

RESEARCH ARTICLE

Google Classroom Usability in Google Play Reviews: A Framework-Guided Content Analysis Using Nielsen's Dimensions

Rafiq Rizalita¹ * , Putut Aji Nalendro¹ , Novi Febrianti² 

¹Department of Mathematics and Natural Science Education, Faculty of Teacher Training and Education, University of Lampung, Bandar Lampung, **Indonesia**

²Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, **Malaysia**

✉ *Corresponding Author: rafiqarizalita@fkip.unila.ac.id

This article contributes to:



ABSTRACT

Google Classroom is used to organize learning activities, yet structured assessments may overlook problems reported during everyday use. This study examined 800 Google Play reviews to identify usability-related concerns, classify them using Nielsen's five usability dimensions, and measure the proportion of feedback that did not fit the framework. Reviews were retrieved with the Python google-play-scraper package using Indonesian language and country settings and sorted from newest to oldest. Framework-guided qualitative content analysis assigned each review one primary category: learnability, efficiency, memorability, errors, satisfaction, or Other. Category frequencies were summarized descriptively. Ratings were concentrated at both extremes, with 347 one-star reviews (43.38%) and 288 five-star reviews (36.00%). Of the 800 reviews, 427 (53.38%) were classified within Nielsen's dimensions, while 373 (46.63%) fell into the Other category. Among framework-mapped reviews, errors accounted for 56.44% and satisfaction for 37.00%, whereas learnability, efficiency, and memorability collectively represented 6.56%. Common error-related reports concerned bugs, failed assignment uploads, login difficulties, crashes, and notification problems. The large Other category indicates that nearly half of the reviews could not be classified within the five dimensions under the study's coding rules. These findings suggest that app-store reviews complement structured usability assessments by highlighting user-reported problems with essential learning tasks. Because the reviews were self-selected and collected at one point in time, the category frequencies should not be generalized to Google Classroom users overall.

KEYWORDS

Content analysis; Digital learning; Google Classroom; LMS; Usability; User reviews

🕒 *Received:* Aug. 16, 2026; *Revised:* Sep. 25, 2026; *Accepted:* Oct. 26, 2026; *Published Online:* Oct. 31, 2026

How to Cite: Rizalita, R., Nalendro, P. A., & Febrianti, N. (2026). Google Classroom Usability in Google Play Reviews: A Framework-Guided Content Analysis Using Nielsen's Dimensions. *Journal of Hypermedia & Technology-Enhanced Learning*, 4(3), 288–303. <https://doi.org/10.58536/j-hytel.267>



1. INTRODUCTION

Digital learning platforms support routine educational activities that depend on timely access to materials, assignments, feedback, and communication. Use of these platforms expanded during the COVID-19 pandemic, increasing institutional reliance on online services [1], [2]. This makes usability an operational concern: access to a platform does not guarantee that students and instructors can complete the tasks it supports [3], [4]. An unsuccessful assignment upload or inability to sign in can interrupt a time-sensitive learning activity even when the platform is otherwise perceived as useful. Assessing digital learning systems therefore requires attention to operational usability as well as pedagogical value, while distinguishing usability from educational effectiveness.

Google Classroom combines course organization, assignments, and communication with other Google services. Research has examined its perceived usefulness, ease of use, acceptance, and pedagogical integration [5], [6], [7], [8], [9]. Structured usability research has also assessed how users interact with the platform [10]. Pawestri and Khasanah examined 77,454 Google Classroom reviews using sentiment analysis and association-rule mining to identify recurring complaint terms, including those related to assignments and uploads [11]. That study examined sentiment and recurring terms rather than classifying reviews by usability dimension. Related work has synthesized findings on Google Meet interface complaints [12], and evaluated interface heuristics and deadline-reminder needs in a university virtual classroom [13]. These studies provide context but do not examine the classification of unsolicited Google Classroom reviews using Nielsen's dimensions.

Usability provides a basis for examining these accounts. ISO 9241-11 defines usability through effectiveness, efficiency, and satisfaction in a specified context of use [14]. Nielsen's framework differentiates learnability, efficiency, memorability, errors, and satisfaction [15]. Mobile-app usability research also shows that device constraints, tasks, and usage settings shape the interactions users experience [16], [17]. A failed upload, difficult menu navigation, and a favorable overall appraisal reflect different aspects of Google Classroom use. Nielsen's dimensions provide explicit categories for organizing these comments, but assigning a review to a dimension is not equivalent to measuring the corresponding attribute through a standardized usability assessment.

App-store reviews capture experiences that users report without researcher prompts. Prior research has used them to identify technical problems, feature requests, and experiences relevant to software development [18], [19], [20]. In learning platforms, such reviews may reveal disruptions to educational tasks not anticipated in surveys or usability tests. However, reviews are voluntarily posted, differ substantially in specificity, and may be influenced by software updates and the circumstances of posting [21], [22], [23]. Reviews of educational apps can also address usefulness, content, compatibility, and support rather than usability alone [24]. Consequently, the frequency of each primary complaint category describes only the analyzed corpus; it does not estimate prevalence among all users or establish a technical cause.

Although studies of Google Classroom acceptance [5], [8], [9] and structured usability [10] provide valuable evidence, the remaining research problem is not whether app-store reviews contain operational complaints, which previous mining studies have already shown [11]. Rather, it concerns which aspects of these unsolicited accounts a theory-guided classification can represent. Sentiment classification and keyword extraction summarize evaluative patterns and recurring terms, whereas a Nielsen-based scheme

distinguishes specific usability attributes and makes its classification boundaries explicit. A review may describe learning support, product requests, or circumstances unrelated to the five attributes; forcing it into one of those categories would overstate the framework's coverage. The proportion assigned to Other therefore indicates the coding scheme's coverage of this corpus. That proportion reflects the corpus and coding rules; it neither tests Nielsen's framework nor identifies the content of uncategorized reviews.

This study applies Nielsen's five dimensions to a fixed corpus of 800 Google Play reviews of Google Classroom. Three questions guide the analysis: (RQ1) How are star ratings distributed in the collected reviews? (RQ2) Which primary usability categories and associated themes are most frequent? (RQ3) What proportion of reviews can be assigned defensibly to Nielsen's dimensions, and what proportion remains in Other? Reporting both the thematic distributions and unclassified reviews complements existing research on acceptance, usability testing, and app-review mining. The conclusions concern only the analyzed reviews and do not represent platform-wide failure rates or measured learning outcomes.

2. LITERATURE REVIEW

2.1. Digital Learning and Learning Management Systems

Digital learning environments combine instructional resources, assessment, and communication, but their educational value also depends on platform quality, institutional support, and pedagogical design. Research on engagement in online learning indicates that access to technology alone does not ensure productive participation [4]. Learning management systems (LMSs) bring course content, assignments, assessment, and communication into a common environment [3]. Studies of e-learning system success distinguish system quality from perceived usefulness and educational outcomes [25], [26]. Usability problems may disrupt access to materials or assignment completion, but these disruptions do not, by themselves, establish effects on learning outcomes.

Google Classroom has been examined through acceptance models, perceptions of usefulness and ease of use, and studies of pedagogical integration [5], [7], [8], [9]. A structured usability evaluation has assessed interaction quality against explicit usability criteria [10]. More recently, a study of 77,454 Google Play reviews used sentiment analysis and association-rule mining to identify recurring terms associated with user complaints [11]. These approaches provide different evidence: surveys measure responses to predefined constructs, usability tests examine selected tasks, and review mining identifies patterns in voluntarily posted text. The present study instead examines how Nielsen's categories classify a smaller corpus of Google Play reviews ordered by recency.

2.2. Usability

Usability concerns users' effectiveness, efficiency, and satisfaction when pursuing specified goals in a particular context. This definition follows ISO 9241-11 [14]; Nielsen's framework additionally distinguishes learnability, efficiency, memorability, errors, and satisfaction [15]. Learnability concerns initial understanding of an interface, efficiency concerns the effort required once it is understood, and memorability concerns how readily users can resume using an interface after a period of non-use. Nielsen's errors attribute addresses errors and recovery, while satisfaction reflects subjective appraisal. In this study, reports of crashes, failed uploads, and authentication problems are treated as user-reported operational failures within the errors category; they are not independently verified system faults or direct measures of error rates.

For learning platforms, interface usability is distinct from instructional effectiveness and learning gains. E-learning evaluation research has long distinguished functional interaction from broader educational experience [27]. Mobile-app usability research likewise identifies the influence of device characteristics, usage context, and interaction requirements [16]. A failed assignment upload, praise for intuitive navigation, and a claim that the platform supports study activities reflect different experiences. The analysis therefore treats perceived usefulness as a favorable appraisal only when warranted by the review wording; it does not equate that appraisal with evidence of improved learning.

2.3. User Reviews as User-Generated Content

App-store reviews combine user-written text with metadata such as star ratings and publication dates. Unlike responses to researcher-designed tasks, these comments reflect what users choose to post publicly [19]. Research has mined such feedback for bug reports, requirements, feature requests, and sentiment [18], [20]. However, reviews are often brief, ambiguous, specific to an application, and dependent on when they were posted [22]. These comments can identify concerns worth investigating, but neither theme counts nor rating distributions capture the experiences of users who did not post reviews.

Reviews of educational applications often combine judgments about technical performance with perceptions of learning usefulness. Studies of mobile-learning app feedback describe concerns involving content, compatibility, support, and functionality [24], while educational-app research has also classified review content using software-quality frameworks [28]. For Google Classroom specifically, Pawestri and Khasanah used sentiment and keyword-based analysis of Google Play reviews to identify recurring problem terms [11]. Their approach differs from the present study's assignment of one primary Nielsen dimension per review. Methodological research nonetheless cautions that either approach depends on collection settings, coding choices, and the content of the retrieved reviews [18], [21], [22]. Retaining an Other category prevents unsupported assignments to specific usability dimensions.

2.4. Analytical Positioning

The analysis combines deductive framework categories with inductively identified themes. Nielsen's five dimensions define the primary classification, while open codes describe the specific concern or appraisal expressed in each review. This separation follows content-analysis guidance on coding units, categories, and decision rules [29], [30]. It allows descriptive comparisons across the corpus, but the resulting frequencies should not be interpreted as standardized usability scores. The Other category quantifies reviews that could not be assigned to predefined categories; it does not describe their content.

3. MATERIALS AND METHODS

3.1. Study Design and Analytical Framework

The study used framework-guided qualitative content analysis with descriptive frequency summaries. Each Google Play review was the unit of analysis, and the principal concern conveyed in its text informed its primary classification. Nielsen's five attributes—learnability, efficiency, memorability, errors, and satisfaction—provided deductive categories [15]; open codes retained the specific issues or appraisals expressed by users. A review was assigned to Other when its principal content did not support a defensible classification within the five dimensions [29], [30]. The study was descriptive and did not estimate population prevalence, infer causal effects, or administer a standardized usability assessment.

3.2. Data Source and Corpus Construction

Google Classroom reviews were collected from Google Play using the Python google-play-scraper package (application identifier: com.google.android.apps.classroom). The retrieval script shown in Figure 1 specified lang='id', country='id', sort=Sort.NEWEST, and count=800. These settings identify the requested language, storefront, and sorting order; they do not verify each reviewer's nationality, educational role, or place of use. The analysis used review text, star ratings, and review publication dates. Because the review stream and application versions change over time, the 800 collected reviews constitute a time-dependent, nonprobability corpus rather than a representative sample of Google Classroom users [22].

```
!pip install google-play-scraper pandas
from google_play_scraper import reviews, Sort
import pandas as pd

result, _ = reviews(
    'com.google.android.apps.classroom',
    lang='id',
    country='id',
    sort=Sort.NEWEST,
    count=800
)

df = pd.DataFrame(result)

print(df[['score', 'content', 'at']].head())

df[['score', 'content', 'at']].to_csv(
    'google_classroom_reviews.csv',
    index=False
)

print("File berhasil disimpan")
from google.colab import files
files.download('google_classroom_reviews.csv')
```

Figure 1. Python retrieval script and parameters used to obtain Google Play reviews

The retained analytic corpus comprised 800 reviews, matching the count specified in Figure 1. No review-level screening log was retained, so the extent of any duplicate removal or exclusion of incomplete or irrelevant reviews cannot be established. Neither the extraction timestamp nor the installed scraper version was preserved. Accordingly, the 800 available reviews serve as the denominator for all corpus-level results, and missing collection metadata limit reproducibility and cannot be reliably reconstructed [18], [22].

3.3. Coding Framework and Decision Rules

Coding began with open codes and proceeded to framework classification. Each review first received a code reflecting its main reported experience; related codes were grouped into themes such as Ease of Use, Fast and Practical, Consistent Interface, Bugs/Glitches, Failed Assignment Uploads, Login Errors, Crashes/Application Exits, Notification Issues, Positive Satisfaction, and Supports Learning. Each review was subsequently assigned one primary category—learnability, efficiency, memorability, errors, satisfaction, or Other—to produce mutually exclusive frequency counts. Classification was based on the dominant content: learning to use the interface (learnability), effort or speed of task completion (efficiency), recalling how to use the interface after previous use (memorability), reported failures or

recovery difficulties (errors), or an overall user appraisal (satisfaction). A favorable comment about support for learning was classified under Satisfaction only when it expressed a positive experience, not an observed learning gain. Reviews without a sufficiently clear fit remained in Other.

A single primary category makes the counts mutually exclusive but may obscure secondary concerns when a comment describes more than one issue. The reported counts therefore represent dominant classifications rather than the total number of mentions of each topic. Likewise, infrequent coding of learnability or memorability reflects what reviewers chose to report, not evidence that these attributes are satisfactory or unimportant.

3.4. Analytical Checks, Trustworthiness, and Reporting

Descriptive counts were compiled for star ratings, primary categories, and themes. The five rating groups and six primary categories each sum to 800 reviews; the error subthemes sum to 241 reviews and the satisfaction subthemes to 158. Corpus percentages use 800 as the denominator, whereas percentages restricted to reviews classified within Nielsen’s dimensions use 427. Values are rounded independently to two decimal places, which may produce minor differences in displayed totals. The retained materials do not document independent second coding or intercoder agreement, so no reliability coefficient is reported. Because review-level links between final codes and star ratings were not preserved, category-by-rating analysis was not possible. Indonesian-language comments are summarized in English without reproducing reviewer identifiers. These reporting choices support cautious interpretation but cannot substitute for an auditable coding record.

4. RESULTS

The results address three questions: the distribution of star ratings, the primary Nielsen classifications and their themes, and the proportion of reviews assigned to Other. Because review-level links between categories and ratings were not preserved, the two distributions are presented independently rather than used to infer associations.

Table 1. Star-Rating Distribution of the Google Classroom Review Corpus

Rating	Number of Reviews	Percentage
1	347	43.38%
2	56	7.00%
3	58	7.25%
4	51	6.38%
5	288	36.00%

Table 1 shows that 347 reviews (43.38%) received one star and 288 (36.00%) received five stars. Together, these ratings accounted for 79.38% of the dataset; the other 165 reviews (20.63%) received two, three, or four stars. This concentration at the endpoints resembles patterns reported in research on voluntary online reviews [23]. However, the distribution cannot establish whether Google Classroom users generally have polarized views. Because review submission is voluntary, the observed pattern may reflect posting behavior as well as usage experiences [19], [22], [23].

4.1. Distribution Across Nielsen's Usability Dimensions

Table 2 presents the mutually exclusive primary classifications. Of 800 reviews, 427 (53.38%) could be assigned to one of Nielsen's dimensions, whereas 373 (46.63%) were classified as Other. These proportions describe the coverage of the stated classification criteria within the corpus. The Other classification does not mean that the remaining reviews were unrelated to usability, nor does it explain why they could not be classified.

Table 2. Distribution of Reviews Across Nielsen's Usability Dimensions

Dimension	Frequency	Percentage
Errors	241	30.13%
Satisfaction	158	19.75%
Learnability	12	1.50%
Efficiency	10	1.25%
Memorability	6	0.75%
Other	373	46.63%
Total	800	100%

Across the full corpus, 241 reviews (30.13%) were classified as errors and 158 (19.75%) as satisfaction; learnability (12; 1.50%), efficiency (10; 1.25%), and memorability (6; 0.75%) were much less frequent. Among the 427 reviews mapped to Nielsen's dimensions, errors comprised 56.44% and satisfaction 37.00%, whereas the other three categories together comprised 6.56%. These proportions describe the dominant issues in the collected comments, not the relative quality or importance of the underlying usability attributes.

4.2. Analysis by Usability Dimension

4.2.1 Learnability

Twelve reviews (1.50% of the full corpus) were classified under learnability, all within the Ease of Use theme (Table 3). These comments described the interface or its menus as easy to understand. These reports indicate that the reviewers found the interface easy to understand; they do not measure how quickly new users learn the system or establish its learnability across the wider user population.

Table 3. Learnability Theme

Theme	Frequency
Ease of Use	12

4.2.2 Efficiency

Ten reviews (1.25% of the corpus) were classified under efficiency, all within Fast and Practical (Table 4). Reviewers described Google Classroom as convenient for routine tasks, including accessing course materials and submitting assignments. These comments indicate perceived convenience for the tasks described; they do not provide task-time measurements or establish overall system efficiency.

Table 4. Efficiency Theme

Theme	Frequency
Fast and Practical	10

4.2.3 Memorability

Six reviews (0.75% of the corpus) were classified under memorability, all in the Consistent Interface theme (Table 5). One review described a familiar interface that was easy to remember, consistent with Nielsen’s memorability attribute [15]. However, the available reviews do not document controlled return-to-use tasks, and six comments cannot characterize memorability among Google Classroom users generally.

Table 5. Memorability Theme

Theme	Frequency
Consistent Interface	6

4.2.4 Errors

Errors were the largest Nielsen category, comprising 241 reviews (30.13% of the full corpus; 56.44% of mapped reviews). As Table 6 shows, Bugs/Glitches accounted for 96 reports (39.83% of error-classified reviews), followed by Failed Assignment Uploads (58; 24.07%), Login Errors (42; 17.43%), Crashes/Application Exits (31; 12.86%), and Notification Issues (14; 5.81%). Failed uploads and login difficulties together accounted for 41.49% of error-classified reviews. These are reports of disrupted interactions, not independently verified software defects or estimates of failure rates across installations.

Table 6. Error-Related Themes

Theme	Frequency
Bugs/Glitches	96
Failed Assignment Uploads	58
Login Errors	42
Crashes/Application Exits	31
Notification Issues	14

Reported examples included unsuccessful assignment submission, difficulty signing in after an update, and unexpected application closure. These incidents concern activities that students and instructors must complete to participate in online learning. Under ISO 9241-11, obstacles to goal completion can affect effectiveness and satisfaction [14]. The reports therefore identify workflows that warrant technical investigation, although the reviews alone cannot establish why the failures occurred.

4.2.5 Satisfaction

Satisfaction comprised 158 reviews (19.75% of the corpus; 37.00% of mapped reviews). Table 7 separates Positive Satisfaction (102 reviews; 64.56% of this category) from Supports Learning (56;

35.44%). Comments in the latter theme described Google Classroom as helpful or as making learning activities easier; they were included because they expressed favorable appraisals. They do not provide evidence of measurable learning gains. Neither theme constitutes a standardized satisfaction score.

Table 7. Satisfaction-Related Themes

Theme	Frequency
Positive Satisfaction	102
Supports Learning	56

5. DISCUSSION

5.1. Reported Operational Problems as the Main Usability Signal

Within the 427 reviews assigned to Nielsen’s framework, operational problems formed the most frequent primary category. The coded error themes concerned unsuccessful uploads, login difficulties, application crashes, notifications, and other reported faults. This pattern is consistent with app-review research showing that public feedback often includes reports of software problems and requirements relevant to maintenance [18], [19], [20]. A separate Google Classroom review-mining study also identified assignment- and upload-related terms in negative reviews [11]. Because app-review results depend on collection conditions and analytic procedures [22], that study provides contextual evidence rather than a direct replication of the present frequencies. User reports do not establish technical causes or failure rates across installations. For learning platforms, these reports identify workflows whose disruption could affect time-sensitive educational activities, consistent with research that treats system quality as distinct from learning outcomes [25], [26].

The dominance of operational complaints should not be read as evidence that Google Classroom is educationally ineffective. A platform may be perceived as useful for learning while presenting serious difficulties during particular interactions. Reviews of mobile-learning applications similarly encompass usability, compatibility, support, and perceived value [24]. The findings identify recurrent problems described by reviewers; determining their technical causes or effects on learning would require additional evidence, such as diagnostic logs, controlled usability tests, or outcome measures.

5.2. Coexistence of Technical Problems and Positive Evaluations

Positive feedback was also common: 36.00% of the corpus comprised five-star reviews, and 37.00% of framework-mapped reviews were coded under satisfaction. These aggregate measures describe different aspects of the corpus and should not be treated as contradictory. Previous Google Classroom research has emphasized perceived usefulness, ease of use, and factors shaping acceptance [5], [8], [9]. In this dataset, satisfaction-related comments also described perceived usefulness for study activities. However, the absence of a preserved review-level link between ratings and codes prevents assessment of whether error reports corresponded to low ratings or favorable appraisals to high ratings.

Star ratings provide an overall evaluation but reveal little about what users valued or found difficult. Ratings may influence educational-app selection [31], yet they are not direct measures of usability, acceptance, or learning outcomes. Previous app-review research also distinguishes textual sentiment and technical concerns from numeric ratings [22], [24]. Here, ratings describe the distribution of expressed

evaluations, whereas the coding identifies the topics that appeared in review text. No association between the two was tested.

5.3. Why Learnability, Efficiency, and Memorability Were Rarely Reported

Learnability, efficiency, and memorability accounted for only 28 reviews (3.50% of the corpus; 6.56% of mapped reviews). Their low frequency indicates limited spontaneous mention, not as evidence that these qualities were unimportant or performed well. Assessing learnability or memorability often requires observation of first use or return after non-use, while efficiency can be assessed through task time and effort [15]. Structured usability methods can examine these attributes more directly [16], [17]. By contrast, failed logins or assignment submissions may readily prompt public complaints.

The pattern of topics therefore reflects, in part, what reviewers choose to report. Few mentions of memorability cannot establish strong memorability, just as repeated upload complaints cannot be used to calculate an overall software failure rate. Review mining and structured Google Classroom usability studies [10] address different kinds of evidence. Combining them could clarify both interaction attributes seldom articulated in public reviews and problems reported under everyday conditions.

5.4. Framework Coverage and the Meaning of the Other Category

Other was the largest primary classification: 373 reviews (46.63%) could not be assigned confidently to the five Nielsen dimensions. This result indicates limited coverage of the selected coding scheme within the observed feedback, not an intrinsic failure of Nielsen's framework. Public reviews may discuss software and educational concerns that do not correspond neatly to a predefined usability taxonomy [18], [20], [22], while framework-based app studies also identify multiple quality domains [28]. Retaining Other avoids assigning a specific usability dimension when the review text does not justify one.

No subcategories were preserved for the 373 reviews classified as Other, so the analysis cannot determine which concerns predominate within that group. The size of Other indicates limited classification coverage but does not reveal its internal themes. A future study could retain the raw time-stamped corpus and code these reviews inductively alongside the deductive framework [18], [21]. Such an approach would allow topic-level interpretation of material that falls outside the initial categories without inventing classifications retrospectively.

5.5. Practical Implications for Digital Learning Platforms

For developers and institutions, the findings identify essential learning workflows associated with user complaints: authentication, assignment uploads, application stability, and notifications. Institutions may need alternative submission or communication arrangements when the platform becomes unavailable. The review themes suggest five priorities for investigation: 1) examining application crashes and recovery; 2) investigating login and authentication failures; 3) testing assignment-upload and resubmission procedures; 4) evaluating the timeliness and reliability of notifications; and 5) comparing reported difficulties across app versions, devices, account states, and network conditions.

The first four priorities follow the types of failures described in the reviews; the fifth requires evidence not present in the dataset. No application logs, device diagnostics, network measurements, or reliable review-to-version links were retained. These are investigative priorities, not verified diagnoses of software defects. Combining review analysis with operational analytics and controlled usability tests would help identify the causes and consequences of recurrent problems [19], [20], [21].

5.6. Empirical and Methodological Contribution

This study classifies Google Classroom reviews using a usability framework and quantifies the coding scheme's coverage. Previous work has examined acceptance and structured usability [5], [8], [9], [10], while a recent review-mining study analyzed sentiment and recurring problem terms [11]. The present analysis instead reports which concerns were assigned to Nielsen's five dimensions and the proportion that could not be classified within the framework. Of 800 reviews, 427 (53.38%) were classified within those dimensions and 373 (46.63%) as Other; errors dominated the mapped reviews. This makes the coding scheme's coverage explicit but does not validate the framework, characterize the uncategorized reviews, or establish the prevalence of operational problems among Google Classroom users. The contribution is a descriptive account of reported usability concerns and the coverage of a predefined coding scheme within this corpus.

5.7. Limitations and Future Research

Six limitations constrain these findings. First, the 800 reviews were voluntarily submitted and retrieved using Indonesian language and country settings; the corpus is not a probability sample, and reviewer nationality and educational roles cannot be determined from the retrieval settings. Second, the extraction timestamp, scraper version, and record-level screening decisions are unavailable, restricting reproducibility. Third, the available materials do not document independent double coding or interrater agreement, limiting assessment of classification reliability. Fourth, assigning only one category per review can omit secondary concerns. Fifth, the lack of a review-level matrix linking codes and ratings prevents analysis of their association. Sixth, no preserved subcategories describe the 373 reviews assigned to Other. Future studies should archive collection metadata and coding decisions, assess coding agreement, retain review-level links, permit multiple codes where justified, and examine unmapped content inductively across collection periods [18], [22], [30].

6. CONCLUSION

This study analyzed 800 Google Play reviews of Google Classroom using Nielsen's usability dimensions as a coding framework. One-star and five-star ratings represented 43.38% and 36.00% of the corpus, respectively. Of the 427 reviews mapped to the framework, 56.44% concerned errors and 37.00% satisfaction, while learnability, efficiency, and memorability were rarely the principal topics. Within the errors category, reviewers reported bugs, unsuccessful uploads, login problems, crashes, and notification issues. These accounts identify recurring difficulties with routine learning workflows, but do not establish how frequently those failures occur across the user population. Positive appraisals also appeared, although their relationship with individual ratings and error reports could not be assessed.

The analysis also shows incomplete framework coverage: 373 reviews (46.63%) could not be classified defensibly within Nielsen's five categories. This result documents the limits of the chosen coding scheme for the collected feedback, rather than demonstrating that the framework is invalid. App-store analysis can complement structured usability assessments by indicating which concerns users choose to report, but self-selection and time-dependent collection restrict generalization. Future research should preserve review-level data, document independent coding procedures, and combine deductive classification with inductive analysis of uncategorized comments and technical diagnostic evidence.

DECLARATIONS

Author Contributions

Rafiq Rizalita: Conceptualization; Methodology; Investigation; Data Curation; Writing – Original Draft. **Putut Aji Nalendro:** Formal Analysis; Validation; Visualization; Writing – Review & Editing. **Novi Febrianti:** Writing – Review & Editing. All authors approved the manuscript and agreed to the author order.

Acknowledgments

Not applicable.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing Interests

The authors declare no competing interests.

Ethical Approval

Formal ethical approval was not obtained. The study analyzed publicly available Google Play reviews without participant recruitment, direct interaction, or intervention. Reviewer usernames and other directly identifying information are not reported.

Informed Consent

Not applicable. The study used publicly available Google Play reviews without recruiting or interacting with participants or conducting an intervention. No private or confidential personal information was intentionally collected or used.

Consent for Publication

Not applicable. The study reports no identifiable individual participants or private personal information.

Study or Protocol Registration

Not applicable. The study did not involve a clinical trial or intervention requiring prospective registration.

Data Availability

The analysis used publicly available Google Classroom reviews retrieved from Google Play. Individual reviews may no longer be accessible in their original form because app-store content changes over time. The processed dataset and supporting analytical materials are available from the corresponding author upon reasonable request, subject to applicable terms of service and data-use requirements.

Generative AI and AI-Assisted Technologies Statement

During manuscript preparation, ChatGPT (OpenAI, GPT-5.6, accessed in 2026) and Claude (Anthropic, accessed in 2026) were used as language and research assistants, primarily for translation and text refinement. Grammarly (accessed in 2026) was used for grammar, spelling, clarity, and English-language proofreading. All AI-assisted material was reviewed and verified by the authors, who remain responsible for the content, accuracy, interpretation, and final manuscript. No AI system or AI-assisted technology is listed as an author.

Preprint and Prior Dissemination

The authors report that this manuscript has not previously been disseminated. It has not been published as a preprint, conference paper or abstract, thesis or dissertation, working paper, repository version, protocol, report, or translated publication. No substantially related manuscript using the same dataset or analysis has been published or submitted elsewhere.

REFERENCES

- [1] S. Dhawan, “Online learning: A panacea in the time of COVID-19 crisis,” *Journal of Educational Technology Systems*, vol. 49, no. 1, pp. 5–22, 2020, doi: [10.1177/0047239520934018](https://doi.org/10.1177/0047239520934018).
- [2] UNESCO, “Education: From disruption to recovery,” UNESCO, 2020. [Online]. Available: <https://en.unesco.org/covid19/educationresponse>
- [3] S. H. Alshammari, “The influence of technical support, perceived self-efficacy, and instructional design on students’ use of learning management systems,” *Turkish Online Journal of Distance Education*, vol. 21, no. 3, pp. 112–141, 2020, doi: [10.17718/tojde.762034](https://doi.org/10.17718/tojde.762034).
- [4] F. Martin and D. U. Bolliger, “Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment,” *Online Learning*, vol. 22, no. 1, pp. 205–222, 2018, doi: [10.24059/olj.v22i1.1092](https://doi.org/10.24059/olj.v22i1.1092).
- [5] J. A. Kumar and B. Bervell, “Google Classroom for mobile learning in higher education: Modelling the initial perceptions of students,” *Education and Information Technologies*, vol. 24, no. 2, pp. 1793–1817, 2019, doi: [10.1007/s10639-018-09858-z](https://doi.org/10.1007/s10639-018-09858-z).
- [6] K. A. Azhar and N. Iqbal, “Effectiveness of Google Classroom: Teachers’ perceptions,” *Prizren Social Science Journal*, vol. 2, no. 2, pp. 52–66, 2018.
- [7] K. R. Heggart and J. Yoo, “Getting the most from Google Classroom: A pedagogical framework for tertiary educators,” *Australian Journal of Teacher Education*, vol. 43, no. 3, pp. 140–153, 2018, doi: [10.14221/ajte.2018v43n3.9](https://doi.org/10.14221/ajte.2018v43n3.9).
- [8] R. A. S. Al-Marouf and M. Al-Emran, “Students acceptance of Google Classroom: An exploratory study using PLS-SEM approach,” *International Journal of Emerging Technologies in Learning (ijET)*, vol. 13, no. 6, pp. 112–123, 2018, doi: [10.3991/ijet.v13i06.8275](https://doi.org/10.3991/ijet.v13i06.8275).
- [9] M. Alotumi, “Factors influencing graduate students’ behavioral intention to use Google Classroom: Case study-mixed methods research,” *Education and Information Technologies*, vol. 27, no. 7, pp. 10035–10063, 2022, doi: [10.1007/s10639-022-11051-2](https://doi.org/10.1007/s10639-022-11051-2).

- [10] A. Priyadi, E. Sedyono, and H. D. Purnomo, “Evaluasi Kebergunaan (Usability) dan Rekomendasi Penggunaan Google Classroom untuk Blended Learning di Perguruan Tinggi,” *Jurnal Sistem Informasi Bisnis*, vol. 11, no. 2, pp. 105–116, 2021, doi: [10.21456/vol11iss2pp105-116](https://doi.org/10.21456/vol11iss2pp105-116).
- [11] A. Y. Pawestri and A. U. Khasanah, “The implementation of text mining to improve Google Classroom performance based on user review,” *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, vol. 13, no. 1, pp. 1–8, 2026, doi: [10.37373/tekno.v13i1.1479](https://doi.org/10.37373/tekno.v13i1.1479).
- [12] O. Pramudia, R. Pratama, S. N. Azizah, and R. Novita, “Analisis Tematik Keluhan Pengguna terhadap UI/UX Google Meet Berdasarkan Kajian Literatur tentang Ulasan Google Play Store,” *Journal Purnama Teknologi Kejuruan*, vol. 1, no. 1, pp. 49–61, 2026. [Online]. Available: <https://journal-purnama.com/index.php/purnatek/article/view/40>
- [13] K. Asafa, E. D. Sari, R. P. Adijaya, and R. P. Rahayu, “Analisis Antarmuka VClass Universitas Lampung: Evaluasi Prinsip Heuristik dan Eksplorasi Kebutuhan Sistem Pengingat Batas Waktu Tugas,” *Journal Purnama Teknologi Kejuruan*, vol. 1, no. 1, pp. 40–48, 2026. [Online]. Available: <https://journal-purnama.com/index.php/purnatek/article/view/28>
- [14] International Organization for Standardization, “*Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts*,” ISO 9241-11:2018, 2018. [Online]. Available: <https://www.iso.org/standard/63500.html>
- [15] J. Nielsen, *Usability Engineering*. San Diego, CA, USA: Academic Press, 1993.
- [16] Z. Huang and M. Benyoucef, “A systematic literature review of mobile application usability: Addressing the design perspective,” *Universal Access in the Information Society*, vol. 22, no. 3, pp. 715–735, 2023, doi: [10.1007/s10209-022-00903-w](https://doi.org/10.1007/s10209-022-00903-w).
- [17] B. A. Kumar, S. S. Chand, and M. S. Goundar, “Usability testing of mobile learning applications: A systematic mapping study,” *International Journal of Information and Learning Technology*, vol. 41, no. 2, pp. 113–129, 2024, doi: [10.1108/IJILT-03-2023-0029](https://doi.org/10.1108/IJILT-03-2023-0029).
- [18] J. Dąbrowski, E. Letier, A. Perini, and A. Susi, “Analysing app reviews for software engineering: A systematic literature review,” *Empirical Software Engineering*, vol. 27, art. no. 43, 2022, doi: [10.1007/s10664-021-10065-7](https://doi.org/10.1007/s10664-021-10065-7).
- [19] D. Pagano and W. Maalej, “User feedback in the AppStore: An empirical study,” in *Proc. 21st IEEE Int. Requirements Engineering Conf. (RE)*, 2013, pp. 125–134, doi: [10.1109/RE.2013.6636712](https://doi.org/10.1109/RE.2013.6636712).
- [20] M. Tavakoli, L. Zhao, A. Heydari, and G. Nenadić, “Extracting useful software development information from mobile application reviews: A survey of intelligent mining techniques and tools,” *Expert Systems with Applications*, vol. 113, pp. 186–199, 2018, doi: [10.1016/j.eswa.2018.05.037](https://doi.org/10.1016/j.eswa.2018.05.037).
- [21] J. Dąbrowski, E. Letier, A. Perini, and A. Susi, “Mining and searching app reviews for requirements engineering: Evaluation and replication studies,” *Information Systems*, vol. 114, art. no. 102181, 2023, doi: [10.1016/j.is.2023.102181](https://doi.org/10.1016/j.is.2023.102181).
- [22] X. Wang, T. Zhang, Y. Tan, W. Shang, and Y. Li, “How to effectively mine app reviews concerning software ecosystem? A survey of review characteristics,” *Journal of Systems and Software*, vol. 213, art. no. 112040, 2024, doi: [10.1016/j.jss.2024.112040](https://doi.org/10.1016/j.jss.2024.112040).

- [23] V. Schoenmueller, O. Netzer, and F. Stahl, “The polarity of online reviews: Prevalence, drivers and implications,” *Journal of Marketing Research*, vol. 57, no. 5, pp. 853–877, 2020, doi: [10.1177/0022243720941832](https://doi.org/10.1177/0022243720941832).
- [24] Y. Singh and P. K. Suri, “An empirical analysis of mobile learning app usage experience,” *Technology in Society*, vol. 68, art. no. 101929, 2022, doi: [10.1016/j.techsoc.2022.101929](https://doi.org/10.1016/j.techsoc.2022.101929).
- [25] D. Al-Fraihat, M. Joy, R. Masa’deh, and J. Sinclair, “Evaluating e-learning systems success: An empirical study,” *Computers in Human Behavior*, vol. 102, pp. 67–86, 2020, doi: [10.1016/j.chb.2019.08.004](https://doi.org/10.1016/j.chb.2019.08.004).
- [26] S. Eom, N. J. Ashill, J. B. Arbaugh, and J. L. Stapleton, “The role of information technology in e-learning systems success,” *Human Systems Management*, vol. 31, no. 3–4, pp. 147–163, 2012, doi: [10.3233/HSM-2012-0767](https://doi.org/10.3233/HSM-2012-0767).
- [27] P. Zaharias and A. Poylymenakou, “Developing a usability evaluation method for e-learning applications: Beyond functional usability,” *International Journal of Human–Computer Interaction*, vol. 25, no. 1, pp. 75–98, 2009, doi: [10.1080/10447310802546716](https://doi.org/10.1080/10447310802546716).
- [28] H. S. Jung, H. Lee, and K. C. Park, “Analysis of UX elements in educational applications for young children and implementation of ISO/IEC 25010 quality standards,” *SAGE Open*, vol. 15, no. 3, 2025, doi: [10.1177/21582440251377385](https://doi.org/10.1177/21582440251377385).
- [29] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology*, 4th ed. Thousand Oaks, CA, USA: SAGE Publications, 2018.
- [30] K. A. Neuendorf, *The Content Analysis Guidebook*, 2nd ed. Thousand Oaks, CA, USA: SAGE Publications, 2017.
- [31] E. Liptrot, A. Montazami, H. A. Pearson, and A. K. Dubé, “Why this app: How user ratings and app store rankings impact parents’ choice of educational apps,” *Computers & Education*, vol. 238, art. no. 105410, 2025, doi: [10.1016/j.compedu.2025.105410](https://doi.org/10.1016/j.compedu.2025.105410).



This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution, and reproduction in any medium or format, provided that appropriate credit is given to the original author(s) and the source, a link to the Creative Commons license is provided, and any modifications are indicated. Unless otherwise specified in a credit line, this article’s images or other third-party material are included under the Creative Commons license. If certain material is not covered by the article’s Creative Commons license and its intended use is not permitted by statutory regulation or exceeds the allowed usage, permission must be obtained directly from the copyright holder. <http://creativecommons.org/licenses/by/4.0/>.

AUTHOR BIOGRAPHIES



Rafiq Rizalita is an academic and lecturer in the Information Technology Education study program at the University of Lampung (Unila), Indonesia. Her research interests include digital education, instructional media development, artificial intelligence in learning, Android-based educational tools, and the development of critical-thinking and leadership skills in technology-enhanced learning contexts. She contributes to research and scholarly publication in educational technology.



Putut Aji Nalendro is an academic, lecturer, and researcher in the Information Technology Education study program at the University of Lampung (Unila). His academic interests include software engineering, multimedia systems, computer networking, and vocational education technology. His work examines the integration of digital technologies in education, including the implementation of the Kurikulum Merdeka in vocational high schools and the development of interactive learning media. He is also involved in coordinating student academic programs and publishing scholarly work.



Novi Febrianti is a lecturer and researcher in electronic engineering education. Her research includes international collaboration, including work with Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), focusing on artificial intelligence and computational approaches in engineering and education.



Publisher's and Journal's Note *Sagamedia Teknologi Nusantara, as the publisher and editor of the Journal of Hypermedia & Technology-Enhanced Learning (J-HyTEL), upholds the highest ethical standards in academic publishing. The journal remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Authors are fully responsible for the originality, accuracy, and integrity of their work. Post-publication ethical concerns will be addressed through corrections, clarifications, or retractions as necessary. The content of this publication has not been approved by the United Nations and does not reflect the views of the United Nations or its officials or Member States. <https://www.un.org/sustainabledevelopment>*
