



Review Process of Textbooks in Balochistan: Structure, Clarity, and Measurability of the PRC and IRC Review Tools

Dr. Mehwish Malghani¹, Dr. Gulab Khan Khilji², Fareed Ahmed Buzdar³

¹Professor, Department of English, Sardar Bahadur Khan Women's University, Quetta, Balochistan, Pakistan,

Email: dr.mehwishmalghani@gmail.com

²Chairman, Balochistan Textbook Board, Balochistan, Pakistan

³Subject Specialist (English), Balochistan Textbook Board, Balochistan, Pakistan

ARTICLE INFO

Article History:

Received: April 20, 2026
Revised: June 16, 2026
Accepted: June 25, 2026
Available Online: July 02, 2026

Keywords:

English textbook evaluation; curriculum review; Balochistan; PRC tool; IRC tool; meta-evaluation; measurability; review instrument design

Corresponding Author:

Dr. Mehwish Malghani

Email:

dr.mehwishmalghani@gmail.com

DOI: <https://doi.org/10.59075/tamsaal.v4i7.48>



ABSTRACT

Textbooks in the province of Balochistan pass through a multi-stage review process administered by two separate institutions: the Bureau of Curriculum and Extension Centre (BoC), which applies the Provincial Review Committee (PRC) Tool, and the Balochistan Textbook Board (BTBB), which applies the Internal Review Committee (IRC) Tool. Because both institutions evaluate the same English textbook manuscripts using structurally similar checklist instruments, questions have been raised about whether the process represents genuine quality assurance or unnecessary duplication of effort. This paper addresses the following question: How do the textbook review tools used by the Balochistan Textbook Board and the Bureau of Curriculum differ in structure, clarity, and measurability of evaluation criteria? Grounded in contemporary meta-evaluation theory (Stufflebeam, 2001; Yarbrough et al., 2011), the study employed systematic comparative document analysis of the two official review tools across four analytical dimensions: structural design, evaluation domains, measurability of indicators, and degree of indicator overlap. The two instruments were found to share an almost identical checklist format using Yes/No responses supported by comments, and to overlap substantially in four foundational domains — curriculum alignment, content accuracy, language quality, and glossary requirements. However, the tools diverge considerably in institutional emphasis: the PRC tool foregrounds curriculum compliance, conceptual progression, and assessment design, while the IRC tool prioritises instructional activities, physical presentation, and cultural and social values. Measurability coding using a five-point rubric (Objective, Operationalised, Semi-objective, Subjective, Highly Subjective) revealed that a substantial share of indicators in both instruments — roughly 60% in the PRC tool and 69% in the IRC tool — are semi-objective or subjective, lacking clearly operationalised thresholds. Indicator-level overlap was estimated at 40–45%, indicating that the two tools are not simple duplicates of one another at the level of content. However, because a manuscript in practice passes through up to four review events — not two — before approval, this content-level finding does not by itself establish that the review process as a whole is free of redundancy; that broader question depends on factors such as inter-institutional coordination and consistency of application that fall outside the scope of a document-based analysis. The paper concludes that the review tools reflect a largely appropriate, if currently uncoordinated, division of quality-assurance emphasis between the two institutions, and offers recommendations — including the development of institutionally exclusive review sections — for operationalising subjective criteria, formalising institutional coordination, and strengthening reviewer training.

Introduction

Textbooks play a central role in shaping teaching and learning within formal education systems. In many contexts, particularly in developing countries, textbooks function not only as instructional materials but as the primary vehicle through which curriculum is delivered in classrooms: teachers rely on them to organise lessons and introduce concepts, while students depend on them as a major source of knowledge and practice. Because of this central role, the quality, accuracy, and pedagogical effectiveness of textbooks significantly influence educational outcomes, making the evaluation and approval of textbook manuscripts a critical responsibility of educational authorities.

To safeguard the quality of instructional materials, education systems typically establish structured review mechanisms built around review tools — instruments that set out criteria for assessing manuscripts on dimensions such as curriculum alignment, content accuracy, language appropriateness, pedagogical effectiveness, and visual presentation. In the province of Balochistan, this function is shared between two institutions: the Bureau of Curriculum and Extension Centre (BoC), which administers the Provincial Review Committee (PRC) Tool, and the Balochistan Textbook Board (BTBB), which administers the Internal Review Committee (IRC) Tool. Every textbook manuscript is expected to pass through review by both bodies before it is approved for classroom use.

The existence of two independently developed review tools evaluating the same manuscripts raises a foundational question: do the PRC and IRC tools duplicate one another, or do they assess genuinely different dimensions of textbook quality? Casual comparison suggests the two instruments share considerable structural and thematic similarity, which could mean that textbook authors and publishers are being asked to satisfy the same criteria twice, adding time, administrative workload, and cost without a commensurate improvement in textbook quality. At the same time, if the tools address different aspects of quality, the two-stage process may represent a legitimate and complementary division of institutional responsibility.

This concern is compounded by a second, related issue: even where evaluation criteria appear well specified on paper, many are expressed in general or ambiguous language for example, that a textbook's language is “appropriate for the age or grade level”, or that its illustrations are “child friendly and appealing”. Such wording provides limited guidance on how a reviewer should reach a judgement, and different reviewers may interpret the same criterion in different ways. Before the reliability of the review reports that these tools generate can be examined, it is necessary to first establish exactly what the tools themselves require how they are structured, what they emphasise, and how measurable their criteria actually are. Establishing this is the specific task, and the exclusive focus, of this paper.

Statement of the Problem

Although the multi-stage textbook review process in Balochistan is intended to ensure quality assurance, concerns exist about its efficiency. The review tools used by the BoC and the BTBB appear, on the surface, to contain many similar evaluation criteria, raising the question of whether repeated evaluation using comparable instruments contributes meaningfully to textbook quality or instead produces duplication of effort. Despite the importance of this question for the design of the provincial textbook review system, little empirical work has systematically

compared the structure, content, and measurability of the two tools side by side. This paper addresses that gap by subjecting the PRC and IRC tools to structured comparative document analysis, in order to determine whether and to what extent the two instruments are duplicative or complementary, and whether their evaluation criteria are sufficiently clear and measurable to support consistent, reliable review.

Significance of the Study

This analysis is significant for several reasons. First, it contributes to a body of research that has more often examined textbooks themselves than the tools and processes used to evaluate them; by analysing the clarity, measurability, and institutional emphasis of two real review instruments, it offers insight into how textbook quality assurance actually operates in practice. Second, it has direct practical value for policymakers: by establishing precisely where the PRC and IRC tools overlap and where they diverge, the analysis provides an evidence base for decisions about whether the current two-stage review process should be preserved, streamlined, or restructured. Third, by quantifying the proportion of semi-objective and subjective indicators in each tool, the study highlights a structural source of inconsistency in textbook evaluation and points toward specific, actionable improvements in instrument design.

Research Question

How do the textbook review tools used by the Balochistan Textbook Board and the Bureau of Curriculum differ in structure, clarity, and measurability of evaluation criteria?

This question is foundational to understanding the architecture of the provincial textbook review system. If the two tools are largely duplicative in structure and content, the multi-stage review process may represent an avoidable inefficiency. Conversely, if the tools are genuinely complementary; addressing different dimensions of textbook quality, the multi-stage process may be justified as a comprehensive, if uncoordinated, quality-assurance framework. This paper resolves the question through comparative analysis across four dimensions: structural design, evaluation domains, measurability of indicators, and degree of overlap between the two instruments.

Theoretical Framework

The study is grounded in the theory of meta-evaluation, which concerns the systematic evaluation of evaluation processes and instruments rather than of the objects those instruments assess. While the concept was originally introduced by Scriven in the late 1960s, the framework applied here draws on its more recent articulation: Stufflebeam's (2001) metaevaluation imperative, and the Joint Committee on Standards for Educational Evaluation's Program Evaluation Standards (Yarbrough et al., 2011), which together represent the field's current consensus framework for judging evaluation instruments. This framework holds that evaluation systems must themselves be judged against five standards: utility, feasibility, propriety, accuracy, and evaluation accountability in order to determine whether they generate reliable, meaningful, and defensible judgements. Applied to textbook review, a meta-evaluation perspective directs attention away from the quality of any individual textbook and toward the quality of the instruments and procedures used to assess textbooks specifically, whether the PRC and IRC tools provide clear, measurable, and properly structured criteria capable of supporting consistent

evaluation. This theoretical lens frames the entire analysis reported in this paper: the object under investigation is not the textbooks reviewed in Balochistan, but the review tools themselves.

Literature Review

The Role of Textbooks in Education

Textbooks provide the structured instructional content that guides teaching and learning in formal education systems, and in many developing-country contexts function as the primary medium through which curriculum reaches the classroom. Curriculum studies have long emphasised that textbooks form the essential link between curriculum policy and classroom practice, translating curricular objectives into materials that teachers and students engage with daily (Mikk, 2000). Because of this central role, poorly designed or inadequately reviewed textbooks can directly undermine students' understanding of subject content, making rigorous evaluation of textbook manuscripts an essential responsibility of curriculum authorities (Pingel, 2010).

Importance and Frameworks of Textbook Evaluation

Scholars of textbook evaluation (Cunningsworth, 1995; Tomlinson, 2011) argue that instructional materials should be judged according to how effectively they support learning objectives and meet the needs of learners and teachers, and that such evaluation should occur both before and during classroom use. Cunningsworth's (1995) framework emphasises the relationship between textbook content and curriculum objectives, appropriateness of language level, and the range of learning activities provided, while Littlejohn's (1998) framework stresses the importance of examining the pedagogical assumptions and types of learning tasks embedded within materials. In institutional practice, these theoretical concerns are operationalised through review tools, checklists or evaluation forms that reviewers use to assess textbook manuscripts against structured criteria, intended to standardise the review process and ensure consistency across reviewers (Pingel, 2010).

Challenges in Textbook Review Processes

Despite established evaluation frameworks, practical implementation of textbook review faces recurring challenges. A central one concerns the clarity and measurability of evaluation criteria: many review tools express criteria in general or ambiguous terms, for instance, that “language is appropriate for the grade level” or that “illustrations are attractive” that provide limited guidance for assessment and rely heavily on individual reviewer judgement (Kline, 2005). When criteria are not clearly operationalised, different reviewers may apply them inconsistently, producing variable judgements about the same manuscript, a concern closely related to the problem of inter-rater reliability (Shrout & Fleiss, 1979). A second, related challenge is that multi-stage review processes involving more than one institution can generate duplication of effort where the institutions' tools cover substantially the same ground without coordination, raising legitimate efficiency concerns even where each institution's evaluation is individually sound.

National Policy Framework for Textbook Review and Approval

Provincial review arrangements do not operate in isolation; they implement a national policy framework. The Government of Pakistan's National Textbook and Learning Materials Policy and

Plan of Action, issued by the Curriculum Wing of the Ministry of Education following the Inter-Provincial Education Ministers' Conference (Government of Pakistan, Ministry of Education, 2007), established the basic architecture within which provincial Textbook Boards operate. The Policy directs Textbook Boards to establish a review panel to examine manuscripts submitted by publishers with regard to curriculum, quality of content, presentation, language, and provincial coverage, and specifies that the review-and-revision cycle between the Textbook Board and the publisher should be completed within 30 days (Government of Pakistan, Ministry of Education, 2007). Once a Textbook Board's review committee is satisfied, the Policy directs that approved manuscripts be forwarded to the Curriculum Wing for a No Objection Certificate confirming compatibility with the National Curriculum, a step also recommended to be completed within 30 days, after which the manuscript is returned to the Textbook Board for final endorsement.

The national policy therefore envisions a single, sequential, time-bound review pathway: one substantive review at the provincial Textbook Board, followed by one compatibility check at the Curriculum Wing, rather than parallel or repeated review of the same content by multiple bodies. Read against this founding framework, the four-pass structure identified in the Balochistan review process in this paper (Section 5.1) represents a departure from, rather than a straightforward implementation of, the sequencing the national policy set out a point returned to directly in Recommendation 1b (Section 6).

Research Gap

Although considerable research addresses textbook evaluation frameworks in the abstract (Cunningsworth, 1995; Littlejohn, 1998; Tomlinson, 2011), comparatively little empirical work has examined how institutional review tools function and relate to one another in practice particularly in provincial or sub-national systems such as Balochistan's, where two separate bodies independently apply their own instruments to the same manuscripts. There is limited evidence on whether such tools are structurally and substantively duplicative, or whether they represent a legitimate division of evaluative labour, and equally limited evidence on how measurable their underlying criteria actually are. This paper addresses that gap directly by subjecting the two official Balochistan textbook review tools to systematic comparative document analysis.

Methodological Approach

This study employed a systematic comparative document analysis, treating the PRC Tool and the IRC Tool as primary source documents and subjecting each to structured examination. This approach was considered appropriate because the research question concerns the properties of the instruments themselves their structure, content, and measurability rather than their application by reviewers or their reception by stakeholders.

Data Sources and Sampling

The data source for this analysis consisted of the two official textbook review tools currently in use in Balochistan: the Provincial Review Committee (PRC) Tool, administered by the Bureau of Curriculum and Extension Centre, and the Internal Review Committee (IRC) Tool, administered by the Balochistan Textbook Board. Purposive sampling was used, as these are the only two

official instruments used in the provincial textbook review process and both were therefore included in full; no further sampling of tools was required.

Analytical Lenses

Four analytical lenses were applied sequentially to each tool:

- Structural analysis: examining format, number of sections, indicator count, and institutional labelling of each tool.
- Domain analysis: mapping each tool's indicators onto thematic evaluation domains (e.g., curriculum alignment, language quality, pedagogical activities, physical presentation, cultural and social values) and coding the degree of overlap between tools as High, Moderate, or Low.
- Measurability coding: classifying each indicator using a five-point rubric: Objective (O), Operationalised (OP), Semi-objective (SO), Subjective (S), or Highly Subjective (HS).
- Overlap and uniqueness analysis: identifying indicators shared across both tools, those unique to the PRC tool, and those unique to the IRC tool.

Measurability Coding Rubric

Code	Category	Description	Example Indicator
O	Objective	Clearly verifiable, observable indicator	“Free from typographical, grammatical and punctuation errors”
OP	Operationalised	Linked explicitly to a curriculum standard or framework	“All SLOs covered along with competencies, standards, benchmarks”
SO	Semi-objective	Interpretable, but lacks a measurable threshold	“Language appropriate to age and level”
S	Subjective	Relies primarily on personal judgement	“Language thought-provoking and stimulating”
HS	Highly Subjective	Reflects aesthetic or personal preference	“Child friendly and appealing graphical elements”

This multi-dimensional approach yields a richer characterisation of the two instruments than a simple side-by-side listing of indicators would allow, and situates the comparison within the broader literature on evaluation instrument design and textbook quality assurance discussed above.

Findings

Structural Comparison of the Two Tools

The first level of comparison examined the overall structure and format of each tool, independent of content.

Table 1: Structural Comparison of the PRC and IRC Tools

Structural Aspect	PRC Tool (Bureau of Curriculum)	IRC (Textbook Board)	Tool	Interpretation
Number of sections/categories	5 sections	6 categories		Near-identical structure
Evaluation format	Checklist: Yes/No/Comments	Checklist: Yes/No/Comments		Identical
Total indicators	~25 indicators	~29 indicators		IRC slightly more detailed
Institutional title	Bureau of Curriculum and Extension Centre, Balochistan	Balochistan Textbook Board, Quetta		Different authorities
Subject-specific version	General (all subjects)	Specific Languages	to	IRC more targeted

Both tools share a strikingly similar structural design, employing identical checklist formats with Yes/No responses supplemented by a Comments column. The PRC tool is organised into five sections while the IRC tool contains six categories, and the indicator counts are comparable (approximately 25 versus 29). One notable difference is that the IRC tool is designed specifically for the subject of Languages, whereas the PRC tool is intended for general application across subjects, meaning the IRC tool contains subject-relevant indicators that a general-purpose instrument would not include. The overall structural convergence of the two tools suggests that the two institutions drew on a common evaluation-framework template even though their instruments were developed independently, and this convergence may itself contribute to stakeholder perceptions of duplication, even where the substantive content of the criteria diverges.

Comparison of Evaluation Domains

The second level of analysis examined the thematic domains covered by each tool and the degree to which the two instruments address the same dimensions of textbook quality.

Table 2: Domain-by-Domain Comparison of the PRC and IRC Tools

Evaluation Domain	PRC	IRC	Overlap	Comparative Interpretation
Curriculum Alignment / SLO Coverage	✓	✓	High	Both tools prioritise SLO alignment; PRC dedicates a full section, IRC opens with it
Content Accuracy	✓	✓	High	Both require factually accurate, up-to-date content and illustrations
Logical/Conceptual Development	✓	Partial	Moderate	PRC has a dedicated section; IRC embeds this within content quality
Language Quality	✓	✓	High	Both assess suitability, grammar, clarity; PRC adds thought-provoking criteria
Pedagogical Activities	Limited /	✓	Low	IRC dedicates a full category to instruction; PRC coverage minimal

Instruction					
Physical / Visual Presentation	✓	✓	Moderate	Both address presentation; IRC covers layout and glossary in more detail	
Assessment Design	✓	✓	Moderate	PRC's assessment section is the most elaborated (7 indicators) vs IRC (2)	
Cultural / Social Values	Limited	✓	Low	IRC dedicates an entire category; PRC only mentions bias-free language	
Glossary	✓	✓	High	Both require a glossary; IRC adds grade-level placement guidance	
Prior Knowledge / Integration	✓	✓	Moderate	Both expect links to prior learning; PRC uses horizontal/vertical integration terms	

Areas of High Overlap

The domain comparison reveals meaningful convergence in four foundational areas: curriculum alignment and SLO coverage, content accuracy, language quality, and the requirement for a glossary. Both tools assign significant weight to ensuring that content aligns with official Student Learning Outcomes and that language is appropriate, accurate, and free of bias, a shared baseline that both institutions apply regardless of their specific institutional roles.

Areas of Divergence

The comparison also reveals clear divergence consistent with the distinct mandates of the two bodies. On cultural and social values, the IRC tool dedicates an entire category to gender bias, religious and cultural inclusivity, tolerance, and diversity, while the PRC tool contains only a single indicator on bias-free language, suggesting that the BTBB, as the body approving textbooks for classroom use across a diverse provincial population, accords greater prominence to cultural suitability than the BoC does in its curriculum-compliance role. On instructional quality, the IRC tool includes a dedicated category on quality of and support for instruction, with indicators for critical-thinking tasks, hands-on and observation-based learning, and reinforcement activities, while the PRC tool does not treat pedagogical activities as a distinct domain. Conversely, on assessment design the PRC tool is considerably more elaborated, with seven detailed indicators on command words, cognitive levels, assessment types, and SLO alignment, compared to the IRC tool's two-indicator treatment reflecting the BoC's stronger institutional emphasis on ensuring textbook assessment aligns with national curriculum frameworks. Together, these differences suggest the PRC tool prioritises curriculum compliance and assessment design, while the IRC tool prioritises instructional design, cultural appropriateness, and physical presentation, a de facto, if informal, division of quality-assurance labour between the two institutions.

Measurability of Evaluation Criteria

The third level of analysis classified each indicator in both tools according to the five-point measurability rubric described in Section 3.2.

Table 3: Measurability Coding of PRC and IRC Tool Indicators

Code	Category	PRC Frequency	IRC Frequency	Example Indicator
O	Objective	3	4	“Free from typographical, grammatical and punctuation errors”
OP	Operationalised	7	5	“All SLOs covered along with competencies, standards, benchmarks”
SO	Semi-objective	9	12	“Language appropriate to age and level”
S	Subjective	4	6	“Language thought-provoking and stimulating”
HS	Highly Subjective	2	2	“Child friendly and appealing graphical elements”
	TOTAL	25	29	

In the PRC tool, semi-objective indicators account for approximately 36% of all criteria (9 of 25), while subjective and highly subjective criteria together account for approximately 24% (6 of 25); objective or operationalised criteria make up the remaining 40%. In the IRC tool, semi-objective indicators are even more prominent at approximately 41% (12 of 29), with subjective and highly subjective criteria at approximately 28% (8 of 29) and only 31% (9 of 29) clearly objective or operationalised. Indicators such as “language appropriate to age and level”, “child friendly and appealing graphical elements”, and “language thought-provoking and stimulating” lack clear measurable thresholds, leaving their interpretation open to individual reviewer judgement.

This has direct implications for inter-rater reliability: when indicators lack operationalised thresholds, different reviewers may apply the same criterion differently, producing varying outcomes for identical manuscripts — a risk compounded in a multi-stage process where the same or similar indicators are applied by different reviewers across two separate institutions. The PRC tool's relatively higher share of operationalised indicators, particularly within its assessment section, reflects the BoC's stronger grounding in formal curriculum frameworks, while the IRC tool's greater share of semi-objective indicators clusters in its categories on physical presentation, instructional activities, and cultural values — dimensions that are inherently harder to measure with precision and may require professional judgement as an irreducible element of quality assessment.

4.4 Degree of Overlap and Uniqueness Between the Tools

The fourth level of analysis examined the extent to which indicators in the two tools overlap or are unique to each institution.

Table 4: Shared and Unique Indicators Across the PRC and IRC Tools

Shared Indicators (Both Tools)	Unique to PRC (Bureau of Curriculum)	Unique to IRC (Textbook Board)
SLO / curriculum coverage	Horizontal and vertical integration of concepts	Interactive questions for critical thinking
Content accuracy and currency	Life-skills integration and stimulation of curiosity	Hands-on / observation-based learning tasks
Language clarity and grade-level suitability	Missing SLOs explicitly identified	Extension activities for reinforcement
Glossary present	Assessment: appropriate command words	Graphical placement on the same page as content
Prior knowledge referenced	Assessment: appropriate number/length of items	Full category on cultural and social values
Assessment aligned with learning outcomes	Author/editor credentials included	Pronunciation and oral communication activities All text types included — narrative, persuasive, expository

Approximately 40–45% of the evaluation criteria overlap between the two tools, clustered predominantly around SLO and curriculum coverage, content accuracy, language quality, and basic assessment alignment — representing a minimum quality baseline both institutions consider essential regardless of their specific role. The unique indicators are equally revealing of institutional identity: the PRC tool uniquely emphasises horizontal and vertical integration of concepts, explicit identification of missing SLOs, author/editor credentials, and detailed assessment-quality criteria, consistent with the BoC's curriculum-compliance mandate; the IRC tool uniquely addresses oral communication and pronunciation activities, a full range of text types, detailed graphical-placement criteria, reinforcement activities, and a comprehensive set of cultural and social-values indicators, reflecting the BTBB's focus on classroom usability and social suitability. This estimated 40–45% overlap suggests the tools are genuinely complementary rather than redundant as complete instruments, although the shared indicators do represent an area where both institutions currently expend evaluation effort on the same quality dimensions without apparent coordination.

Summary of Findings

Finding 1 Structural Identity: Both tools share a structurally identical format; a checklist-based instrument using Yes/No responses and a comments column; indicating a comparable evaluation-framework template. This structural convergence may contribute to perceptions of duplication even where substantive content differs significantly.

Finding 2 Complementary Emphasis: Despite meaningful overlap in foundational domains, the tools diverge considerably in emphasis: the PRC tool prioritises curriculum compliance and assessment design, while the IRC tool weights instructional quality, physical presentation, and cultural appropriateness more heavily, consistent with the distinct mandates of the BoC and the BTBB.

Finding 3 Measurability Limitations: Both tools contain a substantial proportion of semi-objective and subjective indicators (approximately 60% in the PRC tool and 69% in the IRC tool), indicating that the reliability of the review process is potentially compromised by the absence of clearly operationalised, measurable criteria in both instruments.

Finding 4 Overlapping but Not Identical Content: The estimated overlap in specific indicators is approximately 40–45%, indicating that the two tools are not simple duplicates of one another at the level of content though this figure also means that close to half of each tool's indicators are shared, and their near-identical structural format may compound, rather than merely obscure, stakeholder perceptions of duplication. Whether this level of content overlap translates into unnecessary duplication of effort in practice cannot be settled by document analysis alone (see Section 5.1).

Discussion and Implications

Implications for the Multi-Stage Review Process

The existence of two structurally similar tools does not, by itself, confirm that the multi-stage review process involves problematic duplication. Rather, the PRC and IRC tools appear to serve distinct but complementary functions: the BoC evaluates textbooks primarily for curriculum compliance and conceptual development, while the BTBB evaluates them for instructional quality, presentation, and cultural suitability. This de facto division of quality-assurance labour may be appropriate and even beneficial, provided it is explicitly recognised, formally coordinated, and communicated to textbook authors and publishers. However, the indicators that are shared between the tools particularly those on curriculum alignment, content accuracy, and language quality do represent areas of genuinely overlapping evaluation. If reviewers from both institutions assess the same indicators without coordination, the process may generate redundant effort without a commensurate improvement in textbook quality in those specific areas.

It is also worth noting that a manuscript does not, in practice, pass through the PRC and IRC tools only once each. Each institution applies its own tool across two internal passes – an initial review followed by a select review meaning a manuscript is reviewed up to four times (IRC review, IRC select review, PRC review, PRC select review) before approval, using closely related versions of the same two underlying instruments. This four-pass structure was not itself the object of the document-comparison method used in this study, which examined the PRC and IRC tools as two documents rather than as four applied review events, but it is a relevant piece of context for interpreting the practical significance of the 40–45% indicator overlap reported in Section 4.4: the same overlapping content may, in effect, be reviewed more than twice before a manuscript is approved.

Implications for Evaluation Instrument Design

The high proportion of subjective and semi-objective indicators in both tools raises broader concerns for evaluation instrument design. Evidence from other quality-assurance contexts indicates that inter-rater reliability improves markedly when criteria are operationalised with explicit, observable thresholds and illustrative examples (Kline, 2005; Shrout & Fleiss, 1979). The current instruments' reliance on open-ended professional judgement for a majority of indicators risks inconsistent outcomes across reviewers and across institutions, which in turn

undermines the fairness and credibility of the review process from the perspective of textbook authors and publishers. That the PRC tool already achieves a relatively higher share of operationalised indicators, particularly in its assessment section, demonstrates that operationalisation is achievable for this class of criteria and could reasonably be extended across both tools.

Implications for Institutional Coordination

The finding that approximately 40–45% of indicators overlap between the two tools, combined with their near-identical structural format, suggests real value in formally mapping the relationship between the two instruments and making it explicit in procedural guidance for textbook review. A formal memorandum of understanding or joint evaluation framework between the BoC and the BTBB could clarify which aspects of textbook quality are reviewed by which institution, reducing duplication of effort on shared indicators while preserving the complementary functions each tool currently serves.

Findings in Light of the Program Evaluation Standards

Section 1.4 grounded this study in the Program Evaluation Standards (Yarbrough et al., 2011), which judge an evaluation system against five criteria: Utility, Feasibility, Propriety, Accuracy, and Evaluation Accountability. Because the present study relied exclusively on comparative document analysis of the two tools themselves, it is best positioned to speak directly to only some of these standards; the remainder would require data this study did not collect, such as interviews with reviewers or observation of the review process in action. Table 5 maps the findings reported above onto the five standards and is explicit about where the evidence is strong and where it is necessarily limited.

Table 5: Findings Mapped Against the Program Evaluation Standards

Standard	What It Asks	What This Study Found	Evidential Strength
Utility	Do the tools' outputs inform decisions that actually improve textbooks?	The 40–45% indicator overlap and the tools' complementary domain emphases (Section 4.2, 4.4) suggest each tool generates some non-redundant, potentially decision-relevant information; whether reviewers and authors actually act on that information cannot be confirmed from document analysis alone.	Indirect / partial
Feasibility	Are the tools practical to administer given time, cost, and reviewer capacity?	The near-identical checklist format and comparable indicator counts (~25 vs ~29; Table 1) suggest both tools impose broadly similar administrative burden and are procedurally straightforward to apply.	Moderate — inferable from structure

Propriety	Is the review conducted fairly, ethically, and with due regard for those affected?	This study's document-only method cannot assess how reviewers are trained, how conflicts of interest are managed, or how authors experience the process; no claim is made on this standard.	Not assessed — outside method's scope
Accuracy	Do the criteria produce valid, reliable, defensible judgements?	This is the standard the study addresses most directly. The measurability coding (Table 3) shows 60% of PRC indicators and 69% of IRC indicators are semi-objective or subjective, indicating a meaningful risk to reliable, defensible judgement in both tools.	Strong — primary focus of the analysis
Evaluation Accountability	Is the evaluation process itself documented, transparent, and open to review?	The absence of any formal, published mapping between the PRC and IRC tools' respective roles (noted in Section 5.3) suggests the division of labour between institutions is not currently documented or accountable in a way stakeholders can inspect.	Moderate — inferable from absence of coordination documents

Read against this framework, the study's central contribution is strongest on Accuracy and offers reasonable, if indirect, evidence on Feasibility, Utility, and Evaluation Accountability. It makes no claim regarding Propriety, which would require a different method most plausibly, interviews or surveys with reviewers, authors, and institutional staff. This is consistent with the delimitations set out in Section 7: the tools were the unit of analysis, not the people who apply them or are affected by them.

Recommendations

Recommendation 1: Develop a Joint Indicator Mapping Framework

The BoC and BTBB should commission a formal joint mapping exercise to identify all shared, complementary, and unique indicators across the PRC and IRC tools, and use the resulting map to inform a revised division of responsibilities that formally designates which quality dimensions are owned by which body.

Recommendation 1a: Restructure the Tools Around Institutionally Exclusive Sections

Building on the indicator map produced under Recommendation 1, the paper recommends a more structural reform: that the PRC and IRC tools be reorganised so that each contains a set of sections addressed and reviewed exclusively by one institution, with no indicator appearing in both tools' exclusive sections. Concretely, this would mean:

- A BoC-exclusive section (or sections) covering curriculum alignment, SLO coverage, conceptual progression, and assessment design the domains where Section 4.2 shows the

PRC tool already carries the greater share of detailed, operationalised indicators reviewed only by the Bureau of Curriculum and not re-examined by the Textbook Board.

- A BTBB-exclusive section (or sections) covering instructional and pedagogical activities, physical and graphical presentation, and cultural and social values the domains where Section 4.2 shows the IRC tool carries the greater share of detailed indicators reviewed only by the Balochistan Textbook Board and not re-examined by the Bureau of Curriculum.
- A small, jointly owned section limited to the minimum baseline indicators that both institutions currently treat as essential (content accuracy, basic language correctness, and glossary presence; see Section 4.4), reviewed once, with the result shared between institutions rather than assessed twice.

This restructuring directly targets the 40–45% indicator overlap identified in Section 4.4 and the four-pass review structure noted in Section 5.1: rather than relying on informal convergence of institutional focus, it would make the division of responsibility an explicit property of the tools themselves, so that no author or reviewer encounters the same substantive indicator being assessed independently at more than one stage of the process. Because this recommendation would change the content of the instruments rather than simply their coordination, it should be piloted on a limited set of manuscripts and evaluated for its effect on review turnaround time and reviewer workload before being adopted province-wide.

Recommendation 1b: Establish a Single Sequential Review Pathway Coordinated by a Desk Officer

Recommendation 1a addresses what each institution should review; this recommendation addresses how a manuscript should move between them. Rather than the current arrangement, in which a manuscript in effect circulates through four review events (Section 5.1) with limited formal handover between them, the paper recommends a single, sequential review pathway with one point of institutional accountability at each step:

- A manuscript is first received and logged by a designated desk officer, who acts as the single point of custody and tracking for the manuscript as it moves between institutions, records the date and status of each stage, and is responsible for ensuring no stage is skipped or unnecessarily repeated.

This pathway is intended to convert the current four-pass structure into a single pass through each institution's exclusive domain, coordinated through one administrative point of contact. It would also guard against a plausible risk of the present arrangement: where the two institutions reach different judgements on the same or related aspects of a manuscript at different, uncoordinated stages, an author may be required to revise the same material more than once in contradictory directions. Routing the manuscript once through each institution's exclusive section, with a single consolidated report, would remove the structural opportunity for this kind of conflicting, duplicated instruction to arise. Notably, this recommendation is not a departure from established policy but a return to it: the Government of Pakistan's National Textbook and Learning Materials Policy (Government of Pakistan, Ministry of Education, 2007) already envisions exactly this kind of single, sequential, time-bound pathway a substantive review at the Textbook Board followed by a compatibility check at the Curriculum Wing, each within a recommended 30-day window rather than the repeated, uncoordinated review that has developed

in provincial practice. As with Recommendation 1a, this is a procedural change with resourcing implications most notably the designation and training of desk officer staff and should be piloted before wider adoption.

Recommendation 2: Operationalise Semi-Objective and Subjective Indicators

Both institutions should undertake a systematic operationalisation exercise in which each semi-objective and subjective indicator is rewritten to include a precise definition of the quality dimension being assessed, two to three observable and measurable criteria for satisfactory performance, and illustrative examples drawn from actual manuscripts. For instance, the indicator “language is suitable for the grade level” could be rewritten as “sentence length and vocabulary correspond to the recommended readability level for the specified grade”. This exercise should involve reviewers, curriculum specialists, and, where possible, textbook authors and publishers.

Recommendation 3: Introduce Inter-Rater Reliability Protocols

Given the high proportion of subjective indicators in both tools, both institutions should implement structured inter-rater reliability protocols, including calibration exercises in which reviewers independently evaluate the same manuscript and then compare and reconcile their judgements – a standard practice in other high-stakes evaluation contexts (Kline, 2005; Shrout & Fleiss, 1979).

Recommendation 4: Develop Subject-Specific Tool Variants

The IRC tool's language-specific design is a positive precedent for subject-sensitive evaluation. Both institutions should consider developing subject-specific tool variants for core areas such as Mathematics, Science, and Social Studies, ensuring criteria are calibrated to the distinctive quality dimensions of each subject rather than relying solely on a general-purpose instrument.

Recommendation 5: Institutionalise Periodic Review of the Tools Themselves

Both the PRC and IRC tools should be formally reviewed on an annual or biennial cycle to ensure their criteria remain aligned with updated national curriculum frameworks and evolving pedagogical and social standards, ideally overseen by a standing joint committee drawn from both institutions and independent curriculum experts.

Recommendation 6: Strengthen Cultural and Social Values Criteria in the PRC Tool

In light of the IRC tool's comprehensive treatment of cultural and social values, the PRC tool should be expanded beyond its single bias-free-language indicator to encompass gender equity, religious and cultural inclusivity, tolerance, and diversity — ensuring that curriculum-compliance review also incorporates social-values dimensions increasingly recognised as central to textbook quality.

Delimitations

This paper is delimited to the review tools themselves and therefore does not report on how reviewers apply these criteria in practice within written review reports, nor on how stakeholders perceive the review process; these are separate empirical questions requiring different data

sources and methods, and were not part of the present analysis. The analysis was further limited to the two official review tools currently used by the Bureau of Curriculum and the Balochistan Textbook Board; other aspects of textbook development, such as writing, editing, printing, and classroom implementation, fall outside its scope. Because the IRC tool examined here is the version specific to Languages, findings regarding its subject-specificity may not generalise to other subject areas, and because the study is confined to the province of Balochistan, its findings may not be directly generalisable to textbook review systems used elsewhere in Pakistan or in other educational contexts.

A further delimitation follows directly from the theoretical framework adopted in Section 1.4. Judged against the five Program Evaluation Standards (Yarbrough et al., 2011), this study speaks with confidence only to Accuracy, and offers indirect, structurally inferred evidence on Feasibility, Utility, and Evaluation Accountability (Section 5.4). It does not address Propriety whether the review process is conducted fairly and ethically toward reviewers, authors, and other participants because comparative document analysis of the tools themselves cannot generate evidence about how those tools are actually administered or experienced. Assessing Propriety, and strengthening the evidence base for Utility, Feasibility, and Evaluation Accountability beyond inference, would require a different method: interviews or surveys with reviewers, curriculum authors, and institutional staff, together with direct observation of the review process. This limitation is deliberate rather than incidental it reflects the study's specific focus on the properties of the instruments rather than on their application but it means the findings should not be read as a complete meta-evaluation of the Balochistan textbook review system in the full sense intended by the Program Evaluation Standards. In particular, this study cannot determine from document analysis alone whether the four-pass review structure noted in Section 5.1, or the coordination and reviewer-consistency practices that surround it, add proportionate value at each stage; that determination would require reviewer- and process-level data beyond the scope of this paper.

Conclusion

This paper has presented a systematic comparative analysis of the PRC and IRC textbook review tools used by the Bureau of Curriculum and the Balochistan Textbook Board respectively. The analysis demonstrates that the two tools, while structurally identical in format, serve meaningfully different and largely complementary functions within the provincial textbook quality-assurance system: the PRC tool prioritises curriculum compliance, conceptual development, and assessment design, while the IRC tool prioritises instructional quality, cultural appropriateness, and physical presentation. An estimated 40–45% of evaluation indicators overlap between the two instruments, while the remaining 55–60% represent unique quality dimensions that each institution addresses independently.

At the same time, the analysis reveals a significant and shared vulnerability in both instruments: a preponderance of semi-objective and subjective indicators that lack clearly operationalised, measurable thresholds. This characteristic poses a risk to the consistency, reliability, and fairness of the review process and represents the most urgent target for reform. Combined with the absence of formal coordination mechanisms between the two institutions, it creates conditions in which the multi-stage review process may be experienced as duplicative and unpredictable by textbook stakeholders, even where the substantive quality-assurance functions of the two bodies

are genuinely distinct. The recommendations advanced in this paper, centred on indicator operationalisation, inter-rater reliability protocols, joint instrument mapping, and formal institutional coordination offer a practical pathway toward a more coherent, efficient, and credible textbook quality-assurance system in Balochistan, one that preserves the complementary strengths of both the PRC and IRC review processes while addressing the measurability limitations that currently undermine their reliability and perceived legitimacy.

It is important to be precise about what this conclusion does and does not claim. The finding that the PRC and IRC tools are complementary rather than duplicative concerns instrument content specifically, and rests on comparative document analysis alone. It does not establish that the review process as a whole is efficient. As Section 5.1 notes, a manuscript in practice passes through up to four review events, not two, and the 40–45% of indicator content that is shared between the tools may in effect be assessed more than twice before approval. Whether this repetition constitutes wasteful duplication of effort, or a defensible degree of cross-checking, cannot be determined from a comparison of the tools' written content alone; it would depend on factors this study did not measure, such as how consistently the shared indicators are interpreted across review passes and how efficiently the process as a whole is coordinated between institutions. The claim this paper is entitled to make is therefore a narrower one: the tools themselves are not simple duplicates of one another. Whether the process built around them is redundant remains an open question that the recommendations in Section 6 particularly the proposal for institutionally exclusive review sections are designed to address regardless of how that question is ultimately resolved.

Funding: This research was funded by the Research Center of the Balochistan Textbook Board.

References

1. Balochistan Textbook Board. (n.d.). Internal Review Committee (IRC) tool: Review criteria for languages. Balochistan Textbook Board.
2. Bureau of Curriculum and Extension Centre, Balochistan. (n.d.). Provincial Review Committee (PRC) tool for textbook evaluation. Bureau of Curriculum.
3. Cunningsworth, A. (1995). *Choosing your coursebook*. Heinemann.
4. Government of Pakistan, Ministry of Education. (2006). *National curriculum for languages (English), Grades I–XII*. Ministry of Education.
5. Government of Pakistan, Ministry of Education, Curriculum Wing. (2007). *National textbook and learning materials policy and plan of action*. Curriculum Wing, Ministry of Education.
6. Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation*. SAGE Publications.
7. Littlejohn, A. (1998). The analysis of language teaching materials: Inside the Trojan horse. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 190–216). Cambridge University Press.
8. Mikk, J. (2000). *Textbook: Research and writing*. Peter Lang.
9. Pingel, F. (2010). *UNESCO guidebook on textbook research and textbook revision* (2nd ed.). UNESCO.
10. Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428.

11. Stufflebeam, D. L. (2001). The metaevaluation imperative. *American Journal of Evaluation*, 22(2), 183–209.
12. Tomlinson, B. (Ed.). (2011). *Materials development in language teaching* (2nd ed.). Cambridge University Press.
13. Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Corwin Press.