SOCIAL PRISM

Volume 3(6), 2026, 108-124
Online ISSN: 3006-8886
Print ISSN: 3006-8878

Beyond the Review Tool: A Content Analysis of Narrative Review Reports Produced by the Balochistan Textbook Board

Dr. Mehwish Malghani[1], Dr. Gulab Khan Khilji[2], Fareed Ahmed Buzdar[3]

[1]*Professor, Department of English, Sardar Bahadur Khan Women's University (SBKWU), Balochistan, Pakistan, Email: dr.mmalghani@gmail.com*
[2]*Chairman, Balochistan Textbook Board, Balochistan, Pakistan*
[3]*Subject Specialist, Balochistan Textbook Board, Balochistan, Pakistan*

Abstract

Textbook manuscripts submitted to the Balochistan Textbook Board (BTBB) are evaluated by panel of experts whose judgment is recorded largely as narrative, free-text comments in addition to the BTBB review tool. Though the review tools used in this process have received a little attention in earlier work, the descriptive comments themselves given by the reviewers have not been studied directly. This current paper reports a qualitative content analysis of the descriptive reports produced by the reviewers. A total of 574 individual reviewer comments extracted from eleven BTBB Internal Review Committee (IRC) reports of English, Mathematics, and Computer Science manuscripts across Grades 9 to 12, were analysed. The comments were coded under four dimensions namely subject-matter focus, objectivity, evidence base, and actionability. Analysis of the data shows that curriculum and SLO alignment along with general content accuracy are leading concerns of reviewers with (23%) and (23%) values respectively. Additionally it was found that purely subjective, unjustified comments are rare (0.2%), but that a large majority of comments (74%) are semi-objective and only 22% cite a specific, verifiable reference. The third variable, actionability, also shows noticeable differences depending on how the report is structured. Reports that present findings alongside their recommendations tend to provide clearer and more practical guidance than those that list findings and recommendations separately. Overall, the consistency of the reports seems to be influenced more by the subject area and the format of the report than by individual reviewers. The paper presents recommendations for strengthening and improving evidence-based BTBB review practice and suggests recommendations for further research.

Keywords: Textbook review, content analysis, Balochistan Textbook Board, review reports, reviewer feedback, quality assurance, curriculum alignment

Introduction

Textbook review is among the least visible yet most consequential stages of curriculum implementation. The manuscript passes through a thorough review process before its practical use in the classroom. During the process the reviewers catch what is wrong and to tell the author what should change. In the entire process in Balochistan there is a huge responsibility on Balochistan Textbook Board (BTBB) to conduct an internal review of every textbook manuscript submitted for approval.

Received 04 July 2026; Revised 25 July 2026; Accepted 05 August 2026; Published (online) 09 August 2026

Corresponding Author Email: dr.mmalghani@gmail.com
DOI: https://doi.org/10.69671/ socialprism.3.6.2026.162

Online ISSN: 3006-8886
Print ISSN: 3006-8878

Reviewers are required to evaluate the manuscripts on the basis of BTBB Review Tool and to submit a descriptive report too. The narrative report submitted by the reviwers plays a crucial role as it is based on narrative comments referring to specific pages, lessons, or sections of the manuscript rather than as a simple yes or no on a form.

These narrative comments carry a lot of responsibility as translating them can actually transform the only teaching ais used in most of the government schools i.e. book. A checked box tells an author that something is wrong; the comment next to it is supposed to tell them what, and why. A simple comment that “the language is not appropriate” leaves the author guessing at what standard was applied and what change would satisfy it. A remark that “the vocabulary in this passage exceeds the Grade 4 word list specified in the curriculum” helps and guides the author something concrete to act on. In other words, the quality of the review process, depends much on the quality of what reviewers write in addition to the checklist.

Despite the crucial role reviewers’ comments play, they are rarely examined on their own terms. Existing discussions of textbook review in Pakistan tend to focus on the review tools themselves. The discussion mostly focuses on broad indicators such as approval timelines and comparatively little attention has been paid to what reviewers actually write, what issues are raised by them most often, how well their comments are supported by evidence or a clear standard, and whether the recommendations that follow give authors a workable path to revision. The present study looks directly at that gap by examining a sample of narrative review reports produced by BTBB reviewers across three subjects, English, Mathematics, and Computer Science, and four grade levels.

Statement of the Problem

At the BTBB, reviewers assess textbook manuscripts using a structured review checklist. However, the feedback that authors actually rely on during revision comes mainly from the narrative comments written by reviewers rather than from the checklist itself. These comments differ considerably in their quality and level of detail. Some clearly identify specific, verifiable issues and explain what needs to be corrected. Others are much more general, using remarks such as *"needs improvement"* or *"not suitable"* without explaining what the problem is, why it matters, or how it can be addressed.

When comments lack sufficient detail, authors may find it difficult to determine what revisions are expected. As a result, they may interpret the feedback differently or make changes that do not fully address the reviewers' concerns. At the same time, reviewers may focus on different issues from one manuscript to another, reducing consistency across the review process. In such situations, the review system may not achieve its full potential as a quality assurance mechanism, even if the checklist itself is comprehensive and well designed.

Despite the central role that reviewer comments play in textbook revision, little is known about their overall quality. There has been no systematic investigation of how frequently BTBB review reports contain clear, evidence-based, and actionable feedback compared with comments that are vague, unsupported, or difficult for authors to act upon. Similarly, it remains unclear which aspects of textbook manuscripts receive the greatest attention from reviewers and whether patterns differ across subject areas.

Online ISSN: 3006-8886
Print ISSN: 3006-8878

To address this gap, the present study analyses a sample of BTBB narrative review reports. The aim is to examine the nature of reviewers' written comments, distinguish between actionable and non-actionable feedback, and identify the textbook features and subject areas that attract the most reviewer attention. By providing an evidence-based picture of current review practices, the study seeks to contribute to improving the consistency, usefulness, and overall effectiveness of the BTBB manuscript review process.

Objectives of the Study

General Objective

To examine the characteristics and quality of the narrative comments that BTBB reviewers produce during the textbook review process.

Specific Objectives

1. To identify the types of issues most frequently raised in BTBB narrative review comments.
2. To assess the extent to which these comments are objective and supported by evidence or a stated standard, as opposed to general impression.
3. To evaluate how specific and actionable the recommendations contained in these comments are.
4. To examine how consistent BTBB reviewers are in the depth and quality of the comments they write, across reports, subjects, and grade levels.

Research Questions

Main Research Question

What are the characteristics and quality of the narrative review comments produced by BTBB reviewers?

Sub-Questions

1. What types of issues do BTBB reviewers most frequently identify in their narrative comments?
2. To what extent are BTBB reviewers' comments objective and evidence-based rather than based on general impression?
3. How specific and actionable are the recommendations contained in BTBB reviewers' comments?
4. How consistent are BTBB reviewers in the depth and quality of the comments they write across different reports, subjects, and grade levels?

Significance of the Study

This study contributes to a small but growing body of work that treats textbook review not just as a checklist exercise but as an act of professional judgment extending evaluation theory to a type of document, the narrative reviewer comment, that has received little empirical attention in the Pakistani context specifically.

Online ISSN: 3006-8886
Print ISSN: 3006-8878

On a practical level, the findings can give BTBB a clearer, evidence-based picture of where reviewer feedback is already strong and where it tends to fall back on vague or unreferenced language. If, for instance, only a minority of comments turn out to be clearly evidence-based, that is useful information for designing reviewer training or revising internal guidance on how comments should be written and referenced.

For textbook authors, understanding the patterns in the kind of feedback BTBB reviewers tend to give could make it easier to anticipate common concerns and address them before submission, shortening revision cycles.

Finally, the coding framework developed for this study, though built around BTBB's own reports, is written in a way that could be adapted later by this researcher or others to examine review reports produced by other provincial or federal textbook boards, once such a comparison becomes feasible.

Delimitations of the Study

1. This study is limited to narrative review reports produced by the Balochistan Textbook Board. Review reports produced by the Bureau of Curriculum fall outside the scope of this analysis, and no direct comparison between the two institutions' reports is attempted here.
2. The study examines what reviewers wrote, not whether their underlying judgment about the textbook was correct. It evaluates the quality of the review comments themselves, not the quality of the textbooks being reviewed.
3. The analysis is based on a purposive sample of eleven BTBB review reports covering English, Mathematics, and Computer Science manuscripts for Grades 9 through 12, rather than the Board's entire historical archive or all subjects and grade levels it reviews.
4. The study is a document-based content analysis. It does not include interviews or surveys of reviewers about their intentions or reasoning; conclusions are drawn from the text of the reports as written.
5. The coding framework developed here is a research instrument built for the purposes of this study. Developing or formally validating a scoring rubric for institutional adoption by BTBB is beyond the scope of this study, though the findings may inform such an effort in the future.

Literature Review

Textbook Evaluation as Educational Quality Assurance

The international literature generally views textbook evaluation as a continuous quality assurance process rather than a single review conducted just before publication. One of the earliest and most influential contributions was made by Cunningsworth (1995), who proposed a systematic framework for evaluating coursebooks. His framework encourages reviewers to examine areas such as learning objectives, language content, teaching methodology, and the balance of language skills and classroom activities. By providing clear evaluation criteria, it helps reduce reliance on personal opinion and makes the review process more consistent and transparent.

Online ISSN: 3006-8886
Print ISSN: 3006-8878

Building on this perspective, McDonough, Shaw, and Masuhara (2013) argue that textbook evaluation should not be limited to a single stage. They describe it as an ongoing process that includes pre-use, in-use, and post-use evaluation. According to their model, materials should be assessed before they are introduced in the classroom, monitored while they are being used, and reviewed again after implementation to determine how effectively they support teaching and learning. This broader view highlights that textbook evaluation continues throughout the life of a textbook rather than ending with its approval.

This perspective is particularly relevant to the present study because the Balochistan Textbook Board (BTBB) conducts reviews at the pre-use stage. At this point, reviewers examine manuscripts before they are published or used in schools. Consequently, their comments focus on aspects that can be judged from the manuscript itself, such as content, language, organization, alignment with curriculum requirements, and pedagogical suitability. Since these review reports often determine whether a manuscript is revised or approved, the quality and clarity of reviewers' written feedback become especially important. Understanding the nature of these comments is therefore essential for evaluating how effectively the BTBB review process contributes to textbook quality assurance.

Objectivity and Evidence in Reviewer Judgment

A recurring theme in the textbook evaluation literature is the tension between criterion-based judgment and personal opinion. Sheldon (1988), in an early and still widely cited argument, insisted that textbook evaluation should rest on explicit, stated criteria rather than a reviewer’s general impression, precisely because vague criteria produce feedback that different reviewers apply inconsistently and that authors struggle to act on.

The distinction between objective and subjective judgment is central to the coding framework used in this study. BTBB reviewers often provide comments that range from clear, evidence-based observations to more general opinions. To examine these differences systematically, the present study classifies reviewer comments according to the extent to which they are supported by specific evidence and provide actionable guidance for authors.

Ellis (1997) offers a useful perspective by distinguishing among predictive, in-use, and retrospective evaluation. Predictive evaluation takes place before a textbook is introduced in the classroom. At this stage, reviewers cannot rely on evidence from actual teaching and learning because the material has not yet been used. Instead, their judgments depend largely on professional expertise and the reasoning they provide to support their recommendations. This makes the written justification in a review report especially important, as it often serves as the only documented explanation for why a particular revision has been requested.

Tomlinson (2013) and the recent review by Gholampour and Mehrabi (2023) show that textbook evaluation frameworks have become increasingly comprehensive over time. However, they also point out that most existing frameworks focus on evaluating the quality of the textbook itself rather than assessing the quality of the reviewers' written feedback. In other words, they provide guidance on **what** should be evaluated in a textbook but pay far less attention to how reviewers communicate their judgments. This represents an important gap in the literature. The present study addresses this gap by examining the quality, clarity, and actionability of reviewer comments, rather than evaluating the textbook manuscripts alone.

Online ISSN: 3006-8886
Print ISSN: 3006-8878

Textbook Evaluation Research in Pakistan and Balochistan

In Pakistan, one of the most influential studies on textbook evaluation was conducted by Aftab (2012). Her work provides a comprehensive discussion of English language textbook evaluation by adapting well-established international evaluation frameworks to the Pakistani educational and policy context. The study has served as an important theoretical foundation for many later studies on textbook evaluation in Pakistan and highlights the need for systematic and context-sensitive approaches to assessing teaching materials.

A study that is particularly relevant to the present research was carried out by Ahmad and colleagues (2016), who evaluated English textbooks used in Balochistan from the perspective of primary school teachers. Using both quantitative and qualitative methods, they examined how effectively these textbooks met classroom needs and supported teaching and learning. Their findings provide valuable insights into the strengths and weaknesses of the textbooks in actual classroom settings. However, their focus remained on the textbooks after they had been published and implemented rather than on the review process that took place before publication. Consequently, the study does not explore how reviewers reached their decisions or the quality of the written comments that guided textbook revisions.

Taken together, the available Pakistani literature demonstrates that textbook quality is an important and well-established area of educational research. However, most studies have concentrated on evaluating the textbooks themselves or investigating the experiences of teachers and students who use them in classrooms. Very little attention has been given to the institutional processes through which textbooks are reviewed before publication, particularly in Balochistan. Research examining the quality, consistency, and usefulness of reviewers' written comments is especially limited.

This lack of evidence represents an important gap in the literature. The present study seeks to address this gap by focusing on the narrative review reports produced during the Balochistan Textbook Board (BTBB) manuscript review process. Rather than evaluating the textbooks themselves, it examines the nature of reviewers' written feedback and its role in supporting effective textbook revision and quality assurance. It should also be acknowledged that published research specifically on Balochistan's textbook review system remains scarce. Therefore, this review draws on the strongest available national and international literature while contributing new evidence from a context that has received relatively little scholarly attention.

Policy Context: The National Curriculum of Pakistan

Because curriculum and SLO alignment turned out, in this study's results, to be the single largest category of reviewer comment, the policy documents that define those SLOs are as relevant to this review as the academic literature above. Pakistan's curriculum reform since 2020 has moved toward a standards- and outcomes-based model: the Ministry of Federal Education and Professional Training's National Curriculum for English Language, Grades I–XII (2020) sets out student learning outcomes grade by grade, explicitly so that "teacher and learners can use [them] to gauge their progress towards the benchmarks," and the National Curriculum Council's broader National Curriculum Framework (Ministry of Federal Education and Professional Training, 2017) sets the policy expectation that textbooks and review processes across provinces be built around exactly this kind of stated, checkable standard rather than general impression. The Math-

Online ISSN: 3006-8886
Print ISSN: 3006-8878

ematics reports examined in this study make the same expectation explicit at the level of individual practice, with reviewers citing the "National Curriculum 2022–23" by name when checking whether SLOs are fully addressed. Read together, these documents mean that BTBB reviewers are not inventing the objective, standard-referenced style of comment called for in the literature it is already the model the national curriculum policy asks them to apply. The gap this study documents, in other words, is less a gap in guidance than a gap between the standard the policy sets and how consistently reviewer comments cite it in practice.

Research Gap

Taken together, the literature offers well-developed frameworks for evaluating textbooks and a reasonably clear theoretical case for why review criteria should be explicit and evidence-based rather than impressionistic. What it does not offer, either internationally or within Pakistan, is a close empirical look at the actual text of review reports at what reviewers write, how specific it is, and how much of it authors can realistically act on. No published study has examined the narrative review reports produced by BTBB in this way, across multiple subjects and grade levels. This study addresses that gap directly, applying the objectivity and evidence-based criteria established in the wider literature and in national curriculum policy to a body of text that has not previously been analysed: BTBB's own review reports.

Theoretical Framework

This study is grounded in Hattie and Timperley's (2007) model of feedback effectiveness, set out in their widely cited synthesis "The Power of Feedback." The model was developed for classroom feedback to students, but its central claims translate directly to the setting examined here, since a BTBB reviewer's comment on a manuscript is, structurally, the same kind of act as a teacher's comment on a student's work: a piece of expert feedback intended to close the gap between where the work currently stands and where it needs to be.

Hattie and Timperley argue that effective feedback answers three questions for its recipient Where am I going? (the goal or standard being applied), How am I going? (a judgment of current performance against that goal, ideally with evidence), and Where to next? (what specific action would close the gap). Feedback that leaves any of these three questions unanswered, in their account, is markedly less useful to the person receiving it, regardless of how accurate the underlying judgment is. This maps onto the present study's coding framework in a direct and motivated way, rather than as an after-the-fact fit: the objectivity and evidence-base dimensions capture whether a BTBB comment answers "how am I going" with a stated, checkable standard rather than a bare impression, and the actionability dimension captures whether it answers "where to next" with a specific step rather than a general direction.

Hattie and Timperley also distinguish levels at which feedback can operate feedback about the task itself, about the process used to produce it, about self-regulation, and about the person and note that task-level feedback paired with process-level guidance tends to be the most useful for improving the next iteration of the work, while purely evaluative or person-level comments tend to be the least useful. This distinction is used in the discussion of findings to interpret why some BTBB comments (for example, those citing a specific page or SLO) function as strong task-and-process feedback, while others (general impressions with no attached standard) sit closer to the less effective end of their model. The theory is used here as an interpretive lens for feedback

Online ISSN: 3006-8886
Print ISSN: 3006-8878

quality, not as a claim that BTBB reviewers are or should be functioning as classroom teachers; the parallel is with the structure of the feedback act itself, not the setting in which it occurs.

Research Methodology

Research Design

This study uses qualitative content analysis, following Krippendorff's (2018) definition of content analysis as a research technique for making replicable and valid inferences from text, supplemented with descriptive statistics to summarise how frequently different types of comments occur. This combination suits the research questions well: the underlying judgment of what a comment is doing (raising a curriculum issue, offering evidence, suggesting a fix) is qualitative, but counting how often each pattern occurs across a sample of reports requires a systematic, quantifiable coding step.

Data Source and Sample

The data consist of narrative comments contained in eleven Internal Review Committee (IRC) reports produced by BTBB, covering English, Mathematics, and Computer Science manuscripts for Grades 9 through 12. Only BTBB reports are used; reports produced by the Bureau of Curriculum are not part of the data set, consistent with the delimitations set out above. The reports were made available for this study by the researcher's institutional contacts at BTBB; several duplicate file submissions were identified and removed prior to analysis, leaving eleven unique documents.

Unit of Analysis

The unit of analysis is the individual reviewer comment a single, distinct remark addressing one issue rather than the review report as a whole. Across the eleven reports, 574 such comments were identified and coded. A single report typically contains many such comments touching on different pages or aspects of the manuscript, and coding at this finer level allows the study to capture variation within as well as across reports.

Report formats varied in one respect worth noting for methodological transparency: seven of the eleven reports (the English-medium reports) pair each identified issue with a recommendation in the same table row, while the remaining four (Mathematics and Computer Science) present a block of findings followed by a separate, shorter block of recommendations that is not matched item by item. Both formats were retained and coded using the same framework, with the difference addressed explicitly when interpreting the actionability results.

Coding Framework

Each comment was coded along four dimensions:

- Focus: the subject-matter area the comment addresses for example curriculum/SLO alignment, language and grammar, content accuracy, activities and questions, presentation and layout, glossary, assessment, or academic integrity.
- Objectivity: whether the comment is objective (points to a clearly verifiable issue), semi-objective (identifies a plausible concern without a precise measure), or subjective (a gen-

Online ISSN: 3006-8886
Print ISSN: 3006-8878

eral impression with no stated criterion), following the objective–subjective distinction set out by Sheldon (1988).
- Evidence base: whether the comment cites a specific page, unit, SLO code, or prior instance as the basis for the judgment, or offers none.
- Actionability: whether the comment gives the author a specific corrective action, a general suggestion, or no attached recommendation.

Coding was conducted using a detailed, rule-based coding manual developed specifically for this study. The same coding procedures and criteria were applied consistently across all eleven review reports to ensure a systematic and reliable analysis. Independent verification of a subsample by additional coders is discussed as a direction for follow-up work in the limitations section below.

Data Analysis

Coded comments were summarised using frequencies and percentages for each of the four dimensions. This was done both across the full sample and broken down by individual report, to allow comparison across subjects, grade levels, and reviewers.

Ethical Considerations

Institutional permission was obtained from BTBB before review reports were accessed. Reports were treated as internal institutional documents and are reported on in aggregate and thematic terms rather than by naming individual reviewers. The data were used solely for the purposes of this research.

Data Analysis/Results

Overview of the Sample

The analysis draws on eleven BTBB IRC reports covering English, Mathematics, and Computer Science manuscripts for Grades 9 through 12, yielding 574 individually coded comments.

Report	Subject / Manuscript	Comments coded
1	Grade 9 – English (Report 1)	49
2	Grade 9 – English (Report 2)	57
3	Grade 10 – English (Report 1)	51
4	Grade 10 – English (Report 2)	38
5	Grade 11 – English (Report 1)	75
6	Grade 11 – English (Report 2)	34
7	Grade 12 – English	65
8	Grade 9 – Computer Science	27
9	Grade 11 – Computer Science	55
10	Grade 9 – Mathematics	18
11	Grade 10 – Mathematics (chapter-wise)	105

As noted in the methodology, the seven English-medium reports pair each issue with its recommendation directly, while the Mathematics and Computer Science reports separate findings from

Online ISSN: 3006-8886
Print ISSN: 3006-8878

recommendations into two blocks. This difference is kept in view throughout the results below, particularly for RQ3.

RQ1 – Focus of Reviewer Comments

Focus area	Frequency	Percentage
Curriculum / SLO alignment	132	23.0%
Content accuracy / other subject-matter	129	22.5%
Presentation, layout & illustrations	97	16.9%
Activities & questions	90	15.7%
Language, grammar & mechanics	51	8.9%
Glossary / vocabulary list	33	5.7%
Technical / content accuracy	16	2.8%
Content selection (texts/poems/stories)	15	2.6%
Assessment	6	1.0%
Instructional design / pedagogy	3	0.5%
Academic integrity / plagiarism	2	0.3%

Curriculum and SLO alignment and general content accuracy are effectively tied as the leading concerns, together accounting for close to half of all coded comments. Reviewers frequently flag SLO numbers that do not match the content given, competencies missing from a unit, or SLOs stated in partial or non-standard wording rather than the exact phrasing of the national curriculum document; alongside this, they raise a wide range of subject-specific accuracy issues mathematical notation, technical definitions, factual errors particularly in the Mathematics and Computer Science reports. Presentation-related issues and activity/question design each account for roughly one comment in six. Assessment and pedagogy are rarely commented on directly, together accounting for under 2% of the sample, most likely because the manuscripts under review are largely finished textbooks rather than teacher-facing assessment or methodology documents.

Illustrative examples focus:

Curriculum/SLO alignment (Grade 9 – IRC Minutes): Issue – "SLO Number 10". Recommendation – "Also include the relevant portion of the SLO that is reflected on page 127."

Presentation/layout (Grade 9 – IRC Minutes): Issue – "Picture". Recommendation – "Replace the gender-biased image with a gender-neutral one."

Technical/content accuracy (Mathematics Grade 10): "Improve alignment of mathematical steps."

RQ2 – Objectivity and Evidence Base

Category	Frequency	Percentage
Objective	148	25.8%
Semi-objective	425	74.0%
Subjective	1	0.2%
Evidence / reference cited	128	22.3%
No evidence cited	446	77.7%

Online ISSN: 3006-8886
Print ISSN: 3006-8878

Purely subjective comments are close to non-existent (0.2%), meaning reviewers are, in a narrow sense, not simply offering bare opinion. The larger finding is that three in four comments are semi-objective: they identify a real, checkable issue without pointing to the specific page, unit, SLO code, or prior instance that would let an author or later reader verify the claim independently. Only 22% of comments cite that kind of concrete reference, and this figure is markedly lower in the technical subjects than in English, suggesting the evidence gap is not specific to language-focused review.

Illustrative examples objectivity and evidence:

Objective, evidence-cited (Grade 9 – English (Report 1)): Issue – "Proverbs". Recommendation – "Provide different proverbs, as these are already included on page 31." The page reference is exactly what lets an author verify and act on the comment without guessing.

Semi-objective, no evidence cited (Computer Science Grade 11): "Blockchain descriptions are misleading regarding 'public digital ledgers,' as not all blockchains are public." This names a real, checkable problem but does not point to the page or paragraph containing the claim.

Subjective (Computer Science Grade 9) – the only comment coded this way in the full sample: "Title page is dull and not attractive, not a good representation of the contents covered in the book." No standard or reference is offered for what would count as an acceptable title page.

RQ3 – Actionability of Recommendations

Category	Frequency	Percentage
Specific corrective action	369	64.3%
General suggestion	153	26.7%
No recommendation	52	9.1%

Read at face value, 64% of comments carry a specific, direct instruction. This figure needs the report-format caveat introduced earlier: in reports that separate findings from recommendations rather than pairing them, a meaningful share of individually coded findings show no attached recommendation at all most strikingly the Grade 11 Computer Science and Grade 9 Mathematics reports, where this is true of more than half of all findings. The Grade 9 Computer Science report shows a related pattern from a different angle: its findings are worded broadly enough that none were coded as narrowly specific, yet the report closes with a single, holistic instruction to "revise the entire manuscript" rather than fixes attached to each finding. In each case the format not necessarily the reviewer's underlying diligence is what limits how directly usable the feedback is at the level of an individual finding.

Illustrative examples actionability:

Specific corrective action (Grade 10 – English (Report 1)): Issue – "Activity 3 and 4". Recommendation – "Change font Color." Short, but an author can act on it immediately without further clarification.

Online ISSN: 3006-8886
Print ISSN: 3006-8878

General suggestion (Grade 12 – IRC Minutes), general comment: "Illustrations and Pictures should be clear and aligned with the content. Relevancy and clarity should be ensured." The direction is clear in spirit but does not say which illustrations, or what 'clarity' would look like in practice.

No attached recommendation (Computer Science Grade 11): Finding – "Definitions for 'Plagiarism' and 'Podcasts' are outdated." This finding sits in the report's Observations block with no item-level fix in the matching Recommendations block illustrating the format effect discussed above.

RQ4 – Consistency Across Reports

The table below reports two independent breakdowns side by side for each report: an objectivity/evidence pair (drawn from the RQ2 dimensions) and a full three-way actionability breakdown (Specific / General / No recommendation, drawn from the RQ3 dimension). The three actionability columns shown in green sum to 100% within each row; % Objective and % Evidence are a separate dimension and are not meant to sum with them.

Report	N	% Objective	% Evidence	% Specific	% General	% No rec.
Grade 9 – English (Report 1)	49	22.4%	34.7%	61.2%	38.8%	0.0%
Grade 9 – English (Report 2)	57	26.3%	35.1%	70.2%	29.8%	0.0%
Grade 10 – English (Report 1)	51	41.2%	43.1%	74.5%	25.5%	0.0%
Grade 10 – English (Report 2)	38	34.2%	34.2%	52.6%	47.4%	0.0%
Grade 11 – English (Report 1)	75	32.0%	18.7%	72.0%	28.0%	0.0%
Grade 11 – English (Report 2)	34	38.2%	29.4%	52.9%	47.1%	0.0%
Grade 12 – English	65	18.5%	3.1%	76.9%	23.1%	0.0%
Grade 9 – Computer Science	27	44.4%	48.1%	0.0%	100.0%	0.0%
Grade 11 – Computer Science	55	23.6%	3.6%	40.0%	3.6%	56.4%
Grade 9 – Mathematics	18	38.9%	11.1%	16.7%	27.8%	55.6%
Grade 10 – Mathematics	105	6.7%	12.4%	89.5%	0.0%	10.5%

The seven English-medium reports cluster in a comparatively narrow band on every dimension objectivity mostly between 18% and 41%, specificity mostly between 53% and 77% suggesting a reasonable, if imperfect, degree of shared practice among reviewers working in that format. The sharpest differences in the full sample are between subjects and formats rather than between individual reviewers. The Grade 10 Mathematics report combines the lowest objectivity score in the sample (6.7%) with the highest specificity score (89.5%). The Grade 9 Computer Science re-

Online ISSN: 3006-8886
Print ISSN: 3006-8878

port sits at the opposite pole on actionability in a different way: none of its findings were coded as carrying a narrowly specific instruction, but all of them were still coded as containing some form of general suggestion embedded in the finding's own wording, rather than being left with no recommendation at all the report's only "no recommendation" moment is that this guidance stays general throughout. The Grade 11 Computer Science and Grade 9 Mathematics reports, by contrast, are the two where the majority of findings (56% and 56%) genuinely carry no attached recommendation of any kind at the item level, consistent with their findings-then-recommendations format. This points to report template and subject area as at least as important a source of variation as individual reviewer practice.

Discussion

Read against the literature and the theoretical framework set out above, these findings sit closer to Sheldon's (1988) diagnosis than one might hope, and they map fairly precisely onto Hattie and Timperley's (2007) three feedback questions. The single subjective comment identified across the entire sample the Computer Science Grade 9 remark that a title page is "dull and not attractive, not a good representation of the contents" answers none of the three questions: it gives no stated standard, no evidence-based judgment against that standard, and no specific next step. It is, however, the exception that proves the rule; it is notable precisely because so little else in 574 comments reads this way. Far more representative is the Computer Science Grade 11 reviewer's observation that certain plagiarism and podcast definitions are "outdated": this answers "where am I going" and, loosely, "how am I going" by naming a real, defensible concern, but leaves "where to next" unanswered, since no specific citation or replacement source is offered. In Hattie and Timperley's terms, this is task-level feedback without the process-level guidance that would make it fully actionable an author receiving the page-referenced "these proverbs are already included on page 31" has clearly more to work from than one receiving an unreferenced comment about outdated definitions, even when both comments identify equally valid concerns.

The comparison with Ahmad et al. (2016), the closest prior Balochistan-specific study, is also instructive. Their work found that Balochistan's English textbooks had specific, identifiable shortcomings from the classroom teacher's point of view. The present findings suggest one plausible contributing factor sitting upstream of the classroom: if review comments at the manuscript stage are themselves often under-referenced, some shortcomings that a more evidence-based review process might have caught could persist into the published textbook. This is offered as a plausible connection rather than a demonstrated causal claim, since this study did not track individual manuscripts from review comment through to published outcome.

The consistency findings (RQ4) also refine the picture in a useful way. Because the sample spans three subjects and two distinct report formats, it becomes possible to see that BTBB's review quality is not simply a matter of some reviewers being more careful than others. Format itself shapes what a comment can look like: a review process that lists findings and recommendations separately structurally produces less item-level actionability than one that pairs them, regardless of how thorough the underlying reviewing was. The clearest case in the sample is the Grade 11 Computer Science report, where 56% of findings including the outdated "Plagiarism" and "Podcasts" definitions cited earlier sit in an Observations block with no matching entry in the Recommendations block that follows. The Grade 9 Computer Science report shows a related but distinct pattern: every one of its 27 findings is written broadly enough to double as its own general

Online ISSN: 3006-8886
Print ISSN: 3006-8878

suggestion, but the report's only sharply worded instruction is the single closing line "the authors should reconsider and revise the entire manuscript thoroughly considering the points raised above" rather than a specific fix attached to each finding. Compare either of these with the Grade 10 English (Report 1) report, where an issue as small as "Activity 3 and 4" is paired directly with "Change font Color." All three reviewers may have done comparably careful work, but only the row-paired format leaves the author with a page-by-page action list. This is a finding with a fairly direct practical implication, discussed in the recommendations below.

Recommendations

Capacity Building: Training for Reviewers and, to a Lesser Extent, Authors

The single strongest recommendation to come out of this study is that BTBB invest in structured, recurring training with the heavier weight placed on reviewers rather than authors, since it is reviewer practice that this study's data speak to most directly.

- Reviewer training should be treated as the priority. The data show that BTBB reviewers are rarely subjective, but are inconsistent in citing evidence and in writing actionable recommendations; this is a skill gap that structured training can close far more directly than any change to the review form alone. Training should cover how to write a comment that names a specific, checkable reference (a page, unit, or SLO code), and how to phrase a recommendation as a specific instruction rather than a general direction, using real examples from strong- and weak-performing comments such as those illustrated in Section 10.
- Reviewer training should be recurring rather than a one-time induction, ideally scheduled before each major review cycle, and should include a short calibration exercise in which reviewers code the same sample comments and compare results, both to reinforce the standard and to surface the kind of inter-reviewer inconsistency documented under RQ4.
- A shorter, complementary training or orientation session for authors would help them read and act on review feedback more effectively for instance, understanding what an SLO-referenced comment is asking for, and how to respond when a comment is general rather than specific but this should supplement, not substitute for, the reviewer-focused training above.

For BTBB

- Standardise the review report template across subjects so that every identified issue is paired with its own recommendation in the same entry, rather than allowing findings and recommendations to be listed as two separate, unmatched blocks.
- Introduce a simple citation convention for example, requiring a page, unit, or SLO reference alongside any comment to close the evidence gap identified in RQ2, particularly in the technical subjects, where specific references (an equation number, a line of code) are often readily available. This convention should be taught and reinforced through the reviewer training recommended above.

Online ISSN: 3006-8886
Print ISSN: 3006-8878

For Textbook Authors

- Treat curriculum/SLO alignment and general content accuracy as the two issue types most likely to be raised in review, and prioritise self-checking these areas before submission.
- Where a review comment lacks a specific reference, consider proactively requesting clarification from BTBB rather than guessing at the intended fix, given how common semi-objective, under-referenced comments are in the current sample.

For Future Research

- Have members of the research team independently apply the same coding manual to a subsample, so that a formal inter-coder reliability figure can be reported alongside the frequencies presented here.
- Track a smaller set of manuscripts from initial review comments through to the published textbook, to test more directly whether under-referenced review comments correspond to issues that persist into the final product.
- Extend the sample to additional subjects, grade levels, and, where access allows, to Bureau of Curriculum reports, to test whether the subject- and format-driven patterns found here hold more broadly.
- Evaluate the effect of the training recommended above directly, for example by comparing objectivity, evidence, and actionability rates in review reports produced before and after a training intervention.

Limitations of the Study

Several limitations should be kept in mind when interpreting these findings. A few coding-related aspects stand out as areas that later stages of this research programme could usefully strengthen. The coding manual applied here involves some inherently interpretive judgment calls particularly around the line between “semi-objective” and “objective”, and between a genuinely specific instruction and a general one and a natural next step would be for the frequencies reported here to be cross-verified through independent coding of a subsample, with a formal agreement statistic such as percentage agreement or Krippendorff’s alpha reported alongside them. This is a readily addressable next step given the composition of the research team, and is recommended as a direction for follow-up work. The sample, while larger and more varied than initially planned, remains a purposive sample of eleven reports rather than a full census of BTBB’s review output, and the four Mathematics and Computer Science reports use a findings-then-recommendations format that mechanically affects the actionability figures in ways discussed throughout the results. Finally, this study examines the text of review reports only; it does not verify whether reviewers’ underlying judgments about the manuscripts were themselves correct, nor does it track whether authors ultimately acted on the comments received.

Conclusion

This study set out to examine what BTBB reviewers actually write in their narrative review comments, rather than the review tools or checklists that structure the process on paper. Across 574 comments drawn from eleven reports spanning English, Mathematics, and Computer Science manuscripts, the picture that emerges is of a review process that is rarely baselessly opin-

Online ISSN: 3006-8886
Print ISSN: 3006-8878

ionated but is also, more often than not, under-referenced: reviewers tend to identify real problems without consistently showing the specific evidence behind the claim, and how actionable their recommendations are depends considerably on the format a given report happens to use. These are addressable patterns rather than fundamental flaws, and the recommendations above are offered in that spirit as concrete, incremental steps that could make BTBB's already-substantive review process easier for authors to act on.

AI Disclosure Statement

During the preparation of this manuscript, the authors used Claude (Anthropic), an AI language model, to assist with several stages of the research and writing process. This included extracting and organising data from the review reports analysed in this study, applying the study's coding framework to classify reviewer comments, generating the associated descriptive statistics, and drafting portions of the manuscript text, including sections of the literature review, methodology, results, discussion, and recommendations.

All AI-assisted output was critically reviewed, verified, and edited by the authors, who take full responsibility for the accuracy, integrity, and conclusions of this work. The AI tool was not involved in the design of the study, the interpretation of findings in any final sense, or the decision to submit the manuscript for publication, and it is not listed as an author, as it cannot take responsibility for the work in the manner required of an author. The underlying data used in this study are original institutional documents and were not generated or fabricated by AI

Acknowledgement

The research was conducted with financial support from Balochistan Textbook Board (BTBB) Research Centre.

References

1. Aftab, A. (2012). English language textbooks evaluation in Pakistan (Unpublished doctoral thesis). University of Birmingham.
2. Ahmad, I., Rehman, S., Ali, S., & Khan, T. (2016). Pakistani government primary school teachers and the English textbooks of Grades 1–5: A mixed methods teachers'-led evaluation. Cogent Education, 3(1).
3. Cunningsworth, A. (1995). Choosing your coursebook. Heinemann.
4. Ellis, R. (1997). The empirical evaluation of language teaching materials. ELT Journal, 51(1), 36–42.
5. Gholampour, S., & Mehrabi, D. (2023). Literature review of ELT textbook evaluation. MEXTESOL Journal, 47(1).
6. Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112.
7. Krippendorff, K. (2018). Content analysis: An introduction to its methodology (4th ed.). SAGE Publications.
8. McDonough, J., Shaw, C., & Masuhara, H. (2013). Materials and methods in ELT (3rd ed.). Wiley-Blackwell.
9. Ministry of Federal Education and Professional Training. (2017). National curriculum framework for Pakistan. Government of Pakistan.

Online ISSN: 3006-8886
Print ISSN: 3006-8878

10. Ministry of Federal Education and Professional Training. (2020). National curriculum for English language: Grades I–XII. Government of Pakistan.
11. National Curriculum Council. (2022). National curriculum of Pakistan 2022–23. Ministry of Federal Education and Professional Training, Government of Pakistan.
12. Sheldon, L. E. (1988). Evaluating ELT textbooks and materials. ELT Journal, 42(4), 237–246.
13. Tomlinson, B. (Ed.). (2013). Developing materials for language teaching (2nd ed.). Bloomsbury.