in shaping early PP-oriented research, its applicability to educational and language-learning contexts has been questioned. In particular, Oxford (2016) problematized Seligman's assumption that these dimensions operate as independent and non-overlapping components, arguing instead that psychological experiences are inherently interconnected and embedded within broader sociocultural contexts. From this perspective, isolating discrete dimensions of well-being risks oversimplifying the complex and relational nature of human learning and interaction. Oxford (2016) further noted that although PERMA is grounded in a set of character strengths, the model does not clearly articulate how these strengths map onto, or interact with, the individual PERMA dimensions. In addition, several psychological constructs that have been shown to be especially salient in language learning, such as perseverance, learner agency, and self-factors, are not explicitly accounted for within the PERMA model. In response to these conceptual and structural limitations, Oxford proposed the EMPATHICS model, a more comprehensive model designed to capture a wider range of psychological dimensions relevant to educational settings.

While PP models such as EMPATHICS have significantly enriched conceptual discussions of learner well-being in SLA by foregrounding its complexity and multidimensionality, their translation into empirical research remains challenging. As Alrabai and Dewaele (2023) demonstrate, the absence of clear operationalization, extensive construct overlap, and the theoretically interwoven nature of dimensions render direct measurement problematic. These constructs do not operate as discrete empirical entities but require interpretive analytical decisions before they can be rendered researchable, complicating efforts at

validation and instrument development. This challenge may become especially salient when PP models are applied beyond their original scope, such as in the analysis of instructional textbooks, where psychological processes must be inferred from static textual and pedagogical representations rather than from lived learner experience.

English Language Teaching Textbooks as Objects of Analysis

Textbooks occupy a central position in language education research and practice, functioning not only as sources of linguistic input but also as pedagogical scripts that shape classroom interaction, learning trajectories, and representations of language use (Cunningsworth 1995; Richards 2014). Because of this central role, textbook evaluation has long constituted a major strand of applied linguistics research, with established traditions examining linguistic coverage, skills integration, and methodological orientation (Vitta 2021).

Classic textbook-evaluation models emphasize criteria such as authenticity, communicative value, progression, and pedagogical usability, reflecting concerns with both linguistic adequacy and instructional effectiveness (e.g., Littlejohn 1998). More recent approaches have expanded this focus to include intercultural representations (Nushi and Aghaei 2024), learner-autonomy (Reinders and Balçıkanlı 2011), and learner-centered pedagogy, acknowledging the broader social and ideological dimensions of language education (Tomlinson 2011). Within this tradition, textbooks are generally treated as analyzable pedagogical artefacts whose content can be systematically examined using predefined analytical frameworks and evaluation criteria. The extension of textbook evaluation into affective and psychological domains, however, raises important conceptual and methodological questions. Unlike grammatical structures or task types, psychological constructs are not discrete textual features

but must be inferred from patterns of representation, task design, thematic framing, and pedagogical intent.

Expert Judgment and Consultation in Instrument Development

Expert judgment has long played a central role in educational research, particularly in the development and validation of research instruments. During early stages of instrument design, experts are commonly invited to evaluate the relevance, clarity, and representativeness of constructs, thereby contributing to content validity and theoretical alignment. Methodological approaches such as expert panels, Delphi studies, and Q-methodology explicitly rely on expert perspectives as a primary source of knowledge, especially when constructs are abstract, multidimensional, or context-sensitive (Brown 1980; Waltz et al. 2015). Despite its importance, expert consultation in most empirical studies functions largely as a procedural step rather than as an object of inquiry in its own right. Disagreement among experts is typically managed through consensus-building techniques, or exclusion of outlier views (Leising et al. 2024), and published reports tend to foreground only the final outcomes of these processes. As a result, the reasoning processes underlying expert judgments, including why experts disagree, what epistemic assumptions they bring to the task, and how they conceptualize feasibility and evidence, are rarely examined or theorized explicitly.

Taken together, the literature on PP and ELT textbook evaluation reveals a persistent gap. While theoretical models of learner well-being have become increasingly sophisticated, and while textbooks are widely analyzed as pedagogical artefacts, relatively little attention has been paid to the conceptual work required to connect these domains. In particular, there is no research that examines how researchers and experts negotiate the feasibility of operationalizing holistic psychological constructs for textbook evaluation. Addressing this gap, the present study

contributes to a more reflexive and theoretically grounded approach to materials research in applied linguistics. Rather than asking whether textbooks contain specific PP dimensions, it examines how experts reason about the very possibility of identifying such dimensions, thereby making visible a layer of methodological reasoning that is typically left implicit. Thus, the present study is guided by the following question:

How do experts from positive psychology, English language education, and ELT materials development reason about the feasibility and limits of operationalizing positive psychology constructs for ELT textbook evaluation?

Methodology

Research Design

This study adopts a qualitative, interpretive research design to examine how experts reason about the feasibility and limits of operationalizing PP constructs for ELT textbook evaluation. Rather than treating expert consultation as a procedural step toward instrument validation, the present study conceptualizes expert reasoning itself as the primary object of inquiry. This design is consistent with methodological approaches that emphasize reflexivity and interpretive judgment in construct exploration, particularly when dealing with abstract, multidimensional, and context-sensitive psychological constructs (Creswell et al. 2006; Wiesner 2022). By foregrounding this design, the study seeks to make visible the epistemic assumptions, points of convergence and divergence, and evaluative criteria experts employ when assessing whether PP dimensions can be meaningfully identified in textbook content.

Participants

Participants were selected through purposive sampling to ensure conceptual relevance and informed judgment regarding the operationalization of learner well-being in ELT textbooks. Selection criteria included: (a) possession of a doctoral degree in a relevant field; (b) an established record of scholarly engagement with PP, English language education, or ELT materials development, as evidenced by publications, teaching experience, or materials-related work; and (c) professional experience with ELT textbooks, either through direct teaching, evaluation, or materials development.

The expert panel comprised specialists in PP ($n = 3$), English language education ($n = 5$), and ELT materials development ($n = 4$). All participants held doctoral degrees and had between 8 and 25 years of professional experience. The materials-development experts had direct involvement in the design, adaptation, teaching or evaluation of ELT textbooks, while the PP and English language education specialists had published peer-reviewed research on learner well-being, affective factors, or socio-emotional dimensions of language learning. Furthermore, all participants had professional experience teaching English in higher education or ELT contexts. Ten of the twelve participants reported prior experience teaching or evaluating the target textbook series. The remaining two participants had not previously worked with these textbooks; however, they were systematically introduced to the materials prior to data collection and were provided with sufficient time to review relevant units, tasks, and design features before offering evaluative commentary. Participants ranged in age from 37 to 62 years and included both male ($n = 8$) and female ($n = 4$) academics. This diversity in disciplinary expertise, professional role, and career stage was intended to capture a wide range of epistemic perspectives, thereby strengthening the analytical depth and interpretive credibility of the study. All participants

provided informed consent prior to participation. Anonymity was ensured through the use of numerical identifiers, and all data were stored securely.

Table 1: Demographics

Participant	Primary Area of Expertise	Years of Experience	Gender
P1	PP	8	Male
P2	English Language Education	14	Female
P3	ELT Materials Development	11	Male
P4	English Language Education	25	Male
P5	ELT Materials Development	12	Male
P6	PP	8	Female
P7	English Language Education	16	Male
P8	ELT Materials Development	15	Female
P9	English Language Education	11	Male
P10	PP	9	Female
P11	ELT Materials Development	12	Male
P12	English Language Education	17	Male

Data collection

Data were collected through a recursive expert consultation process combining semi-structured interviews and structured written feedback. The written feedback took the form of asynchronous

email-based responses, and it was designed to elicit experts' analytic reasoning about the feasibility of operationalizing learner well-being constructs for ELT textbook evaluation. Specifically, experts were asked to provide written commentary on (a) the theoretical definitions of the EMPATHICS dimensions as articulated by Oxford (2016), and (b) the extent to which these dimensions could plausibly be inferred from textbook content, such as task design, thematic framing, or pedagogical prompts. Experts were invited to comment on which dimensions they considered more or less observable in textbooks, what types of textual or visual features might serve as indicators, and what methodological challenges or inferential limits they perceived.

This written feedback complemented the interviews by allowing participants time for reflective, theory-informed responses and by supporting the iterative refinement of follow-up prompts in subsequent consultation rounds. All interviews were conducted in participants' L1 (Persian) to minimize linguistic constraints and to facilitate precise articulation of complex theoretical and psychological constructs. Interviews lasted between 35 and 45 minutes, were audio-recorded, transcribed, and subsequently translated into English. This multi-modal approach allowed experts to articulate their perspectives in depth and to revisit earlier positions as the consultation progressed.

At the outset, experts were introduced to the theoretical definitions of the nine EMPATHICS dimensions as articulated by Oxford (2016). These definitions were accompanied by the researchers' initial, research-informed reflections on how each dimension might plausibly manifest in ELT textbook content. These reflections were explicitly framed as tentative and exploratory, serving only as prompts for expert evaluation rather than as analytic claims. Experts were then invited to respond to three guiding prompts:

1. To what extent do you consider each EMPATHICS dimension observable in ELT textbooks?
2. What types of textual or visual features, if any, could plausibly indicate the presence of these dimensions?
3. What challenges do you anticipate in reliably identifying these constructs across different textbook series?

These prompts were designed to elicit expert reasoning about observability, inferential boundaries, and methodological feasibility, rather than to solicit evaluative judgments about specific textbooks. Importantly, the consultation was not treated as a single, static elicitation of expert opinion. Instead, it functioned as a dialogic and iterative process. Experts were given opportunities to reflect on and, where appropriate, revise their responses after engaging with synthesized summaries of emerging points across the panel. This recursive structure supported deeper reflection and allowed underlying assumptions to surface more explicitly.

Data Analysis

Data analysis followed a reflexive thematic analysis approach (Braun & Clarke, 2006, 2019) and proceeded through several iterative stages. First, all interview transcripts and written feedback responses were read repeatedly to achieve familiarization with the data. Second, initial codes were generated inductively, focusing on segments in which experts articulated their reasoning about (a) the observability of EMPATHICS dimensions in textbook content, (b) the types of textual or pedagogical evidence they considered legitimate, and (c) perceived methodological constraints on observing psychological constructs in textbooks. Through iterative comparison, these codes were clustered into higher-order themes reflecting shared epistemic positions, points

of disagreement, and underlying evaluative criteria. These themes were then reviewed, refined, and redefined through repeated engagement with the full dataset to ensure internal coherence and clear distinction between themes.

Analytic attention was directed not only to areas of convergence but also to moments of hesitation, contradiction, and disagreement, which were treated as analytically productive rather than as methodological problems. Throughout the analysis, reflexive memoing was used to document interpretive decisions and to bracket the researcher's prior assumptions about textbook evaluation and PP models.

To ensure reliability of coding procedure, analytic interpretations (i.e., the researchers' thematic understandings of experts' responses) were grounded in two data sources, namely semi-structured interviews and written feedback, and were subjected to repeated review to ensure consistency and transparency. Rather than seeking reliability in the positivist sense, coding and analysis were undertaken collaboratively, with the two researchers independently coding an initial subset of the data to enhance rigor. Differences in coding were resolved through iterative and reflexive discussions, facilitating critical examination of assumptions, consensus on thematic delineation, and the co-construction of a richer understanding of the data.

Although the study was guided by the EMPATHICS model (Oxford, 2016), it was not used as a coding scheme or set of predefined analytic categories. Rather, EMPATHICS functioned as a conceptual lens that sensitized the analysis to issues of learner well-being and operationalization, while allowing themes to be developed inductively from the data. Thematic analysis was therefore not conducted from scratch in a purely inductive sense, nor was it deductive or framework-driven. Instead, the analysis adopted a theory-informed reflexive

thematic approach, focusing on how experts reasoned about the feasibility, evidentiary boundaries, and limits of applying EMPATHICS constructs to ELT textbook evaluation.

Results

Analysis of the data revealed three interrelated themes that capture how experts reasoned about the feasibility and limits of operationalizing PP constructs for ELT textbook evaluation. These themes reflect both areas of convergence and points of divergence in expert judgment, as well as underlying epistemic assumptions about textbooks, psychological constructs, and inference.

Differential Perceived Observability of EMPATHICS Dimensions

The expert consultation data indicated clear variation in participants' assessments of how observable different EMPATHICS dimensions are within ELT textbooks. Nine out of 12 experts reported relatively high confidence in identifying the presence of specific dimensions, particularly *emotion and empathy*, *meaning and motivation*, and *intelligences*. These dimensions were consistently associated with textbook features that participants described as explicit, task-embedded, or structurally visible.

Experts frequently referred to tasks that required learners to express personal opinions, feelings, or experiences as indicators of *emotion and empathy*. According to these participants, such tasks provided direct textual cues that could be identified without extensive inference. P2 stated, "When learners are asked to talk about their feelings, preferences, or personal experiences, I think emotion and empathy are reasonably visible at the task level" (Interview). Similar observations were made by other participants, who emphasized the presence of prompts inviting reflection, discussion, or personal response. Analytically, this position reflects a view of emotion and empathy as interactionally cued constructs that can be textually instantiated through

explicit prompts for personal response. Since such tasks explicitly position learners as emotional and experiential agents, experts considered them less dependent on contextual mediation and therefore more amenable to textbook-based identification. In EMPATHICS terms, the feasibility of operationalizing emotion and empathy was linked to the presence of overt discursive invitations for affective engagement, rather than to inferred psychological states.

The dimensions of *meaning and motivation* were also described as observable when textbook units included explicit goals, thematic coherence, or references to learning purposes. Participants noted that goal-oriented content, such as tasks framed around future use of language or personal relevance, provided identifiable indicators. As one of the participants explained, “Meaning and motivation are often built into units through themes, purposes, or end goals, especially when learners are encouraged to reflect on why they are doing an activity” (P7, written feedback). This reasoning aligns with EMPATHICS’ emphasis on meaning-making and motivational orientation as pedagogically scaffolded rather than internally assumed (Oxford, 2016). From this perspective, observability was attributed to the explicit articulation of purpose within the material, allowing motivation to be inferred from curricular framing.

Similarly, *intelligences* were discussed in relation to task design and modality. Experts pointed to variation in activity types such as visual, verbal, collaborative, or reflective tasks as potential indicators of this dimension. One participant commented, “Multiple intelligences can be inferred when tasks deliberately vary in modality... visual, verbal, collaborative, reflective... even if the term itself is never mentioned” (P9, interview). Here, observability was associated with structural diversity in task formats rather than explicit labeling.

In contrast, other EMPATHICS dimensions were described by several experts as less readily observable in textbook content. Constructs such as *time, agency and autonomy, hardiness*

and habits of mind, and *self-factors* generated more cautious responses. These dimensions were often characterized as internally oriented or context dependent, making them less directly accessible through textual features alone. One expert summarized this position by stating, “Some of these constructs are too internally situated. You can’t always see perseverance or self-belief just by looking at a unit” (P4, interview). Such comments reflect an EMPATHICS-informed distinction between psychological dimensions that can be prompted through explicit discourse and those that emerge through sustained engagement, self-regulation, and learner interpretation over time. In particular, experts viewed hardiness and habits of mind as well as self-factors as constructs that presuppose repeated challenge, feedback, and personal meaning-making—conditions that extend beyond what a single textbook task can transparently represent. Similar comments were made regarding *time* and *agency*, with experts noting that these dimensions may rely on instructional framing or classroom implementation rather than being explicitly represented in textbook tasks.

Altogether, experts differentiated between dimensions that could be associated with identifiable textual elements and those that were described as less visible due to their reliance on learner interpretation, sustained engagement, or contextual mediation. This differentiation was repeatedly articulated in participants’ descriptions of what they considered observable within the scope of textbook evaluation.

Context Dependence and Mediation as Constraints on Inference

The second theme concerned experts’ views on the role of context and mediation in determining whether psychological constructs could be inferred from ELT textbook content. Across the dataset, participants differed in how they assessed the relationship between textual features and

psychological dimensions, particularly with regard to the degree to which inference was warranted in the absence of classroom enactment.

The EMPATHICS dimension of *time* was repeatedly referenced as illustrative of this issue. Seven participants indicated that time-related orientations could be identified through tasks that prompted learners to engage with planning, goal setting, or future-oriented reflection. These participants described such tasks as providing at least indirect access to the construct. As one expert noted, “If students are asked to plan their study time or think about future goals, then ‘time’ is at least implicitly there” (P6, written feedback). Similar comments emphasized that references to deadlines, progress tracking, or future language use could serve as textual indicators.

However, other experts expressed reservations about treating such features as evidence of the *time* dimension, emphasizing that these activities often relied on instructional framing that was not explicitly specified in the textbook. One participant stated, “Those tasks only work if the teacher actually frames them that way. In many books, that mediation is assumed rather than spelled out” (P11, interview). Another expert similarly commented, “You might see a task about goals, but whether it really becomes about time management depends entirely on how it’s handled in class” (P5, interview).

This concern was not limited to the dimension of *time*. Participants extended the same reasoning to other constructs, noting that dimensions such as *agency and autonomy*, as well as *hardiness and habits of mind* were often contingent on how tasks were implemented rather than on the task text itself. P8 observed, “Agency only shows up if learners are given real choices and space to act. A textbook can suggest that, but it doesn’t control what actually happens” (Written

feedback). Another expert remarked, “Perseverance is something you see over time, not in a single activity. A unit alone doesn’t tell you much” (P3, interview).

These comments reflect a recurring expert position that some of EMPATHICS dimensions cannot be straightforwardly inferred from task prompts alone. From this perspective, textbooks often imply well-being–supportive practices (e.g., reflection, goal setting, or learner choice) without explicitly specifying how these practices should be pedagogically framed. As a result, the psychological affordances associated with EMPATHICS constructs remain contingent on teacher interpretation and enactment rather than being textually observable. This reliance on assumed mediation was repeatedly cited as a key limitation of textbook-based analysis, as it complicates claims about whether learner well-being is meaningfully represented in the material itself or merely presupposed at the level of classroom practice.

Across responses, participants frequently distinguished between what textbooks *offer* and what learners *experience*. Several experts emphasized that textbooks may include prompts that invite reflection or engagement, but that such invitations do not guarantee psychological engagement. As one expert stated, “The book can invite reflection, but it can’t guarantee that reflection happens. That’s a big limitation when we try to analyze well-being through materials alone” (P1, interview). Another participant similarly noted, “Textbooks create possibilities, not outcomes. That makes it tricky to say a construct is really there” (P10, written feedback). Some experts also highlighted the role of teachers as mediators between text and learner experience. P12 explained, “A lot of the psychological work happens in the teacher’s explanation or follow-up, not in the printed page” (Interview). Another added, “If we ignore mediation, we risk overestimating what the textbook itself is doing” (P4, written feedback).

Overall, participants' responses reflected differing positions on how much contextual mediation could be assumed when analyzing textbook content. While some experts accepted indirect or implicit cues as sufficient for identifying psychological dimensions, others maintained that reliance on mediation introduced uncertainty that limited the feasibility of inference. These positions were articulated through repeated references to classroom enactment, teacher framing, and the distinction between textual prompts and psychological processes.

Divergent Epistemic Orientations Toward Inference

The third theme concerned differences in experts' orientations toward inference in textbook evaluation, particularly regarding what types of evidence were considered sufficient for identifying PP dimensions in ELT textbooks. Across the dataset, experts articulated two broadly contrasting positions: a preference for restricting analysis to explicitly stated features, and an acceptance of inferred or implicit psychological themes embedded in textbook content.

Five out of 12 experts expressed a cautious approach to inference, emphasizing the importance of limiting analysis to overt textual signals. These participants described explicit instructions, clearly articulated task aims, or directly stated learner actions as the primary basis for identifying PP dimensions. One expert stated, "If a construct is not explicitly signaled in the task or instruction, I would hesitate to code it. Otherwise, we risk projecting our own interpretations onto the material" (P3, interview). Another participant similarly commented, "I think we should only work with what is clearly there on the page. Once we start guessing intentions, it becomes subjective" (P6, interview).

This conservative stance was often linked to concerns about consistency and replicability. Some experts noted that reliance on implicit cues could lead to variation in interpretation across

analysts. As one participant explained, “Two researchers might read the same task very differently if the criteria are not explicit” (P2, written feedback). Another added, “Without clear signals, it’s difficult to ensure that different coders would reach the same conclusion” (P10, interview). These comments position EMPATHICS constructs as legitimate analytic categories only insofar as they are rendered textually visible. From this perspective, psychological dimensions such as agency, perseverance, or motivation are considered methodologically valid indicators only when they are overtly encoded in the material, minimizing interpretive distance between text and construct. Furthermore, these concerns reflect an epistemological preference for minimizing subjectivity by privileging explicit textual evidence over inferred psychological meaning. Consequently, experts adopting this stance viewed the operationalization of PP dimensions in textbook evaluation as feasible only under narrowly defined evidentiary conditions.

In contrast, other experts (N=4) described an interpretive orientation toward textbook evaluation, viewing inference as an unavoidable aspect of analyzing educational texts. These participants argued that psychological dimensions are frequently communicated indirectly through themes, narratives, and task outcomes rather than through explicit labeling. One expert stated, “Textbooks communicate values and orientations implicitly all the time. Ignoring that level would mean missing how psychology actually works in educational texts” (P8, written feedback). Another participant remarked, “A lot of what matters psychologically is embedded in the storylines, topics, and situations, not in explicit instructions” (P7, interview).

Experts endorsing this position often referred to broader patterns within units rather than isolated tasks. One participant noted, “You may not see agency written as a word, but you see it in the kinds of choices learners are repeatedly invited to make” (P5, interview). Another

commented, “Psychological orientations are conveyed through recurring patterns of tasks, topics, and learner roles across units rather than through isolated activities” (P12, written feedback).

On the other hand, three participants acknowledged the tension between these orientations without aligning exclusively with one position. Some described adjusting their stance depending on the dimension under consideration. As one participant stated, “For some constructs, like emotion, I’m comfortable with inference. For others, like hardiness, I’m much more cautious” (P9, interview). Another expert similarly observed, “It really depends on how close the construct is to language use itself” (P4, written feedback). These responses underscore that feasibility judgments are not uniform even within individuals, but are dynamically calibrated to the perceived proximity of specific EMPATHICS dimensions to observable pedagogical practice. This conditional positioning further illustrates that disagreements about inference reflect nuanced epistemic reasoning rather than categorical methodological divides.

In sum, experts consistently framed their positions in terms of evidence boundaries rather than disagreement over the value of PP constructs themselves. Participants repeatedly referenced concerns about interpretive distance, analytic transparency, and the relationship between textual features and psychological meaning. These responses indicate that experts approached inference in textbook evaluation with differing assumptions about what constitutes acceptable evidence and how psychological constructs should be identified within instructional texts.

Discussion

The present study set out to examine how experts in applied linguistics and PP reason about the feasibility and limits of operationalizing PP constructs for ELT textbook evaluation. The findings illuminate three interrelated dimensions of this reasoning: perceived observability, contextual

mediation, and epistemic orientation toward inference. Together, these dimensions help explain why translating holistic PP models such as EMPATHICS into textbook evaluation remains methodologically challenging.

First, experts' differential assessments of the observability of EMPATHICS dimensions underscore that not all psychological constructs are perceived as equally amenable to textbook evaluation. Dimensions such as *emotion and empathy*, *meaning and motivation*, and *intelligences* were consistently associated with identifiable textual features, including personal expression tasks, thematic coherence, and task modality. These observations align with textbook research traditions that treat tasks, prompts, and unit organization as analyzable artefacts (Littlejohn 1998; Tomlinson 2011). At the same time, experts' hesitation regarding constructs such as *time*, *agency and autonomy*, *hardiness and habits of mind*, and *self-factors* reflects the limits of extending textbook evaluation to constructs that are internally situated, longitudinal, or interaction-dependent. This distinction highlights an important methodological implication: feasibility is not an inherent property of a theoretical model but emerges at the intersection of construct characteristics and analytic medium.

Second, the findings demonstrate that expert judgments were strongly shaped by assumptions about mediation and classroom enactment. Across responses, experts repeatedly distinguished between what textbooks *offer* and what learners *experience*, emphasizing that textbooks can at best indicate pedagogical affordances rather than instantiate psychological processes. This distinction resonates with long-standing cautions in SLA and materials research regarding the gap between designed pedagogy and enacted practice (Ellis 1997; Richards 2014). The dimension of *time* served as a focal example, with some experts accepting indirect textual cues as sufficient, while others questioned whether such cues could be meaningfully interpreted

without explicit instructional framing. These divergent positions suggest that operationalization decisions are inseparable from researchers' views on how much mediation can be assumed in materials-based studies.

Third, the study revealed contrasting epistemic orientations toward inference, ranging from conservative approaches that privilege explicit textual signals to interpretive approaches that accept implicit psychological meanings embedded across tasks and units. Importantly, these orientations were not framed by experts as disagreements about the value of PP, but as differences in evidentiary thresholds and analytic responsibility. Experts advocating conservative approaches emphasized consistency, transparency, and replicability, while those endorsing interpretive approaches emphasized the communicative and ideological nature of educational texts. This finding echoes broader methodological debates in applied linguistics regarding the legitimacy of inference in discourse-based and materials-based analysis (Taylor 2013). It also suggests that disagreements about feasibility often reflect deeper epistemological commitments rather than technical shortcomings.

Taken together, these findings help explain why PP-informed textbook studies frequently struggle with operational clarity. EMPATHICS, as Oxford (2016) emphasizes, was proposed as a holistic theoretical model rather than a measurement or evaluation tool. The present study shows that when such a model is brought into contact with textbook evaluation, experts must negotiate boundaries between theory and artefact, between affordance and enactment, and between explicit evidence and interpretive inference. Making these negotiations visible addresses a methodological gap identified in both PP-in-SLA research (Alrabai and Dewaele 2023) and textbook evaluation scholarship, where expert consultation is often treated as a procedural step rather than an analytic resource.

The study contributes to the literature in three main ways. First, it reframes expert consultation as a site of methodological knowledge rather than merely a validation mechanism. Second, it demonstrates that feasibility judgments are structured, patterned, and theoretically informed, rather than idiosyncratic. Third, it highlights that operationalization of these psychological construct for textbook evaluation is not a neutral translation process but a form of epistemic work shaped by assumptions about learning, materials, and evidence. These contributions are particularly relevant for researchers seeking to extend PP models into materials research without collapsing theoretical richness into overly reductive indicators.

The findings suggest that ELT textbooks should not be treated as self-contained carriers of learner well-being. Some of the psychological constructs were described by experts as highly dependent on pedagogical mediation. For teachers, this underscores the importance of instructional framing, task sequencing, and classroom interaction in activating the potential well-being affordances of textbook materials. Rather than assuming that psychological support is embedded in tasks by default, teachers may benefit from critically interpreting and adapting materials to align with learners' needs and classroom contexts.

From a learner perspective, the study highlights that experiences of motivation, agency, or emotional engagement are not guaranteed by textbook content alone. Learners' well-being is shaped by how tasks are enacted, interpreted, and sustained over time. Awareness of this variability may encourage greater learner reflection and agency in engaging with materials, particularly when learners are invited to make sense of task purposes and relevance rather than merely completing activities.

For teacher educators, the findings point to the need for explicit training in materials literacy and evaluative judgment. Given the variation in experts' interpretations and inferential

thresholds, pre-service and in-service teachers may benefit from learning how to apply well-being-oriented models critically and transparently in preparation of teaching materials. Teacher education programs can use these findings to move beyond checklist-based evaluation models and instead foster reflective engagement with psychological constructs in pedagogical materials.

The results indicate that learner well-being cannot be meaningfully embedded through isolated activities or implicit assumptions alone. Developers seeking to align materials with models such as EMPATHICS should consider how psychological dimensions are supported across units through coherent design logic, explicit pedagogical guidance, and opportunities for sustained learner engagement.

Although the present study is conceptual and methodological in nature, its findings have several practical implications for ELT professionals engaged in textbook evaluation, teacher education, and curriculum design. First, the observed variation in experts' judgments highlights the need for evaluators to adopt explicit and theory-grounded criteria when applying learner well-being models such as EMPATHICS. In practice, this suggests that claims about the presence of psychological dimensions in textbooks should be accompanied by transparent justification rather than assumed on the basis of intuitive interpretation. Second, the findings underscore the importance of professional development for teachers and materials evaluators. Given that several EMPATHICS dimensions were described as dependent on pedagogical mediation, teachers play a critical role in actualizing—or constraining—the well-being potential of textbook tasks. Teacher education programs may therefore benefit from training practitioners to critically interpret textbook affordances rather than treating materials as self-sufficient carriers of psychological support. Third, for textbook developers and curriculum planners, the results indicate that learner well-being cannot be meaningfully embedded through isolated activities

alone. Instead, dimensions such as agency, time orientation, and perseverance require coherent design logic across units and explicit pedagogical guidance. Awareness of these design demands can help developers avoid superficial inclusion of PP concepts and promote more principled integration. Finally, at a broader level, the study cautions against the uncritical adoption of learner well-being models in evaluative contexts. By foregrounding expert disagreement and inferential limits, the findings encourage practitioners to treat well-being-oriented evaluation as a reflective and negotiated process rather than a checklist-driven procedure.

This study contains some limitations that should be acknowledged. First, it adopts a qualitative, interpretive design focused on expert reasoning rather than empirical validation of learner outcomes, which limits the generalizability of the findings beyond similar research-design and textbook evaluation contexts. Second, expert judgments are necessarily shaped by participants' disciplinary backgrounds and epistemological orientations, meaning that alternative expert communities might reason differently about feasibility and inference. Third, because the analysis centers on reasoning about operationalization rather than classroom enactment or learner experience, the study does not claim that the identified constructs can or cannot be realized pedagogically in practice. Finally, the focus on textbook evaluation may underrepresent interactional and contextual dimensions of learning that emerge only through classroom-based or longitudinal research.

Conclusion

The present study aimed to explore how experts reason about the feasibility of operationalizing complex psychological constructs, particularly those associated with PP and learner well-being, within textbook evaluation. Rather than treating expert judgment as a neutral or purely technical procedure, the study foregrounded it as an interpretive practice shaped by epistemological

assumptions about evidence, context, and learning. The findings suggest that disagreements among experts are not best understood as methodological failure or lack of clarity, but as reflections of deeper theoretical commitments regarding what can legitimately be inferred from pedagogical artefacts such as textbooks. In this sense, feasibility emerged not as a fixed property of constructs themselves, but as a judgment negotiated at the intersection of theory, method, and analytic purpose.

Several directions for future research follow from these conclusions. First, comparative studies could examine how expert reasoning differs across epistemological traditions (e.g., sociocultural and critical), thereby mapping how feasibility judgments vary systematically rather than idiosyncratically. Such work would further clarify how methodological disagreements are rooted in theoretical orientations. Second, future research could extend the present focus on expert reasoning by triangulating it with classroom-based or learner-centered data, exploring how experts' assumptions about inference align, or clash, with pedagogical enactment and learner experience. This would help bridge the persistent gap between textbook evaluation and interactional research. Third, similar reflexive approaches could be applied to other domains of applied linguistics, such as assessment or curriculum development, where complex constructs are frequently operationalized with limited scrutiny of underlying assumptions. Finally, future studies might experiment with methodological designs that intentionally preserve, rather than resolve, expert disagreement, treating divergence as analytic data that illuminates the conceptual boundaries of constructs.

Funding

The authors state that the study was conducted without any external financial support.

Authors' Contribution

Both authors contributed equally to this work.

Disclosure of interest

The authors report no conflicts of interest.

Acknowledging the use of Generative AI

In preparing this essay authors used ChatGPT-5.2 in order to improve the manuscript's linguistic clarity and overall language quality. The authors take full responsibility for the content of the publication.

Acknowledgements

The authors sincerely thank the participants for their time and engagement, and their colleagues for ongoing support during the research process. They also appreciate the editor and anonymous reviewers for their valuable and constructive feedback.

References

- Arabai, Fakieh, and Jean-Marc Dewaele. 2023. "Transforming the EMPATHICS Model Into a Workable E4MC Model of Language Learner Well-Being". *The Journal for the Psychology of Language Learning* 5 (1):1-14.
<https://jpll.org/index.php/journal/article/view/167>.
- Braun, Virginia, and Victoria Clarke. 2019. "Reflecting on Reflexive Thematic Analysis." *Qualitative Research in Sport, Exercise and Health* 11 (4): 589–597.
<https://doi.org/10.1080/2159676X.2019.1628806>
- Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101.
<https://doi.org/10.1191/1478088706qp063oa>

- Brown, Steven R. 1980. *Political Subjectivity: Applications of Q Methodology in Political Science*. New Haven, CT: Yale University Press.
- Creswell, John W., Ron Shope, Vicki L. Plano Clark, and Denise O. Green. 2006. "How Interpretive Qualitative Research Extends Mixed Methods Research." *Research in the Schools* 13 (1): 1–11.
- Cunningsworth, Alan. 1995. *Choosing Your Textbook*. London: Macmillan Heinemann.
- Dewaele, Jean-Marc, Xinjie Chen, Amado M. Padilla, and James Lake. 2019. "The Flowering of Positive Psychology in Foreign Language Teaching and Acquisition Research." *Frontiers in Psychology* 10: 2128. <https://doi.org/10.3389/fpsyg.2019.02128>
- Ellis, Rod. 1997. "The Empirical Evaluation of Language Teaching Materials." *ELT Journal* 51 (1): 36–42. <https://doi.org/10.1093/elt/51.1.36>
- Leising, Daniel, Heinrich Liesefeld, Susanne Buecker, Andreas Glöckner, and Stefanie Lortsch. 2024. "A Tentative Roadmap for Consensus Building Processes." *Personality Science* 5. <https://doi.org/10.1177/27000710241298610>
- Littlejohn, Andrew. 1998. "The Analysis of Language Teaching Materials: Inside the Trojan Horse." In *Materials Development in Language Teaching*, edited by Brian Tomlinson, 190–216. Cambridge: Cambridge University Press.
- Lomas, Tim, Lea Waters, Paige Williams, Lindsay G. Oades, and Margaret L. Kern. 2020. "Third Wave Positive Psychology: Broadening towards Complexity." *The Journal of Positive Psychology* 16 (9): 1–12. <https://doi.org/10.1080/17439760.2020.1805501>

- MacIntyre, Peter D., Tammy Gregersen, and Sarah Mercer. 2019. "Setting an Agenda for Positive Psychology in SLA: Theory, Practice, and Research." *The Modern Language Journal* 103 (1): 262–274. <https://doi.org/10.1111/modl.12544>
- Nushi, Musa, and Alireza Aghaei. 2024. "Intercultural Representation in ELT Textbooks: The Case of *Touchstone* Series." *MEXTESOL Journal* 48, no. 1. <https://doi.org/10.61871/mj.v48n1-9>
- Oxford, Rebecca L. 2016. "Toward a Psychology of Well-Being for Language Learners: The 'EMPATHICS' Vision." In *Positive Psychology in SLA*, edited by Peter D. MacIntyre, Tammy Gregersen, and Sarah Mercer, 10–88. Bristol, UK: Multilingual Matters. <https://doi.org/10.21832/9781783095360-003>
- Reinders, Hayo, and Cem Balçıkanlı. 2011. "Do Classroom Textbooks Encourage Learner Autonomy?" *Novitas-ROYAL (Research on Youth and Language)* 5 (2): 265–272.
- Richards, Jack C. 2014. "The ELT Textbook." In *International Perspectives on Materials in ELT*, edited by Sue Garton and Kathleen Graves, 19–36. London: Macmillan.
- Seligman, Martin E. P. 2011. *Flourish: A Visionary New Understanding of Happiness and Well-Being*. New York: Free Press.
- Taylor, Stephanie. 2013. *What Is Discourse Analysis?* London: Bloomsbury Academic. <https://doi.org/10.5040/9781472545213>
- Tomlinson, Brian. 2011. "Introduction: Principles and Procedures of Materials Development." In *Materials Development in Language Teaching*, edited by Brian Tomlinson, 1–32. Cambridge Language Teaching Library. Cambridge: Cambridge University Press.

Vitta, Joseph P. 2021. “The Functions and Features of ELT Textbooks and Textbook Analysis: A Concise Review.” *RELC Journal* 54 (3): 856–863.

<https://doi.org/10.1177/00336882211035826>

Waltz, Thomas J., Byron J. Powell, Monica M. Matthieu, Matthew J. Chinman, Jeffrey L. Smith, Enola K. Proctor, Laura J. Damschroder, and JoAnn E. Kirchner. 2015. “Innovative Methods for Using Expert Panels in Identifying Implementation Strategies and Obtaining Recommendations for Their Use.” *Implementation Science* 10 (suppl. 1): A44.

<https://doi.org/10.1186/1748-5908-10-S1-A44>

Wang, Yongliang, Ali Derakhshan, and Lawrence Jun Zhang. 2021. “Researching and Practicing Positive Psychology in Second/Foreign Language Learning and Teaching: The Past, Current Status and Future Directions.” *Frontiers in Psychology* 12: 731721.

<https://doi.org/10.3389/fpsyg.2021.731721>

Wiesner, Claudia. 2022. “Doing Qualitative and Interpretative Research: Reflecting Principles and Principled Challenges.” *Political Research Exchange* 4 (1).

<https://doi.org/10.1080/2474736X.2022.2127372>