

**Milica M. Kočović Pajević\***

State University of Novi Pazar, Department of Philology

<https://orcid.org/0000-0002-2997-8438>

**Jelena V. Grubor\*\***

State University of Novi Pazar, Department of Philology

<https://orcid.org/0000-0003-1518-7704>

## **ESP TEXTBOOK EVALUATION: GETTING A GRIP ON OLD PROBLEMS AND CHALLENGES**

Since ESP literature pivots around “dos and don’ts” concerning textbook qualities and due to scarcity of research into the criteria of textbook evaluation, the paper explores key challenges when a scientifically rigorous perspective is adopted (cf. evaluators, evaluation process/methods/tools, validity, reliability and practicality). The main goal hence is to point to the critical issues of ESP textbook evaluation that need to be resolved prior to undertaking any empirical testing, driven by the underlying idea that any evaluation tool must be operationalizable, valid and reliable so as to provide valuable data. Due to the complexity of the aspects discussed, both theoretical and methodological, further research is needed either to support or dispute the proposed approach.

**Key words:** ESP textbooks; textbook evaluation; reliability, validity, practicality.

### **1. INTRODUCTION**

Despite long-standing critiques of their limitations, English for specific purposes (ESP) textbooks play a pivotal role in ESP curricula (Mishan, 2022), as they do in English language teaching (ELT) contexts in general. Although no single coursebook can be ideal for every class and every context (Azarnoosh et al., 2018; McGrath, 2002), studies show that the majority of teachers regularly rely on coursebooks (e.g. Mishan, 2022; Tomlinson, 2012).

---

\* mkocovicpajevic@np.ac.rs

\*\* bram.english@yahoo.co.uk

The overarching reason behind this adherence to and/or reliance on textbooks lies in the fact that they provide relevance and structure. In terms of advantages, they integrate domain content and language, provide ready-made syllabi and workable materials, save planning time for teachers (Gilmore, 2007; Harwood, 2010; Richards, 2001; Ushioda, 2013). Textbooks provide teachers and/or language instructors with a framework to rely on in terms of appropriate and consistent language level, authentic samples of language (e.g. audio and/or video materials, articles from domain-specific magazines etc.), photocopiable supplementary materials (e.g. role plays reflecting real-life scenarios relevant to the field of study). They also reduce planning time and help novice teachers and/or those new to the respective ESP field through a teacher's book. Furthermore, Richards (2001) highlights that textbooks provide a planned, systematic coverage of language components and skills, offering continuity across lessons, as well as that they can give learners a sense of progress and security by offering tangible evidence of what they are learning. Through textbook learners also have a framework to rely on in terms of self-study, frame of reference, practice, revision, and preparation for tests and/or examinations.

Regarding disadvantages, textbooks fall short of keeping pace with rapidly changing fields and/or professional domains (e.g. IT, healthcare, engineering) which are evolving so quickly that textbooks might become obsolete shortly after their publication. In addition, textbooks might put limits on teacher and learner autonomy by imposing a strict plan and content to follow. Concerning creativity, reliance on fixed schemes and pre-packaged tasks can curb teachers' initiative to design open, creative activities and limit learners' opportunities for creative language use and problem-solving (Allwright, 1981; Richards, 2001). In addition, some textbooks might present inauthentic or outdated contexts by incorporating reading and listening materials that are oversimplified or abridged/adapted or outdated (Gilmore, 2007). In the most extreme case, they cannot possibly fulfil every learner's needs. These drawbacks may place extra burden particularly to ESP teachers working within a learner-oriented paradigm, which is geared towards learners, their needs and/or affinities, as well as directed at their learning outcomes and/or achievement (Elkasović & Čolakovac, 2023; Huang, 2016).

All the stated therefore underscores the need for a systematic approach to ESP textbook evaluation, since evaluation may facilitate tailoring materials to learners' needs and curriculum aims (Al Abdullah, 2022). In practice, evaluation in the field of ESP appears to have often been *ad hoc* and/or checklist based, focusing on teacher's first impressions, neglecting a more analytical approach. In theory, the existing evaluation models typically rely on broad EFL frameworks (e.g. Sheldon 1988, Cunningsworth 1995; Shak et al., 2022), which disregard specific ESP-relevant features and requirements (e.g. professional relevance, domain-specific lexis, task authenticity). In a word, ESP-specific evaluation demands for rigorous, theory-driven

approach and/or empirically grounded evidence, rather than reliance on impressionistic data within “dos and donts” lists.

In order to raise awareness of the need for the creation of valid and reliable evaluation criteria, operationalizable frameworks and/or measurement instruments, the main goal of the paper is to identify and systematize the main challenges in ESP textbook evaluation that need to be addressed before undertaking any empirical research. By clarifying issues related to evaluators, evaluation procedures, and by ensuring the key preconditions for an effective research instrument (i.e. validity, reliability, practicality) are met, the paper seeks to lay the groundwork for developing frameworks and tools that can be consistently applied across contexts. Addressing these challenges is crucial as research findings risk being impressionistic, nonreplicable, and ultimately of limited value for advancing ESP pedagogy without rigorous and operationalizable evaluation criteria. To do so, we will provide a taxonomy of challenges that need tackling prior to devising operationalizable, valid and reliable evaluation tools. By addressing the most relevant issues concerning ESP textbook evaluation in turn, we will attempt to contribute to the respective field by providing some potential solutions to the discussed problems. Hence, the issues emphasized in the paper and possible solutions are intended to provide a well-defined basis for subsequent empirical studies and for refining evaluation practices across various ESP environments.

## **2. CHALLENGES THAT NEED ADDRESSING**

ESP textbook evaluation faces a wide range of challenges. In comparison to the general field of ELT, ESP researchers and/or practitioners encounter an additional layer of complexity: not only do materials need to be appropriate in terms of language level and content and learner-oriented, but they must also be aligned with the rapidly evolving content of a specific professional field (Basturkmen, 2010; Marcu, 2020). As such, the evaluation of a textbook should extend beyond a checklist of general language teaching features. In line with recent trends in ELT, it requires skill integration and learner motivation as well, but also careful analysis of domain-specificity. Apart from this most common general issue, there is another central problem that refers to bridging the gap between academic expectations for scientific rigor, and strong tendency to use impressionistic, often intuitive, methods of measurement in practice (Azarnoosh, et al., 2018; Ellis, 1997).

In general, these challenges may reflect both theoretical dilemmas (e.g. how to define and measure abstract constructs, such as learner engagement), but also practical constraints, such as limited time that practitioners have for their textbook evaluation, need for and/or lack of evaluation training and/or institutional support for thorough evaluation, together with recognizing the potential of textbooks as well as their drawbacks. Evaluators may also face the challenge of ensuring that

textbooks are sensitive to diverse cultures and learner populations, as appropriacy and motivation, for example, in one educational, social and/or cultural setting may not translate effectively to another. Another difficulty lies in coordinating input from multiple stakeholders, such as language teachers, subject-matter experts, curriculum planners, policy makers, learners, whose overriding priorities and/or perspectives on what constitutes an effective ESP textbook may be in almost diametrical opposition (Bachman & Palmer, 2004; Santika & Wirza, 2021; Tomlinson, 2003).

In sum, the challenges we discuss within the taxonomy given below are framed around six key questions. Each question highlights issues deriving from research and/or classroom experience/practices. Plainly put, the following subsections briefly outline the most prominent challenges that emerge when evaluation is approached from a scientifically grounded and/or pedagogically relevant perspective. Therefore, each subsection may serve both as a critical reflection as well as a guide for improving current ESP evaluation practices.

### **2.1. Challenge 1: To use or not to use?**

Given the prominent role of textbooks in ESP (Kočović Pajević, 2024; Mishan, 2022; Tomlinson, & Masuhara, 2018), the main question that arises is whether to use them altogether or not.

The advocates against textbooks point to many disadvantages of using them. They argue that textbooks can be static, culturally biased, or ill-suited to local needs (Allwright, 1981; Nunan, 1988). Basturkmen (2010) and Gilmore (2007), for instance, underline that many ESP books lack truly immersive professional tasks that learners often find not to represent real-life scenarios. Others maintain that textbooks dictate a fixed syllabus, leaving little room for learner-driven content (Allwright, 1981; Richards, 2001; Tomlinson, 2012). Accordingly, some ESP instructors consider the idea of simply eschewing coursebooks altogether in favor of tailor-made materials.

Nevertheless, abandoning textbooks completely is often unfeasible and impracticable. Coursebooks provide coherence, integrated skills development, and a “trustworthy source” of materials and support for structuring classes, especially for novice or overworked teachers. Mishan (2022) reports that teachers, especially those new to the profession, welcome the security and structure of ready-made textbooks. Textbooks offer well-presented materials, they can be suitable for students’ needs and provide students with the possibility not only to look ahead in terms of the content, but to refresh their knowledge as well (O’Neill, 1982). Moreover, studies investigating teachers’ use of textbooks show that approximately 92% of the tested teachers “regularly use” a coursebook and around 65% “always or frequently” use them (Mishan, 2022; Tomlinson, 2012). Finally, from an administrative perspective, selecting an ESP textbook streamlines curriculum planning and saves time and finances.

In short, while the reality suggests that teachers need to supplement and/or adapt coursebooks, the entire rejection of textbooks is rather unrealistic. The debate over *whether* to use textbooks hinges on balancing their practical benefits against their pedagogical drawbacks. Despite the disadvantages stated above, authors are in general agreement that advantages outnumber potential drawbacks (Azarnoosh, & Ganji, 2014; Cunningsworth, 1995; Kočović Pajević, 2024).

In practical terms, systematic and in-depth evaluation of a textbook leads to sounder choices, regarding the selection of the most optimum one, because adequate evaluation provides an overview of its positive and negative sides and opens room for material improvement and/or needed supplementary materials. In a word, using a textbook in an ESP classroom is both necessary and justifiable because overall its advantages outweigh its limitations.

## **2.2. Challenge 2: To evaluate or not to evaluate?**

The next question to consider when a course makes use of a textbook is whether to evaluate it formally or else intuition suffices, as evaluation typically involves impressionistic or predictive evaluation (cf. Ellis, 1997). In spite of the fact that teacher intuition and classroom experience *are* valuable, teachers can be subjective and/or inconsistent, particularly across different teaching contexts and/or student populations.

Simply put, experience and research support the idea of a systematic approach to evaluation. Since textbook adoption is an “executive educational decision” with major pedagogical and financial stakes, basing it on data turns out prudent (Azarnoosh et al., 2018). Specifically, checklists and other evaluation tools are explicitly designed to uncover strengths and weaknesses of a coursebook prior and upon its use (Al Abdullah, 2022). Numerous authors emphasize that checklists can help *pre-use* evaluation (predictive evaluation) to select the most suitable book, as well as *post-use* evaluation (retrospective evaluation) to diagnose how well a selected book meets learners’ needs or else whether they enable them to attain language mastery (e.g. Dudley-Evans, & St John, 1998; Ellis, 1997; McDonough, Shaw, & Masuhara, 2013; Tomlinson, 2012). Similarly, this temporal stand is essential for making sure that materials continue to meet learners’ needs throughout the course, especially in fast-changing professional fields like medicine or IT. Nevertheless, only by combining pre-use, in-use and post-use evaluation, along with quantitative and qualitative methods, can we obtain a clearer perspective. Since evaluation is an intrinsic part of teaching and must be ongoing (Rea-Dickins & Germaine, 1994; Sheldon, 1988), discarding evaluation may lead to misalignment between course content and learners’ goals.

In practice, though, formal evaluation is often reduced to a minimum. Some teachers rely solely on publishers’ claims and/or cursory reviews, and others may feel that critiquing a chosen book is a “waste of time” due to a large pool of textbooks in the

global ELT market. As previously noted, impressionistic evaluation is often the most commonly employed type, which is not remotely sufficient for an informed opinion on the book and/or assessment of its strong and weak sides. Moreover, as Mishan observes “coursebooks formats have continued to replicate each other in a washback effect, whereby new coursebooks are based on previous commercially successful models”, implying that many “new” textbooks tend to share the same features with “older” ones (Mishan, 2022, p. 493).

In summary, recent studies in ESP contexts show that structured evaluation yields useful insights (e.g. Al-Abdullah, 2022; Atigh, & Khabbazi, 2021; Kočović Pajević, 2024; Mollanazar et al., 2023) and the ESP literature accentuates that evaluation should be cyclical. In other words, one evaluates a textbook *before* use to facilitate its selection, and *after* use to improve teaching and course content, as well as to meet the needs of as many learners as possible. Furthermore, systematic post-use evaluations can generate valuable feedback not only for teachers but also for material developers and authors, thereby contributing to more specific, more customized and field-relevant ESP textbooks.

Consequently, the question is not *whether* to evaluate, but *how rigorously* and *continuously* to do so. The conclusion that imposes itself is that textbooks need to be evaluated as systematically as possible, ideally prior and during their use, with objectives, methods and values of the teaching program in mind (Cunningsworth, 1995), alongside special attention to learners’ needs and course settings (its length, number of learners etc.).

### **2.3. Challenge 3: Who to evaluate?**

Another question to be tackled is to do with the *who* element, i.e. the person conducting the evaluation (evaluator).

Traditional approaches to L2 textbook evaluation often assume that experienced teachers can judge a book’s quality on their own (e.g. Atigh, & Khabbazi, 2021; Mukundan et al., 2011). Notwithstanding that teachers and curriculum designers bring in classroom expertise and awareness of students’ needs, they also have personal biases. Preferably, evaluation involves multiple stakeholders. Recent studies take the triangulating perspective: combining feedback from ESP instructors, subject-matter experts, and, importantly, learners themselves (e.g. Kočović Pajević, 2024; Sari, & Sari, 2020). By obtaining opinions from language instructors and field professionals as well as neutral (external) evaluators, educational and real-world priorities can be thus balanced. This reduces individual bias and provides a more nuanced view of textbook qualities. Likewise, learners’ perceptions can reveal issues their teachers might have missed. Preliminary research in different ESP contexts indicates, for example, that many students want more focus on language form in addition to subject matter content or that they find certain topics irrelevant (Litz, 2005; Sari, & Sari,

2020; Warti, 2020). Incorporating student feedback helps ensure that the textbook meets learners' evolving goals.

Another important consideration is that the effectiveness of any tool depends on the evaluator's familiarity with ESP contexts and intended learner population (Harwood, 2010; McGrath, 2002). This is especially significant with external evaluations and/or evaluators so that it can be ensured that the evaluation is as unbiased as it may be.

In short, textbook evaluation is most effective when conducted collaboratively, or more precisely, when including perspectives of all relevant agents (i.e. evaluator's, ESP instructor's, ESP learners') in the evaluation process that can be analysed and compared. Thus, instructors bring in pedagogical insights, based on their methodological knowledge and/or teaching experience, thereby providing insights into pedagogical suitability, classroom applicability, and methodological soundness. ESP subject specialists (e.g. external evaluators) ensure domain relevance, alignment with successful teaching approaches and/or theory-driven and empirically grounded data, or more simply whether materials are aligned with disciplinary content, terminology, professional relevance. Furthermore, learners, as the ultimate beneficiaries, provide individual feedback on perceived usefulness for their future professions, motivational dimensions, their preference of engagement levels, the extent to which a coursebook fulfils their needs, and the like. Finally, involving different evaluators is beneficial for professional development, since it may prompt both teachers and subject experts alike to reflect more deeply on their practices and assumptions.

#### **2.4. Challenge 4: How to evaluate?**

Yet another question to deal with is concerned with the method of evaluation, or else *how* to perform evaluation in terms of measurement instruments. This aspect is particularly challenging as well as significant because when different data collection instruments and/or methods are used for the same construct, they may yield different results.

Broadly put, literature offers a variety of tools to choose from: checklists, rating scales (i.e. continuous scales), questionnaires, interviews, and/or their blend (e.g. open-cloze questions used in combination with a type of scale: e.g. Likert-type, frequency, semantic differential, etc.). Since different researchers use different terms interchangeably, we will further define each said data collection tool primarily to avoid possible ambiguities.

Naturally, different instruments are used for data collection depending on the type of data being measured (Phillip, 2024), but regarding the subject matter of this paper, the most used tools are checklists, also termed evaluation questionnaires (e.g. e.g. Cunningsworth, 1995; McDonough, & Shaw, 2003; McGrath, 2002; Richards, 2001; Sheldon, 1988). *Checklists* are predefined sets of specific criteria, principles,

guidelines or questions, or else categories for judgement, against which an object of evaluation is tested (Fuentes, & Martínez, 2018; Wilson, 2013). ESP checklists include dozens of features to be ticked off and that can be filled out by teachers and/or evaluators (e.g. Cunningsworth, 1995; Sheldon, 1988). Regarding textbook evaluation, they may be used as part of criteria development, namely, key categories that a coursebook is meant to include (e.g. Kočović Pajević, 2024; McGrath, 2002; Mukundan et al., 2011; Sheldon, 1988), given in variable forms and levels of comprehensiveness. In terms of measure, they approximate to dichotomous scales due to their binary range of values (i.e. “yes/no”, “present/not present”). Although checklists allow side-by-side comparisons of multiple books and help ensure that all important items are considered, they fall short of providing different degrees of presence if taken in their typical binary form. Differently put, they may gloss over subtleties or lack descriptive levels for each feature, which is why they would be better used as a type of scale that features a wider range of degrees (e.g. five- or seven-point scale). Plainly put, they are useful in terms of having a criteria framework (i.e. as a starting point), as well as for evaluation criteria development, but they would be better used in parallel with other research tools (cf. Kočović Pajević, 2024; Litz, 2005).

On the other hand, *rating scales* (self-reports) include different degrees of frequency, levels of (dis)agreement, thus offering participants a wider range of points to choose from and allowing for more nuanced differences (cf. Albert et al., 2010; Grubor, 2012). These scales are commonly checked for psychometric properties, which adds on to their validity (e.g. Grubor, 2021, 2020, 2012; Kočović Pajević, 2024; Phillip, 2024). In general, rating scales have been used in ESP contexts to assess students’ and language instructors’ and/or teachers’ opinions of their textbooks (e.g. Atigh, & Khabbazi, 2021; Kočović Pajević, 2024; Litz, 2005).

Finally, although questionnaires are typically used as an umbrella term for different types of research instruments in applied linguistics (Dörnyei, & Dewaele, 2022),<sup>1</sup> for the sake of distinguishing between different types of data collection tools, in this paper they relate to an instrument comprising a set of items that are predominantly open-cloze questions that are analyzed qualitatively (e.g. via content analysis). In the ESP context these tools have been mainly used to define predominant qualities of a textbook as well as to obtain more detailed insights into textbook users’ opinions (e.g. Litz, 2005; Mukundan, Hajimohammadi, & Nimehchisalem, 2011). As for interviews (cf. Brinkmann, 2014; Loewen, & Plonsky, 2016), which are also treated qualitatively, the interviewer has a predefined list of questions for their interviewees, given in a predetermined order (structured interview), as in Atigh and Khabbazi (2021), Harbi (2017), Karimnia and Jafari (2017), coupled with follow-up questions arising from conversation (semi-structured interviews), as in Kočović Pajević (2024), Karimnia

<sup>1</sup> Questionnaires are also interchangeably used with surveys as another generic term for different types of data collection tools (Brown, 2001).

and Jafari (2017).<sup>2</sup> Another qualitative data collection method is focus groups. Thus, besides individual interviews, some ESP studies employed focus groups to facilitate discussion among groups of textbook users, thereby generating richer data through interaction (e.g. Krueger, & Casey, 2015).

Moreover, digital tools and learner analytics are increasingly being used to assess textbook effectiveness, particularly when materials are delivered in hybrid or online ESP environments (e.g. Sari, & Sari, 2020; Soto, Fabre-Merchán, & Amaya-Díaz, 2025).

All things considered, recent trends in applied linguistics call for mixed-methods evaluation (Atigh, & Khabbazi, 2021); hence, methods are often combined in research. For example, classroom observations may be used to determine how the textbook is used, interviews with teachers or students and/or focus groups can be employed so as to obtain insights into their use, and the quantitative method to judge effectiveness of a textbook on student performance and/or achievement. A combined approach can therefore capture both the quality of materials (e.g. coverage of skills) and their effects (e.g. learner engagement or motivation). By combining quantitative and qualitative methods, we can therefore obtain more reliable and detailed insights into characteristics of a textbook.

### **2.5. Challenge 5: Validity, reliability, and practicality**

One of the most important issues that needs careful consideration pertains to the general assessment requirements: validity, reliability and practicality of a tested construct and/or instrument. Ensuring that an evaluation tool meets these standards is essential, as shortcomings in any of these areas can undermine the credibility of the findings and thus lead to misguided decisions about textbook adoption and/or revision.

Even when operating within a framework, ensuring rigorous evaluation is indeed challenging. Validity entails that the evaluation accurately reflects the textbook's quality. For internal validity, criteria must truly represent important aspects of good ESP materials. Involving multiple experts in defining the criteria (e.g. ESP researchers, curriculum designers) helps ensure content validity. Piloting a checklist and seeking feedback can confirm construct validity. Soto et al. (2025) stress that evaluation should check text accuracy and alignment with needs analysis, tying the criteria against an actual learning objective. Therefore, to ensure validity, evaluation criteria must be clearly linked to the underlying constructs they purport to measure (e.g. authenticity, task usefulness) through content-and-construct validation procedures, such as expert reviews and pilot testing of the rubric items (Bachman & Palmer, 2004; Dörnyei, 2007). Every item of a scale, questionnaire etc. needs to reflect a concrete component of a textbook (Tomlinson, 2012).

<sup>2</sup> Unstructured interviews appear to be generally absent in the context of this paper.

Reliability entails that different evaluators (or the same evaluator at different times) yield consistent results. Establishing reliability requires detailed operational definitions, evaluator training, and empirical checks like inter-rater agreement statistics (e.g. Cohen's Kappa, Intraclass Correlation Coefficient (ICC), and/or percentage agreement) to confirm consistency across users and occasions (Dörnyei, 2007; Hughes, 2003). This can be improved by providing clear rubrics and training or instructions for evaluators. For instance, if several teachers rate "authenticity", their judgments should be calibrated to the employed scale and their ratings need to yield similar results. Inter-rater reliability can be checked by having multiple evaluators score the same sample materials and computing agreement. Using mostly objective or relatively objective items (counts or benchmarks) also boosts reliability since these involve less ambiguity and provide a clear frame of reference within which evaluators can work. Lastly, reliability also relates to research instruments in terms of how closely related a set of items are as a group, whether items of a scale, for instance, make up a construct and whether they measure what they are devised and/or meant to measure. With this respect different statistical reliability analyses are performed such as Cronbach's alpha, a measure of internal consistency, Guttman's lambda, McDonald's omega.

Practicality, however, is often in discord with validity and/or reliability. A checklist with hundreds of items (though exhaustive) will be too burdensome for regular use, whereas a very simple checklist might miss important aspects. Thus the key is in balance: to include core, high-impact criteria while keeping the tool usable (Skierso, 1991; Tomlinson, 2003). Many authors suggest tailoring the length of evaluation instruments to teachers' time constraints, and ensuring each item is clearly defined (Al-Abdullah, 2022; Litz, 2005; Jusuf, 2018). Practicality, consequently, demands a judicious selection of high-impact criteria (i.e. balancing comprehensiveness with usability), so that regular classroom teachers can apply an instrument within realistic time and resource constraints (Brown, 2018; Richards, 2001). The same is true for rating scales completed by ESP textbook users (e.g. learners and ESP instructors).

Ultimately, criteria and scoring rubrics should be created in consultation with experts; the developed tool needs piloting (both in terms of statistical testing for psychometric properties and qualitative reflections of textbook users: learners and/or ESP teachers); the instruments might require further improvements based on preliminary findings; once they display good psychometric properties, they may be used across multiple ESP courses. Such a mixed-method cycle echoes recommendations by Richards (2001) and Dörnyei (2007) for language assessment – using both statistical analysis and user feedback to improve evaluation tools. By doing so, the evaluation becomes not just a one-off judgment but part of a curriculum review and/or evaluation process.

## **2.6. Challenge 6: Acknowledging learners' individual needs**

The last but crucial question that needs addressing is to delve into learners' needs that may concern coursebook content, design, layout and/or individual differences (IDs) among learners.

ESP courses are often aimed at diverse students. Any evaluation must therefore consider how well a textbook can be adapted to individual and/or group needs, along with its flexibility in terms of adaptation, supplementation and/or self-study options. Most authors agree that a thoroughly conducted needs analysis is indispensable (Al-Abdullah, 2022; Gu, Bo, & Ren, 2019; Koltai, 2018; Minodora, 2015; Sheldon, 1988) and thus advise that textbooks be accompanied by ongoing needs analysis and feedback collection (Graves, 2000; Sheldon, 1988; Tomlinson, 2003). This is a challenging task, since learners in an ESP class have similar needs (in terms of their profession), but differ in other aspects (e.g. language level, age, background, learning styles etc.). Nevertheless, their professional needs are the most important aspect and therefore ought to be assessed in a systematic way.

Research into ESP students' needs reveals that ESP students consistently report that materials need to be directly relevant to future professional activities, to provide field-specific vocabulary and realistic tasks, to offer opportunities so that learners can practice real-world communication, to be flexible enough to accommodate diverse proficiency levels as well as to be engaging and motivating (Basturkmen, 2010; Gu, Bo, & Ren, 2019; Rajović, 2021). Regarding earlier checklists, they frequently failed to reflect the stated needs. Many rather focused on general pedagogical features, such as clarity of instructions, layout, or cultural appropriateness (e.g. Cunningsworth, 1995) and often underemphasized contextual relevance and needs analysis. In contrast, more recent checklists tend to integrate needs analysis components more often, reflecting a tendency towards learner-centered evaluation, both in ELT and ESP (e.g. Basturkmen, 2010; Mukundan et al., 2011). Such checklists typically address relevance, authenticity, engagement, and flexibility as core dimensions of quality.

What is more, different factors are inextricably intertwined with learners' experience with a coursebook. Those are learner motivational drives, cultural fit of the content, their appeal to learners, and so on. For example, if students do not find topics or examples engaging, motivation drops (Dörnyei, & Ushioda, 2011). Richards (2001) emphasizes the importance of appeal, variety and interest as core criteria in evaluating materials. McKay (2003) underlines that the cultural content of materials strongly influences learner attitudes and can affect motivation and identity. Radić-Bojanić and Topalov (2016) show how textbook topics influence perceived relevance and student engagement.

In addition, ESP textbook evaluation is meant to solicit learner feedback on content relevance and interest. Studies show that students value materials that connect with their future work (e.g. Belcher, 2006; Dudley-Evans & St. John, 1998; Liang, 2023), as well as materials that are perceived as interesting to learners. Mishan

and Timmis (2015) argue that materials should “spark curiosity and engagement,” noting that perceived dullness is a frequent reason why teachers abandon otherwise linguistically sound materials. Interest has been one of the main criteria for choosing textbooks for many years. One of the pioneers to include interest as one of the three major criteria for judging coursebooks (besides usability, and balance) was Grant (1987). Nevertheless, investigating the coursebook potential to engage learners seems to be absent in empirical research, despite its relevance in the general context of ELT studies (e.g. Radić-Bojanić, & Topalov, 2016; Vitta, 2023; Wen-Cheng et. al., 2011).

Overall, ignoring IDs in textbook evaluation (such as individual learner needs, motivation, attitudes), poses a risk of overlooking the “human dimension” of language (Mihăescu, 2008). Besides research work, those factors can be explored by ESP instructors themselves in the form of short questionnaires and/or focus-group discussions, thereby ensuring that the final decision about the choice of a coursebook acknowledges individual learner needs and preferences.

### 3. CONCLUDING REMARKS

ESP textbook evaluation is inherently complex, which further highlights the need for rigor and practicality. Textbooks have been and will be crucial in ESP, so systematic evaluation cannot be optional, since it is essential for aligning materials with course goals.

This paper has delineated six key challenges and sketched a framework on how they can be addressed. Overall, one of the key issues that will be the subject matter of further research is to identify underlying components of ESP textbooks, to develop evaluation criteria following strict procedures, to check their validity and reliability, and consequently to classify them by their measurability so as to allow for comparisons with different types of evaluation instruments as well as across textbook users. Another conclusion that may be derived from a large body of research within applied linguistics studies in general is that mixed-methods research design is preferable in order to obtain as valid and informative results as possible. In doing so, evaluators can operationalize even abstract qualities (like authenticity) while still capturing indispensable subjective insights.

Recent studies support this integrated view. For example, Soto et al. (2025) underscore that evaluation must include checks of text accuracy and suitability to learning situations, ongoing feedback is necessary to make sure that a textbook aligns with evolving student needs. Al-Abdullah (2022) showed how checklist data can directly inform curriculum revisions and teacher training to remedy textbook gaps. Crucially, by combining teacher, expert, and learner perspectives, evaluation becomes richer and more valid (cf. section 2.3).

Finally, future studies could also explore how evaluation instruments can be refined and validated across different ESP domains while still retaining flexibility

to adapt to local contexts. There is also a need for more empirical studies that test the effectiveness and reliability of mixed-method evaluation models, particularly in under-researched areas such as vocational ESP, CLIL-integrated courses, or digital/AI-enhanced materials. Another promising direction lies in developing learner-centered evaluation tools, where learners themselves play a more active role in assessing the relevance, usability, and motivational impact of the textbook. Technology-mediated evaluation might be another potential direction for future research, exploring how digital analytics, corpora, or AI-based tools can support or automate parts of the evaluation process. Cross-cultural and cross-institutional comparisons would also be valuable, as they could reveal how well evaluation tools transfer across diverse educational contexts and ESP domains.

In practical terms, a more rigorous ESP evaluation is beneficial to class outcomes. Coursebooks can be updated or supplemented when needed, ESP teachers can become more reflective, and learners thus get materials that match their learning purposes. Moreover, the increasing availability of digital resources and corpus-based tools opens new possibilities for designing evaluation instruments that are both more efficient and more responsive to learners' developing needs.

As many authors concur, textbook evaluation is vital because the information gained along the way can facilitate decision-making concerning the use of textbooks and ESP materials. By embracing systematic evaluation frameworks, which are grounded both in empirical research and classroom experience, ESP programs can ensure that their textbooks remain assets rather than liabilities.

## References

- Al-Abdullah, S. (2022). *ESP textbook evaluation: The case of Kuwaiti students of business administration*. *English Language Teaching*, 15(10), 78–85.
- Albert, B., Tullis, T., & Tedesco, D. (2010). *Beyond the usability lab: Conducting large-scale online user experience studies*. Burlington: Elsevier.
- Allwright, D. (1981). What do we want teaching materials for? *English Language Teaching Journal*, 36(1), 5–18.
- Atigh, R., & Khabbazi, S. (2021). The evaluation of ESP textbooks based on students' perspectives. *International Journal of Linguistics, Literature and Translation*, 4(5), 56–62.
- Azarnoosh, M., Khosrojerdi, M., & Zeraatpishe, M. (2018). Evaluation of ESP textbooks. In M. Azarnoosh, M. Zeraatpishe, A. Faravani, & H. R. Kargozari (Eds.), *Issues in materials development* (pp. 129–143). Brill. [https://doi.org/10.1163/9789004387379\\_008](https://doi.org/10.1163/9789004387379_008)
- Azarnoosh, M., & Ganji, M. (2014). ESP book evaluation: The case of management course book. *International Journal of Secondary Education*, 2(4), 61–65.

- Bachman, L. F., & Palmer, A. S. (2004). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Basturkmen, H. (2010). *Developing courses in English for specific purposes*. London: Palgrave Macmillan.
- Belcher, D. (2006). English for specific purposes: Teaching to perceived needs and imagined futures in worlds of work, study and everyday life. *TESOL Quarterly*, 40(1), 133–156.
- Brinkmann, S. (2014). Unstructured and semi-structured interviewing. In P. Leavy (Ed.), *The Oxford handbook of qualitative research* (online edn.). Oxford Library of Psychology. <https://doi.org/10.1093/oxfordhb/9780199811755.013.030>, accessed on 27 June 2025.
- Brown, H. D. (2018). *Language assessment: Principles and classroom practices* (3<sup>rd</sup> edn.). Pearson Education.
- Brown, J.D. (2001). *Using surveys in language programs*. Cambridge: Cambridge University Press.
- Cunningsworth, A. (1995). *Choosing your coursebook*. London: Heinemann.
- Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation* (2<sup>nd</sup> edn.). London: Pearson.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies* (pp. 238–256). Oxford: Oxford University Press.
- Dörnyei, Z. & Dewaele, J.-M. (2022). *Questionnaires in second language research. Construction, administration and processing* (3<sup>rd</sup> edn.). New York: Taylor & Francis.
- Dudley-Evans, T., & St John, M.J. (1998). *Developments in ESP: A multi-disciplinary approach*. Cambridge: Cambridge University Press.
- Elkasović, S., & Čolakovac, J. J. (2023). The teaching of ESP lexis revisited: The effect of authentic text adaptation on vocabulary retention. *ESP Today*, 11(1), 75-95. <https://doi.org/10.18485/esptoday.2023.11.1.4>
- Ellis, R. (1997). The empirical evaluation of language teaching materials. *ELT Journal*, 51(1), 36–42.
- Fuentes, E.M. & Martínez, J.R. (2018). Design of a checklist for evaluating language learning websites. *Porta Linguarum*, 30, 23–41.
- Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40(2), 97–118.
- Grant, N. (1987). *Making the Most of Your Textbook*. London: Longman.
- Graves, K. (2000). *Designing language courses: A guide for teachers*. Boston: Heinle & Heinle.
- Grubor, J. (2021). The context of learning English as a foreign language (AL2LC scale). *Folia Linguistica et Litteraria*, 36, 291–308. <https://doi.org/10.31902/fl.36.2021.17>

- Grubor, J. (2020). What makes up the attitude to learning English: EFLS-ALE scale. G. Jašović et al. (Eds.), *Science beyond boundaries 3: Language and languages* (pp. 31–51). Kosovska Mitrovica: Faculty of Philosophy, University of Priština.
- Grubor, J. (2012). *Stavovi prema učenju engleskog kao stranog jezika i njihov uticaj na postignuće* [neobjavljena doktorska disertacija]. Beograd: Filološki fakultet, Univerzitet u Beogradu.
- Gu, H., Bo, H., & Ren, L. (2019). Developing ESP teaching materials based on the analysis of information engineering majors' needs. *Open Journal of Social Sciences*, 7(10), 121–131.
- Harbi, A.A.M. (2017). Evaluation study for secondary stage EFL textbook: EFL Teachers' Perspectives. *English Language Teaching*, 10(3), 26–39.
- Harwood, N. (Ed.). (2010). *English language teaching materials: Theory and practice*. New York: Cambridge University Press.
- Huang, Y.-H. (2016). Examining the usefulness of an ESP textbook for information technology: Learner perspectives. *World Academy of Science, Engineering and Technology, International Journal of Educational and Pedagogical Sciences*, 10(8), 2808–2815.
- Hughes, A. (2003). *Testing for language teachers* (2<sup>nd</sup> edn.). Cambridge: Cambridge University Press.
- Jusuf, H. (2018). The models of checklist method in evaluating ELT textbooks. *Al-Lisan*, 3(2), 17–35. <https://doi.org/10.30603/al.v3i2.422>
- Karimnia, A., & Jafari, F.M. (2017). Critical ESP textbook evaluation: The case of visual arts textbook. *Sustainable Multilingualism*, 11(1), 219–236.
- Kočović Pajević, M. (2024). *Evaluacija udžbenika za engleski kao jezik struke neposrednih korisnika u obrazovnom procesu* [neobjavljena doktorska disertacija]. Kragujevac: Filološko-umetnički fakultet, Univerzitet u Kragujevcu.
- Koltai, A. (2018). *Understanding learner needs: A qualitative study of ESP learner needs in EU English courses in Hungarian tertiary education* [unpublished doctoral dissertation]. Budapest: Faculty of Pedagogy and Psychology, Eötvös Loránd University.
- Krueger, R. A., & Casey, M. A. (2015). *Focus Groups: A Practical Guide for Applied Research* (5th ed.). Thousand Oaks, CA: SAGE.
- Liang, H. H. (2023, October). *Exploring EFL students' perception of evaluation of an ESP textbook and practitioners' performance* [Conference presentation]. International Joint Conference of APLX, ETRA40, and TESPA 2023: *Advancing empowerment and resilience: Language, culture, policy and pedagogy*, Taipei, Taiwan.
- Loewen, S., & Plonsky, L. (2016). *An A-Z of applied linguistics research methods*. London: Palgrave.

- Litz, D.R.A. (2005). Textbook evaluation and ELT management: A South Korean case study. *Asian EFL Journal*, 48(1), 1–53.
- Marcu, N. A. (2020). Challenges in the selection of ESP materials. *Acta Marisiensis. Philologia*, 2(1), 1–4. <https://doi.org/10.2478/amph-2022-0034>
- McKay, S. L. (2003). *The Cultural Basis of Teaching English as an International Language*. TESOL Quarterly, 37(2), 233–257.
- McDonough, J., Shaw, C., & Masuhara, H. (2013). *Materials and methods in ELT: A teacher's guide* (3<sup>rd</sup> edn.). Malden/Oxford/Chichester: Wiley-Blackwell.
- McDonough, J., & Shaw, C. (2003). *Materials and methods in ELT: A teacher's guide* (2<sup>nd</sup> edn.). Malden: Blackwell Publishing Ltd.
- McGrath, I. (2002). *Materials evaluation and design for language teaching*. Edinburgh: Edinburgh University Press.
- Mihăescu, L. (2008). The subjective dimensions of the human resources' evaluation in educational system. *Revista Economica*, 41(4), 49–57.
- Minodora Otilia, S. (2015). English for specific purposes: Past and present. *Annals of the Constantin Brancusi University of Targu Jiu. Economy Series*, 1(1), 54–55.
- Mishan, F. (2022). *The global ELT coursebook: A case of Cinderella's slipper? Language Teaching*, 55, 490–505. <https://doi.org/10.1017/S0261444820000646>
- Mishan, F., & Timmis, I. (2015). *Materials Development for TESOL*. Edinburgh: Edinburgh University Press.
- Mollanazar, H., & Salimi Beni, M. (2023). ESP materials evaluation: Toward communicative and cultural competence in legal and journalistic texts. *Translation Studies of Iran*, 21(1), 75–102.
- Mukundan, J., Hajimohammadi, R., & Nimehchisalem, V. (2011). Developing an English language textbook evaluation checklist. *Contemporary Issues in Education Research*, 4(6), 21–27.
- Nunan, D. (1988). *The learner-centered curriculum*. Cambridge: Cambridge University Press.
- O'Neill, R. (1982). Why use textbooks? *ELT Journal* 36(2), 104–111.
- Philip, M. (2024). Questionnaire, rating scale and checklist – How do they differ? *Methods in Psychology*, 10, 100145. <https://doi.org/10.1016/j.metip.2024.100145>
- Radić-Bojanić, B., & Topalov, J. (2016). Textbooks in the EFL classroom: Defining, assessing and analyzing. *Collection of Papers of the Faculty of Philosophy*, 46(3), 137–153.
- Rajović, J. (2021). *Engleski jezik struke u srednjim školama tehničkog usmerenja na Kosovu i Metohiji: analiza potreba učenika*. [neobjavljena doktorska disertacija]. Kosovska Mitrovica: Univerzitet u Prištini sa privremenim sedištem u Kosovskoj Mitrovici, Filozofski fakultet.
- Rea-Dickens, P., & Germaine, K. (1994). *Managing evaluation and innovation in language teaching*. London: Routledge.

- Richards, J. C. (2001). *Curriculum development in language teaching*. New York: Cambridge University Press.
- Santika, S., & Wirza, Y. (2021). Evaluation of an ESP textbook entitled *Practice your English competence*. In *Proceedings of the Thirteenth Conference on Applied Linguistics (CONAPLIN 2020)* (pp. 562–567). Atlantis Press. <https://doi.org/10.2991/assehr.k.210427.085>
- Sari, L.I., & Sari, R.H. (2020). ESP coursebook evaluation from the perspectives of teachers, cadets, and graduates: The case of Maritime English. In *Advances in social science, education and humanities research* (vol. 434) (pp. 56–60). Amsterdam: Atlantis Press.
- Shak, J., Gardner, S., & Arunasalam, D. (2022). Evaluating English language textbooks: A synthesis of published checklists. *RELC Journal*, 53(1), 5–20. <https://doi.org/10.1177/0033688220964793>
- Sheldon, L. (1988). Evaluating ELT textbooks and materials. *ELT Journal*, 42(2), 237–246.
- Skierso, A. (1991). Textbook selection and evaluation. In M. Celce-Murcia (Ed.). *Teaching English as a second or foreign language*. (pp. 432–453). Boston: Heinle & Heinle.
- Soto, S. T., Fabre-Merchán, P., & Amaya-Díaz, I. (2025). *ESP textbook design and development: Some considerations for language teachers embarking on ESP textbook creation*. *Revista InveCom*, 5(1), e501004. <https://doi.org/10.5281/zenodo.10888759>
- Tomlinson, B., & Masuhara, H. (2018). *The complete guide to the theory and practice of materials development for language teaching*. Hoboken, NJ: Wiley Blackwell.
- Tomlinson, B. (2012). *Materials development for language learning and teaching*. *Language Teaching*, 45(2), 143–179.
- Tomlinson, B. (2003). *Developing materials for language teaching*. London: Continuum.
- Ushioda, E. (2013). Motivation and ELT: An overview. *ELT Journal*, 67(1), 3–23.
- Vitta, J.R. (2023). 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>
- Warti, W. (2020). ESP students' needs analysis based on the learning-centred approach. *Wanastra: Jurnal Bahasa dan Sastra*, 12(2), 164–170.
- Wen-Cheng, W., Chien-Hung, L., & Chung-Chieh, L. (2011). Thinking of the textbook in the ESL/EFL classroom. *English Language Teaching*, 4(2), 91–96.
- Wilson, C. (2013). *Checklists: Credible checklists and quality questionnaires: A user-centered design method*. Waltham: Elsevier.

**Milica M. Kočović Pajević  
Jelena Grubor**

**EVALUACIJA UDŽBENIKA ZA ENGLISKI JEZIK STRUKE:  
KAKO PRISTUPITI STARIM PROBLEMIMA I IZAZOVIMA**

Rezime

Budući da se literatura iz oblasti EJS pretežno bazira na razlozima „za i protiv“ u pogledu toga šta optimalan udžbenik treba da sadrži ili ne i da nedostaju istraživanja kriterijuma za evaluaciju udžbenika, rad se bavi ključnim izazovima (v. evaluator, evaluacioni proces/metod/instrument, validnost, pouzdanost i praktičnost) imajući u vidu naučno zasnovan pristup. Osnovni cilj rada je stoga ukazati na kritična pitanja koja se moraju razrešiti pre bilo kakve empirijske provere, vodeći se osnovnom idejom da svaki evaluacioni instrument mora biti operacionalizovan, validan i pouzdan kako bi pružio značajne podatke. Zbog kompleksnosti aspekata problematizovanih u ovom radu, kako teorijskih tako i metodoloških, neophodna su istraživanja kako bi se potvrdio ili opovrgao predloženi pristup.

**Ključne reči:** evaluacija udžbenika; udžbenici za engleski kao jezik struke; pouzdanost, validnost i praktičnost