INTERNATIONAL JOURNAL OF RESEARCH AND SCIENTIFIC INNOVATION (IJRSI)
ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

# Stakeholder Perception Gaps in Web-Based Programming Education: An ISO/IEC 25010 Evaluation in Resource-Constrained Schools

***Ian L. Tiao**

**Bestlink College of the Philippines**

***Corresponding Author**

**DOI: https://doi.org/10.51244/IJRSI.2026.1306000475**

**Received: 05 July 2026; Accepted: 10 July 2026; Published: 17 July 2026**

## ABSTRACT

Web-based programming platforms are increasingly used to expand access to computing education, yet technical quality may be perceived differently by the experts who assess systems, the teachers who implement them, and the students who use them under everyday constraints. This study examined stakeholder differences in the perceived quality of a web-based digital learning resource for programming education in resource-constrained schools in Antique, Philippines. A cross-sectional, quantitatively driven evaluation involved 105 purposively selected respondents: 65 senior high school students, 30 ICT teachers, and 10 information technology professionals. The instrument operationalized eight ISO/IEC 25010:2011 software-quality characteristics: functional suitability, performance efficiency, compatibility, usability, reliability, security, maintainability, and portability. Because variance homogeneity was not supported across the eight dimension scores, Welch's one-way analysis of variance was used, followed by Games–Howell pairwise comparisons. Internal consistency was also examined for each dimension. Significant stakeholder differences were found across all eight quality dimensions (all $p < .05$ in the recomputed analyses), with effect sizes indicating meaningful between-group variation. IT professionals generally reported the most favorable evaluations, whereas students provided the most critical ratings, particularly for qualities experienced directly in authentic use conditions. The findings show that expert validation alone is insufficient for judging educational technology readiness: software that appears technically strong may still generate a weaker experience for learners facing differences in devices, connectivity, digital literacy, and classroom conditions. The study argues for multi-stakeholder quality assurance and context-sensitive evaluation before large-scale deployment of web-based programming resources.

**Keywords:** ISO/IEC 25010, programming education, stakeholder perceptions, web-based learning, resource-constrained schools

## INTRODUCTION

Programming education requires learners to coordinate conceptual knowledge, procedural fluency, debugging, and problem solving. Web-based learning resources can widen access to practice, feedback, and structured content, particularly where specialized software or laboratory resources are limited. However, the effectiveness and readiness of such systems cannot be inferred from functionality alone. Educational technologies operate within socio-technical environments shaped by bandwidth, device availability, user expertise, classroom routines, and institutional support.

A recurring weakness in educational technology evaluation is the assumption that a favorable expert assessment represents the experience of all users. Information technology professionals may focus on architecture, conformance, security, and technical behavior; teachers may emphasize instructional fit and classroom manageability; students encounter the system as end users under actual learning conditions. These perspectives are complementary rather than interchangeable. A technically acceptable platform may therefore receive lower ratings from learners who experience latency, navigation difficulty, device constraints, or unclear interaction flows.

The ISO/IEC 25010:2011 product-quality model provides a structured framework for examining software product quality through characteristics that include functional suitability, performance efficiency, compatibility, usability, reliability, security, maintainability, and portability. In educational settings, these dimensions offer a useful basis for comparing how stakeholder groups judge the same platform. Yet comparatively little evidence has examined whether expert, teacher, and student evaluations systematically diverge in resource-constrained programming-education contexts.

This study addresses that gap by comparing the evaluations of IT professionals, ICT teachers, and senior high school students who assessed a web-based digital learning resource used for programming education in Antique, Philippines. The study asks: (1) How do the three stakeholder groups evaluate the resource across the eight ISO/IEC 25010:2011 quality dimensions? (2) Do the groups differ significantly in their evaluations? and (3) Which stakeholder contrasts account for the observed differences? The central contribution is an empirical test of the stakeholder-perception gap: the proposition that technical experts, implementers, and end users may reach systematically different judgments about the same educational technology.

### Conceptual Framework

The study adopts ISO/IEC 25010 as its principal quality framework. Functional suitability concerns whether functions meet stated needs; performance efficiency concerns response behavior and resource use; compatibility concerns coexistence and interoperability; usability concerns effective and satisfactory interaction; reliability concerns consistent operation and recovery; security concerns protection, integrity, accountability, and authenticity; maintainability concerns the ease of analysis and modification; and portability concerns adaptation and deployment across environments.

Standards-version note. The instrument used in the source study reflects the eight-characteristic product-quality model of ISO/IEC 25010:2011. The 2023 revision defines a nine-characteristic product-quality model; therefore, this article reports the instrument against the 2011 framework rather than retrospectively relabeling it as a 2023 instrument.

The framework is interpreted through a stakeholder lens. IT professionals evaluate the platform with technical expertise and may test it under relatively controlled conditions. Teachers evaluate whether it can be incorporated into instructional practice. Students evaluate the system through repeated, task-oriented use in heterogeneous real-world conditions. Consequently, disagreement among groups is treated not as measurement noise but as potentially substantive evidence about context, implementation, and user experience.

## METHODOLOGY

### Research Design and Setting

The study used a cross-sectional, quantitatively driven evaluation design supplemented in the broader dissertation project by open-ended feedback. The present article focuses on the reproducible ISO/IEC 25010 survey data. The setting comprised selected schools in the province of Antique, Philippines, where the digital learning resource was evaluated in relation to programming and ICT instruction.

### Participants and Sampling

Purposive sampling was used to recruit respondents with direct relevance to the evaluation. The verified raw dataset contained 105 respondents: 65 senior high school students enrolled in ICT-related tracks, 30 ICT teachers, and 10 IT professionals. Teacher participants were selected for relevant ICT teaching experience, while IT professionals contributed technical and software-oriented perspectives. The three-group design intentionally represented end users, instructional implementers, and technical evaluators. Because the sampling frame was purposive and the IT-professional subgroup was small ($n = 10$), the study is interpreted as a context-

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

specific stakeholder comparison rather than a population-representative estimate. The sample was not designed to support school-, municipality-, or region-level generalization.

### Instrument

The structured questionnaire operationalized eight ISO/IEC 25010:2011 software-quality characteristics using 31 item-level indicators in the available raw dataset. Responses were recorded on a five-point acceptability scale, with higher scores representing more favorable evaluations. Dimension scores were calculated as the respondent-level mean of the items belonging to each quality characteristic. The raw-data audit found no missing cells across the 31 ISO items. The instrument did not separately quantify internet bandwidth or stability, device ownership and specifications, prior programming experience, or digital-literacy level. These factors are therefore treated as plausible contextual explanations requiring direct measurement in future work, not as variables tested in the present analysis.

### Data Analysis

Analyses were recomputed from the original response worksheet rather than copied from dissertation summary tables. Descriptive statistics were calculated for each stakeholder group and quality dimension. Internal consistency was estimated using Cronbach's alpha. Brown–Forsythe/median-centered Levene tests indicated unequal variances across all eight dimensions; therefore, Welch's one-way ANOVA was selected as the omnibus test. Significant omnibus effects were followed by Games–Howell pairwise comparisons, which do not assume equal variances or equal group sizes. Eta squared ($\eta^2$) was reported as a descriptive effect-size estimate. Statistical significance was evaluated at $\alpha = .05$.

### Ethical and Data-Protection Considerations

Participation procedures in the source study emphasized voluntariness, confidentiality, and secure handling of responses. For the present reanalysis, direct identifiers were excluded from analytical outputs. Names and email addresses were not used in scoring, statistical testing, tables, or manuscript reporting.

## RESULTS

### Dataset Verification and Reliability

The raw worksheet confirmed the intended sample of 105 respondents (65 students, 30 teachers, and 10 IT professionals). No ISO item responses were missing. Cronbach's alpha across the eight dimension scales ranged from 0.755 to 0.908, indicating that reliability varied by dimension and should be interpreted at the scale-specific level rather than assumed from a single overall coefficient.

### Stakeholder Evaluations and Omnibus Differences

Table 1 presents the recomputed group means and Welch ANOVA results. Statistically significant differences were observed across all eight ISO/IEC 25010:2011 dimensions. The overall pattern was consistent with a stakeholder perception gap: IT professionals generally assigned the highest ratings, students the lowest, and teachers typically occupied an intermediate position.

| Dimension | IT Pro M | Student M | Teacher M | Welch F | df1 | df2 | p / $\eta^2$ |
|---|---|---|---|---|---|---|---|
| Functionality | 4.83 | 4.34 | 4.49 | 7.28 | 2.00 | 31.46 | $p = 0.003$; $\eta^2$=0.065 |
| Performance Efficiency | 4.83 | 4.24 | 4.57 | 15.78 | 2.00 | 38.31 | $p < .001$; $\eta^2$=0.129 |
| Compatibility | 4.95 | 4.34 | 4.58 | 24.05 | 2.00 | 50.80 | $p < .001$; $\eta^2$=0.110 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Usability | 4.83 | 4.32 | 4.61 | 11.81 | 2.00 | 34.37 | $p < .001$; $\eta^2=0.115$ |
| Reliability | 4.75 | 4.35 | 4.50 | 4.10 | 2.00 | 28.12 | $p = 0.027$; $\eta^2=0.052$ |
| Security | 4.86 | 4.34 | 4.55 | 10.90 | 2.00 | 33.40 | $p < .001$; $\eta^2=0.095$ |
| Maintainability | 4.86 | 4.30 | 4.47 | 15.02 | 2.00 | 38.83 | $p < .001$; $\eta^2=0.083$ |
| Portability | 4.87 | 4.30 | 4.53 | 10.06 | 2.00 | 30.90 | $p < .001$; $\eta^2=0.101$ |

The pattern is substantively important because the groups did not merely differ on a single technical attribute. Differences extended across functionality, efficiency, compatibility, usability, reliability, security, maintainability, and portability. This breadth suggests that stakeholder position influences the evaluation of the platform as a whole.

**Pairwise Stakeholder Contrasts**

Games–Howell comparisons were used to identify the stakeholder pairs underlying significant omnibus effects. Table 2 reports statistically significant contrasts only. Positive mean differences indicate that Group 1 rated the dimension more favorably than Group 2.

Table 2. Statistically Significant Games–Howell Pairwise Comparisons

| Dimension | Group 1 | Group 2 | Mean difference | p |
|---|---|---|---|---|
| Functionality | IT Professional | Student | 0.490 | 0.003 |
| Functionality | IT Professional | Teacher | 0.344 | 0.049 |
| Performance Efficiency | IT Professional | Student | 0.597 | < .001 |
| Performance Efficiency | Student | Teacher | -0.331 | 0.013 |
| Compatibility | IT Professional | Student | 0.612 | < .001 |
| Compatibility | IT Professional | Teacher | 0.367 | 0.002 |
| Usability | IT Professional | Student | 0.513 | < .001 |
| Usability | Student | Teacher | -0.285 | 0.014 |
| Reliability | IT Professional | Student | 0.396 | 0.028 |
| Security | IT Professional | Student | 0.519 | < .001 |
| Security | IT Professional | Teacher | 0.302 | 0.041 |
| Maintainability | IT Professional | Student | 0.558 | < .001 |
| Maintainability | IT Professional | Teacher | 0.387 | 0.016 |
| Portability | IT Professional | Student | 0.569 | < .001 |
| Portability | IT Professional | Teacher | 0.333 | 0.047 |

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

## DISCUSSION

The study provides evidence that software-quality judgments in educational technology depend on who is evaluating the system and under what conditions. Across all eight ISO/IEC 25010:2011 dimensions, the three stakeholder groups differed significantly. The general ordering—more favorable evaluations from IT professionals, more critical evaluations from students, and teacher ratings often between the two—supports the argument that technical validation cannot substitute for authentic end-user evaluation.

The student perspective is particularly important. Students encounter the platform while completing learning tasks, often on heterogeneous devices and networks and with varying levels of digital fluency. Their ratings therefore capture friction that may remain invisible during expert review. Lower student evaluations should not automatically be interpreted as lack of appreciation or technical understanding; they may signal real barriers at the point where software quality becomes educational experience.

Teacher evaluations provide a second form of situated evidence. Teachers mediate between technical capability and classroom implementation. Their judgments are shaped by instructional fit, manageability, learner support requirements, and the feasibility of integrating the resource into existing routines. The intermediate pattern observed in many dimensions is consistent with this bridging role.

Contextual factors that may explain perception gaps. The quantitative design identifies where stakeholder ratings diverge but cannot determine the causal source of those differences. In resource-constrained schools, four factors warrant particular attention: internet connectivity, device availability and capability, prior programming experience, and digital literacy. Unstable connectivity may be experienced by students as poor performance efficiency or reliability; lower-specification or shared devices may affect compatibility, portability, and interaction quality; limited programming experience may increase perceived usability burden; and differences in digital literacy may shape navigation, error recovery, and interpretation of security cues. Because these variables were not separately measured, these explanations remain hypotheses rather than empirically established mediators. Future studies should record them at respondent level and test whether they account for stakeholder differences.

The results have implications for the use of ISO/IEC 25010:2011 in educational technology. The standard offers a valuable taxonomy, but the meaning of a quality score changes with the evaluator. A high expert score may establish that a platform is technically credible; it does not establish that the platform is equally usable, reliable, or efficient for learners in constrained environments. Quality assurance should therefore be multi-stakeholder by design.

### Implications for Resource-Constrained Schools

For schools with uneven connectivity and device access, deployment decisions should prioritize the gap between controlled technical performance and actual user experience. Developers should test low-bandwidth behavior, older or lower-specification devices, recovery after interruptions, navigation clarity, and the visibility of security and system-status information. Schools should also combine platform deployment with orientation, teacher support, and fallback learning activities when connectivity is unstable.

### Implications for Developers and Evaluators

A practical evaluation sequence is recommended: expert technical review, teacher implementation review, student task-based evaluation, and iterative revision based on disagreement among groups. Rather than averaging all stakeholder ratings into a single score, evaluators should inspect divergence. A large gap can identify where technically sound design fails to translate into equitable user experience.

### Limitations

Several limitations constrain interpretation. First, purposive sampling and the single-province setting limit population-level generalization; findings should be understood as evidence from the participating schools rather than estimates for all Philippine schools. Replication should use larger, more diverse samples across multiple schools, municipalities, and regions, with stronger representation of technical evaluators than the present IT-professional subgroup (n = 10). Second, the stakeholder groups were unequal in size. Robust unequal-variance methods were used, but larger and more balanced groups would improve precision and permit subgroup analyses. Third, the study relies on self-reported perceptions. It does not establish that perceived quality corresponds to objective system performance or educational effectiveness. Future evaluations should triangulate ISO/IEC 25010 ratings with task-based usability testing (e.g., completion rate, time on task, error rate), learning analytics (e.g., logins, persistence, completion, help-seeking), programming achievement (e.g., pre/post tests, code correctness, debugging performance), and software indicators (e.g., response time, uptime, failure rate, resource consumption, and behavior under constrained bandwidth). Fourth, contextual variables that could explain stakeholder differences—including internet connectivity, device availability and specifications, prior programming experience, and digital literacy—were not separately measured. The present study therefore cannot test whether these factors mediate or moderate perception gaps. Fifth, the quantitative comparisons identify differences but provide limited evidence about why participants formed those judgments. Follow-up interviews, focus groups, think-aloud usability sessions, and open-ended incident reports should be used to identify concrete navigation, access, classroom-integration, and reliability challenges. Sixth, the available workbook did not contain populated raw responses for the separate learning-outcome block. Consequently, learning-outcome means and previously reported correlations were deliberately excluded because they could not be independently reproduced. Finally, the study concerns one web-based resource, so transferability should be tested across platforms and instructional contexts.

## CONCLUSION

The evaluation of a web-based programming resource by 105 students, teachers, and IT professionals revealed statistically significant stakeholder differences across all eight ISO/IEC 25010:2011 quality dimensions. The consistent divergence demonstrates that educational software quality is not adequately represented by expert judgment alone. Technical evaluators, classroom implementers, and student end users observe different aspects of the same system, and those differences become especially consequential in resource-constrained environments. Multi-stakeholder evaluation should therefore be treated as a core quality-assurance requirement for web-based programming education. The most useful signal may not be a high overall rating, but the gap between those who validate the technology and those who must learn through it. A next-stage evaluation should use a multi-site mixed-methods design with larger and more balanced stakeholder samples, direct measurement of connectivity, device capability, programming experience, and digital literacy, and triangulation of perception scores with task-based usability, learning analytics, programming achievement, and objective performance indicators. Interviews or focus groups should then explain the mechanisms behind the largest quantitative gaps. This design would move beyond documenting disagreement toward identifying actionable causes and testing whether platform revisions improve both technical quality and learner experience.

**Declarations**

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

Competing interests: The author declares no competing interests.

Ethics statement: The study followed the ethical standards for educational research outlined by the Rizal Technological University Graduate School. Participation was voluntary, respondents were informed of the study purpose and their right to withdraw without consequence, and confidentiality safeguards were applied. The study documentation available for this manuscript does not identify a separate research-ethics committee approval number; therefore, no approval number is reported.

Informed consent: Informed consent was obtained from respondents before data collection. For student participants who were minors, parental or guardian consent and student assent were obtained.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

Data availability: The analytical dataset contains participant-related information and is not publicly posted. A de-identified version may be made available by the author upon reasonable request, subject to applicable privacy, institutional, and consent restrictions.

Author contributions (CRediT): Ian L. Tiao: Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing, and Project Administration.

Dissertation disclosure: This article is derived from the author's doctoral dissertation submitted to Rizal Technological University. The present manuscript reports a focused reanalysis of the stakeholder software-quality evaluation dataset and has been substantially restructured for journal publication.

## REFERENCES


1. Barrot, J. S., Llenares, I. I., & del Rosario, L. S. (2021). Students' online learning challenges during the pandemic and how they cope with them: The case of the Philippines. Education and Information Technologies, 26, 7321–7338. https://doi.org/10.1007/s10639-021-10589-x
2. Bond, M., Bedenlier, S., Marín, V. I., & Händel, M. (2021). Emergency remote teaching in higher education: Mapping the first global online semester. International Journal of Educational Technology in Higher Education, 18, Article 50. https://doi.org/10.1186/s41239-021-00282-x
3. International Organization for Standardization. (2011). Systems and software engineering—Systems and software Quality Requirements and Evaluation (SQuaRE)—System and software quality models (ISO/IEC 25010:2011). ISO.
4. International Organization for Standardization. (2023). Systems and software engineering—Systems and software Quality Requirements and Evaluation (SQuaRE)—Product quality model (ISO/IEC 25010:2023). ISO.
5. Rojas, H., Renteria, R., Martinez Duran, V., Tomaylla Gutiérrez, Y., Ibarra Cabrera, M. J., & Aminuddin, A. (2025). Mapping the evolution and future directions of ISO/IEC 25010: A bibliometric and thematic analysis. Engineering, Technology & Applied Science Research, 15(5), 27530–27541. https://doi.org/10.48084/etasr.11772
6. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478. https://doi.org/10.2307/30036540