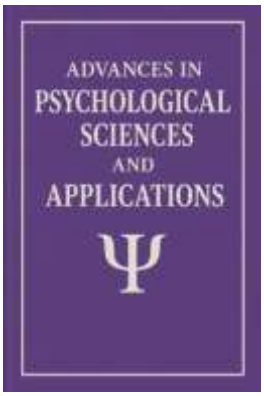
ADVANCES IN PSYCHOLOGICAL SCIENCES AND APPLICATIONS

IISTR



Student Satisfaction with Instructional Delivery

Descriptive Analysis, Reliability, and Improvement Priorities

[1]Robertus Adi Sarjono Owon, [1]Marini Abd Razak*

Affiliations:
[1] SMP Katolik Virgo Fidelis Maumere, Indonesia
[2] University Tun Abdul Razak, Kuala Lumpur, Malaysia
*Correspondence: marini@unirazak.edu.my

ABSTRACT

Student satisfaction surveys can provide information that brief observations cannot capture, particularly about the consistency of students' learning experiences. This study analyzed students' satisfaction with instructional delivery, evaluated the instrument's internal consistency and preliminary structure, and identified areas for improvement. Data comprised 233 student responses to 22 five-point Likert-type items. We retained a response when student and teacher identifiers were available in the source dataset, it was not duplicated, all recorded categories were valid, and at least 18 items were answered. A total of 221 complete and 12 partial responses met these criteria; no response was excluded. Analyses included response distributions, means, confidence intervals, top-two-box and bottom-two-box percentages, Cronbach's alpha, McDonald's omega, the Kaiser–Meyer–Olkin measure, Bartlett's test, and parallel analysis. The overall mean was 4.26/5 (teacher-cluster bootstrap 95% CI: 4.19–4.32), equivalent to an index of 81.57/100. Satisfied or very satisfied responses accounted for 91.9%, whereas dissatisfied or very dissatisfied responses accounted for 6.5%. Internal consistency was high ($\alpha = 0.905$; $\omega = 0.919$). A KMO value of 0.882 and Bartlett's test ($p < 0.001$) supported factorability, while parallel analysis indicated one factor. Priority areas were systematic and progressive presentation of material, adaptation to students' needs, connections among content, science, technology, and real life, holistic assessment, and end-of-lesson reflection. The findings support the use of student feedback for formative school improvement but do not support teacher ranking because rater counts were highly unequal.

Keywords
Formative Evaluation
Instructional Quality
Reliability
Student Feedback
Student Satisfaction

Article history
Received 5 October 2026
Revised 2 November 2026

APSA
Advances in Psychological Sciences and Applications

Vol. 2, No. 03, pp. 168-181

Copyright: © 2026 by the authors. Licensee IISTR, Yogyakarta, Indonesia. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA) license.

Accepted 5 November 2026



journal.iistr.org/index.php/APSA
DOI: 10.56741/IISTR.apsa.002755

Introduction

Instructional quality is not merely a teacher attribute; it is an experience shaped by clear objectives, presentation of content, pedagogical relationships, opportunities to participate, assessment, and follow-up. Frameworks of effective teaching indicate that teacher support, classroom management, cognitive activation, and instructional clarity complement one another [1]–[4]. In learner-oriented policy, information about these experiences should support continuous improvement rather than simply satisfy reporting requirements [5], [6].

Students are informants who directly and repeatedly experience classroom instruction. Their ratings can capture patterns that may not be visible during a single classroom observation. Studies in primary and secondary education indicate that student perceptions can provide reliable measures of teaching quality and relate to external indicators, although validity depends on the instrument, aggregation, context, and intended use [7]–[11]. Student surveys should therefore be treated as one source of evidence within an evaluation system that also incorporates observations, learning outcomes, teacher reflection, and administrative evidence [12], [13].

Student satisfaction must also be distinguished from teacher popularity. Educationally meaningful satisfaction concerns the learning experience: whether instruction is clear, relevant, inclusive, challenging, and helps students progress. Behavioral, emotional, and cognitive engagement are associated with participation and persistence in learning [14]–[16]. Positive teacher–student relationships and learner-centered support are likewise associated with engagement and achievement [17], [18]. Satisfaction scores can therefore serve as an early signal, but they must be interpreted through specific pedagogical indicators.

The instrument covered readiness to learn, alignment of objectives and content, contextual relevance, sequencing, participation, differentiation, media use, holistic assessment, interaction climate, language, reflection, summarizing, and follow-up. This scope is consistent with constructive instructional design [19], differentiated instruction [20], [21], multimedia use [22], [23], and formative assessment [24]–[29]. However, including many items does not itself prove that an instrument measures distinct subscales. Evidence about internal structure and reliability is required before dimension scores are treated as validated constructs [30].

The initial survey report summarized only some indicators and relied primarily on means. That approach can conceal distributional variation, negative responses, missing data, and unequal numbers of raters per teacher. Beyond reporting overall satisfaction, this study examines

whether favourable aggregate ratings conceal specific areas of dissatisfaction and whether the instrument's conceptual dimensions are supported by preliminary structural evidence. Its contribution lies in integrating item-level response distributions, reliability analysis, and uncertainty estimates to inform school-level improvement while recognising the limitations of teacher-level comparisons. This study aimed to: (1) describe student satisfaction across all 22 indicators; (2) assess internal consistency and the instrument's preliminary structure; (3) identify priority areas for improvement; and (4) define interpretive limits so that results are used fairly and responsibly.

Methods

This study employed a cross-sectional survey design and secondary-data analysis. The unit of analysis was one student's evaluation of one teacher. The raw data were obtained from an electronic survey form and comprised 233 responses. The survey was administered in April 2026 at Virgo Fidelis Junior High School in Maumere, East Nusa Tenggara, Indonesia. Students were recruited using a multistage random sampling procedure. Of 570 eligible students invited to participate, 233 submitted responses, yielding a response rate of 40.88%. Each submitted record corresponded to one student's evaluation of one teacher. Because the source dataset contained student and teacher names, all analyses were conducted and reported in aggregate form; no personal names are presented in the tables, figures, or discussion. Given its cross-sectional nature, the study describes students' perceptions at the time of the survey and does not support causal inference.

Data cleaning was conducted before statistical analysis. A response was eligible when student and teacher identifiers were present, the record was not duplicated, the timestamp was readable, all recorded answers fell within the five permitted response categories, and at least 18 of the 22 items (≥80%) were completed. Missing responses were not imputed, and each respondent's mean score was calculated from answered items only. Of 5,126 potential item-level responses, 14 were missing (0.27%). In total, 221 records (94.85%) were complete and 12 (5.15%) were partially complete. All 233 records met the predefined inclusion criteria and were retained for analysis.

The questionnaire was adapted from a teacher supervision instrument administered by educational supervisors and consisted of 22 statements assessing student satisfaction with instructional delivery. The questionnaire was administered in Indonesian. For descriptive interpretation, the 22 statements were grouped into five conceptual categories. These categories were used only as substantive groupings and were not assumed to represent validated subscales.

Each item used a five-point response scale: very dissatisfied = 1, dissatisfied = 2, moderately satisfied = 3, satisfied = 4, and very satisfied = 5. For each item, the valid response count, mean, standard deviation, 95% confidence interval, top-two-box percentage (satisfied + very satisfied), and bottom-two-box percentage (very dissatisfied + dissatisfied) were calculated. The overall score used all valid item responses, and a 0–100 index was calculated as ((mean − 1)/4) × 100. Although individual item responses were ordinal, means were used as supplementary descriptive statistics because the scale contained five response categories and the full response distributions were also presented. This is an acceptable descriptive approach to Likert-type data when interpretation does not depend on the mean alone [31], [32].

Uncertainty surrounding the overall score was estimated using 4,000 teacher-cluster bootstrap replications to preserve dependence among responses concerning the same teacher [33], [34]. Internal consistency was assessed using Cronbach's alpha and McDonald's omega. Alpha was reported because of its widespread use, while omega was included because alpha relies on assumptions that are often not fully met [35]–[38]. Internal structure was examined using the Kaiser–Meyer–Olkin (KMO) measure, Bartlett's test of sphericity, and parallel analysis [39]–[42]. Factor analysis was exploratory; therefore, the five conceptual categories were used only as substantive summaries and were not treated as validated subscales.

Potential areas for improvement were identified by triangulating relatively low mean scores, relatively high bottom-two-box percentages, and pedagogical importance. Importance–Performance Analysis was not used because importance was not measured separately. Teacher-level comparisons were not conducted as inferential tests or formal rankings because rater counts varied widely and most teachers had small rater groups. These decisions follow validity principles for score use: the strength and consequences of an inference should be aligned with the available evidence [30], [43].

The analysis used data previously collected by the school for instructional evaluation; no additional intervention or data collection was undertaken for this secondary analysis. Data minimization was applied by excluding student and teacher identities from all analytical outputs and by avoiding cross-tabulations that could permit re-identification. Working data were used only for the stated analytical purpose, and results were reported at the school level. No identifiable information is presented in this manuscript.

No demographic variables were analyzed because they were not required for the study objectives and their exclusion reduced disclosure risk. Consequently, subgroup differences and measurement fairness across demographic groups could not be examined.

Results

A. Data Quality and Coverage

All 233 responses were included in the analysis. Of these, 221 contained answers to all items and 12 contained answers to 18–21 items. The instrument covered 23 teachers. The number of student raters ranged from 1 to 34 per teacher, and 15 of the 23 teachers had fewer than 10 raters. This imbalance does not directly invalidate the school-level aggregate, but it substantially increases uncertainty at the teacher level and makes ranking inappropriate.

Table 1. Data and key-result summary

Indicator	Result
Responses received/used	233 / 233 (100.00%)
Responses excluded	0 (0.00%)
Complete responses	221 (94.85%)
Partial responses	12 (5.15%)
Missing answers	14 of 5.126 (0.27%)
Overall mean	4.26 of 5
Teacher-cluster bootstrap 95% CI	4.19–4.32
Satisfaction index	81.57 of 100
Top-two-box	91.9%
Bottom-two-box	6.5%

B. Distribution of Satisfaction across 22 Indicators

Overall, 91.9% of answers fell in the satisfied or very satisfied categories. Although the general pattern was positive, negative responses were not evenly distributed across items. Figure 1 therefore conveys more information than a single grand mean. Items with the largest red segments merit further investigation even when their means remain near the satisfied category.

The two highest-rated indicators were alignment of content with learning objectives and instruction delivered in accordance with learning objectives (both 4.48). At least 95.7% of responses to each were satisfied or very satisfied. By contrast, systematic and progressive presentation of material (4.02) and adaptation to students' conditions and characteristics (4.03) had the lowest means and bottom-two-box percentages of approximately 13%. The range between the lowest and highest item means was 0.46. Although all means remained above four, the distributions indicate that student experiences were less consistent for content sequencing and differentiation.

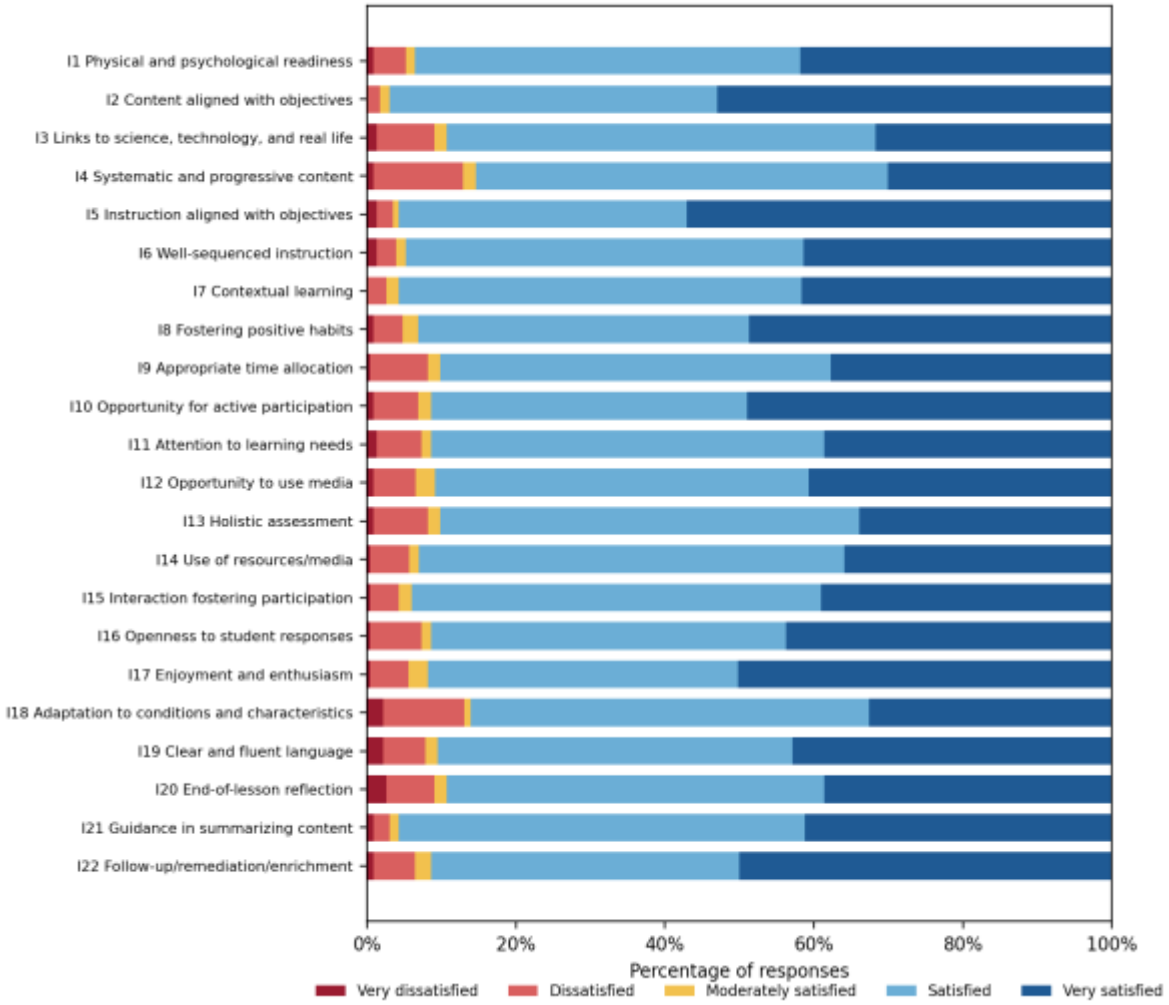


Fig. 1. Response distributions for the 22 satisfaction indicators.

Table 2. Statistics for the 22 satisfaction indicators

Code	Indicator	n	Mean	SD	Top 2 (%)	Bottom 2 (%)
I1	Physical and psychological readiness	232	4.29	0.77	93.5	5.2
I2	Content aligned with objectives	232	4.48	0.62	97.0	1.7
I3	Links to science, technology, and real life	233	4.11	0.87	89.3	9.0
I4	Systematic and progressive content	233	4.02	0.94	85.4	12.9
I5	Instruction aligned with objectives	233	4.48	0.74	95.7	3.4
I6	Well-sequenced instruction	232	4.31	0.74	94.8	3.9
I7	Contextual learning	233	4.35	0.65	95.7	2.6
I8	Fostering positive habits	232	4.36	0.78	93.1	4.7
I9	Appropriate time allocation	233	4.19	0.84	90.1	8.2
I10	Opportunity for active participation	233	4.33	0.85	91.4	6.9
I11	Attention to learning needs	233	4.21	0.84	91.4	7.3
I12	Opportunity to use media	231	4.24	0.82	90.9	6.5
I13	Holistic assessment	233	4.15	0.84	90.1	8.2
I14	Use of resources/media	231	4.23	0.75	93.1	5.6
I15	Interaction that fosters participation	233	4.28	0.72	94.0	4.3
I16	Openness to student responses	233	4.27	0.83	91.4	7.3
I17	Enjoyment and enthusiasm	233	4.36	0.80	91.8	5.6

Code	Indicator	n	Mean	SD	Top 2 (%)	Bottom 2 (%)
I18	Adaptation to conditions and characteristics	230	4.03	0.98	86.1	13.0
I19	Clear and fluent language	231	4.23	0.90	90.5	7.8
I20	End-of-lesson reflection	233	4.16	0.93	89.3	9.0
I21	Guidance in summarizing content	233	4.33	0.69	95.7	3.0
I22	Follow-up/remediation/enrichment	232	4.34	0.84	91.4	6.5

C. Means, Confidence Intervals, and Priorities

Fig. 2 presents item means with 95% confidence intervals. The dashed line marks the overall mean. The five lowest-scoring items were selected as initial priorities because they combined relatively low scores with notable negative-response proportions. Differences among points should be interpreted diagnostically rather than as definitive rankings.

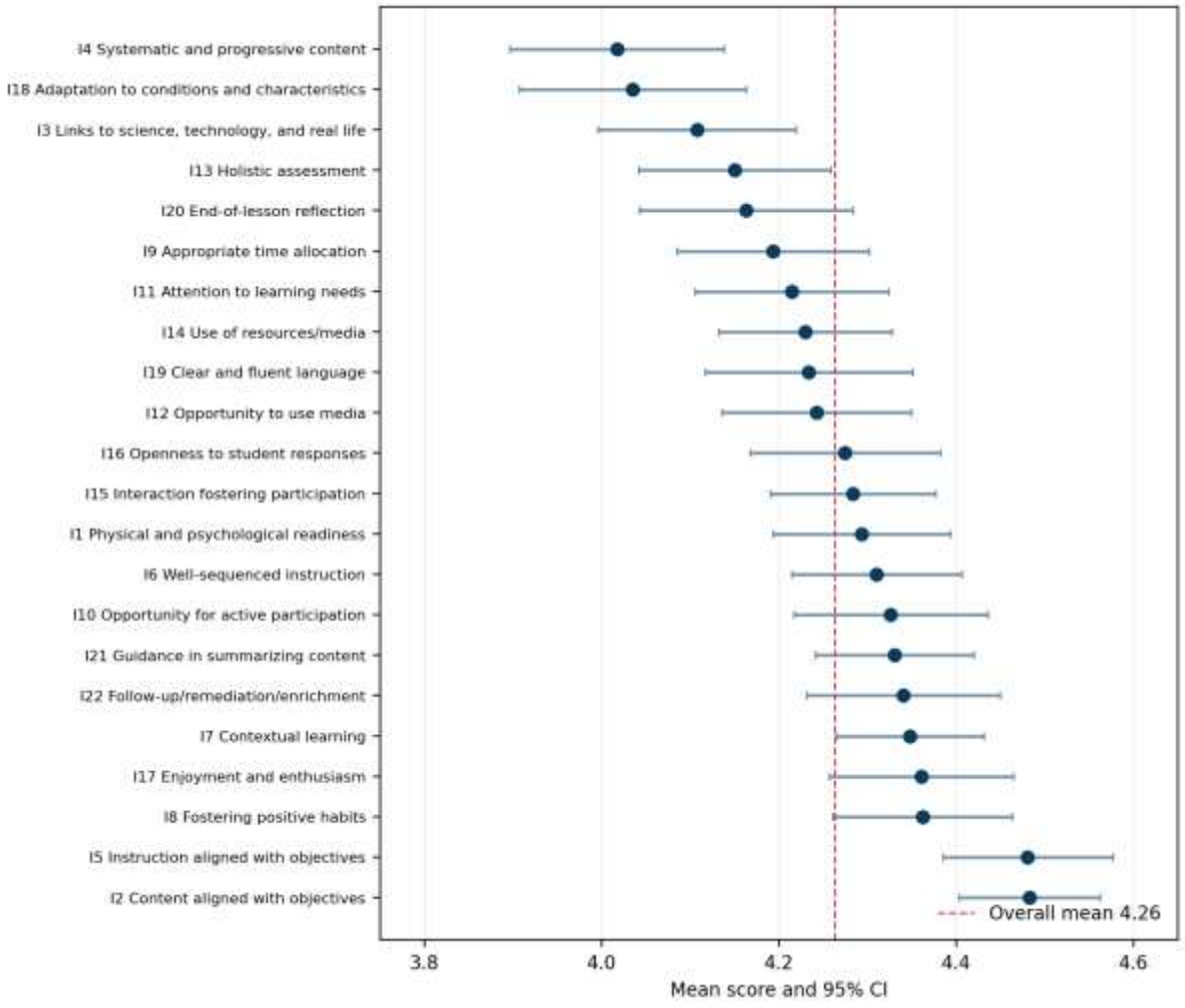


Fig. 2. Means and 95% confidence intervals for each indicator.

Table 3. Priority indicators and improvement directions

E-ISSN: 3090-7071

Priority	Indicator	Mean	Bottom 2 (%)	Improvement direction
1	Systematic and progressive content	4.02	12.9	Sequence concepts from concrete and simple to abstract and complex.
2	Adaptation to conditions and characteristics	4.03	13.0	Use brief diagnostic assessment and options for support/tasks.
3	Links to science, technology, and real life	4.11	9.0	Connect concepts with local issues, technology, or student experience.
4	Holistic assessment	4.15	8.2	Balance evidence of knowledge, attitudes, and skills using clear criteria.
5	End-of-lesson reflection	4.16	9.0	Use reflective prompts, exit tickets, and visible follow-up.

D. Reliability and Internal Structure

The scale showed high internal consistency, with Cronbach's alpha of 0.905 and McDonald's omega of 0.919. A KMO value of 0.882 indicated adequate shared variance for factor analysis, while Bartlett's test was significant, $\chi^2(231) = 1951.5$, $p < 0.001$. Parallel analysis retained one factor.

Table 4. Psychometric summary

Measure	Value	Interpretation
Cronbach's alpha	0.905	High internal consistency
McDonald's omega	0.919	High internal consistency
KMO	0.882	Adequate correlations for factor analysis
Bartlett	$\chi^2(231.0)=1951.5$; $p<0.001$	Correlation matrix differs from an identity matrix
Parallel analysis	1 factor	Structure is simpler than the five conceptual dimensions

E. Substantive Dimension Summary

Five groups of indicators were used to support program interpretation, not to claim that the instrument contained five factors. All groups had high mean scores; however, the presentation and relevance group and the differentiation and media group contained priority indicators requiring follow-up.

Table 5. Substantive dimension summary

Substantive dimension	Mean	Top 2 (%)	Alpha
Preparation and objectives	4.36	94.1	0.686
Presentation and relevance	4.23	91.5	0.690
Engagement and climate	4.31	92.2	0.755
Differentiation and media	4.18	90.4	0.642
Assessment and follow-up	4.25	91.6	0.670

Discussion

The principal finding was a positively rated learning experience: the mean was 4.26/5 and the index was 81.57/100, with more than nine of every ten responses in the satisfied or very

satisfied categories. The teacher-cluster bootstrap interval was also relatively narrow. These results indicate generally favorable student perceptions across many aspects of instructional delivery. However, the index is not a measure of academic attainment and does not establish that all instructional practices were effective. Satisfaction reflects the interaction of experience, expectations, and context; it should therefore be triangulated with classroom observation, learning assessments, and teacher reflection [12], [13].

The high scores must also be interpreted in light of the questionnaire design. Every question combined a description of practice with the phrase "satisfies you," requiring students to judge practice and satisfaction simultaneously. Although efficient, this wording may produce a halo effect in which a general impression of a teacher influences many responses. The absence of a "not able to judge" option may also lead students to select a middle or positive category when they have not observed a practice. Future revisions should separate the frequency of a practice from its quality, use single-focus statements, and include a not-able-to-judge option. These changes would make the results more actionable and strengthen content validity [30], [43].

Relative strengths appeared across communication, support, and instructional-delivery indicators. Theoretically, clear language and openness to student responses promote a psychologically safe pedagogical relationship. Meta-analyses show that teacher–student relationships and learner-centered instruction are positively associated with engagement and achievement [17], [18]. A climate in which students feel able to respond is also essential for formative assessment because teachers can adapt instruction only when student understanding becomes visible [24], [25]. Strong areas should therefore be maintained through routine practice rather than considered completed tasks.

The response distribution shows why aggregate figures should be read alongside item-level patterns. A mean above four can coexist with a small group of students having negative experiences. That group remains important for service improvement. Bottom-two-box percentages for progressive presentation and adaptation to student characteristics approached 13%, approximately one in eight valid responses. This does not demonstrate failure by any particular teacher; it indicates uneven learning experiences. The school should seek explanations through class discussion, focused observation, or safe open-ended questions to determine whether difficulties concern pace, language, prerequisites, access to media, or task format.

The first priority is systematic and progressive presentation of material. Instruction that proceeds from prior knowledge toward new understanding can reduce unnecessary cognitive load and support integration of information [22]. Improvement may include stating objectives and prerequisites, breaking complex content into manageable steps, using worked examples, checking understanding at each stage, and gradually withdrawing support. Clear sequencing does

not require one-way instruction; rather, it can create a safer structure for exploration and discussion.

The second priority is adapting instruction to students' conditions, needs, and characteristics. Effective differentiation does not require a completely separate plan for every student. It uses evidence about readiness, interests, and learning profiles to vary support, process, or products. Research syntheses indicate that differentiation can be beneficial when it is implemented systematically and supported by teacher competence [20], [21]. Schools can begin with brief diagnostic assessment, flexible grouping, tiered materials, task options, and additional support for students who need it.

The third priority concerns connecting content with science and technology and real-life contexts. Relevance helps students understand the usefulness of knowledge and can strengthen cognitive engagement [14]–[16]. Technology should follow pedagogical objectives; meta-analytic evidence indicates that technology has an effect when it enriches learning strategies rather than merely replacing an older medium [23]. Local examples, authentic problems, demonstrations, data from the surrounding environment, and small projects may be more meaningful than simply adding digital presentations.

Holistic assessment and end-of-lesson reflection were also priorities. Formative assessment is effective when evidence of learning is used to close the gap between current performance and intended goals, rather than merely to produce marks [24]–[29]. Teachers can use simple rubrics for relevant knowledge, skills, and attitudes; provide specific feedback; and offer opportunities for revision. Reflection may use exit tickets, the prompt "what do I understand and what is still unclear?", or a one-minute summary. The decisive issue is not the format but whether the information changes subsequent teaching.

Operationally, holistic assessment should not become an attempt to assess every domain in every activity. Teachers can select evidence aligned with the learning objective, explain criteria before work begins, and record only sufficient evidence. End-of-lesson reflection must also be linked to a visible response in the next lesson. If students repeatedly reflect but observe no change, the exercise will be perceived as procedural. A statement such as "based on yesterday's exit tickets, we will revisit this concept today" demonstrates that student voice is being used and strengthens feedback literacy [27], [28].

High alpha and omega values indicate that responses across items were reasonably consistent. Reliability, however, is not sufficient evidence of validity. Alpha may increase simply because a scale contains many or highly similar items [35]–[38]. The simpler structure indicated by parallel analysis suggests a strong general satisfaction factor. Future research should use polychoric correlations and ordinal confirmatory factor analysis in an independent sample,

assess convergent and discriminant validity, and test invariance across grade levels or student groups [38], [41], [44].

The one-factor result also affects reporting. A total score can summarize the general tendency in this period, but dimension scores should not be assigned overly precise diagnostic labels. Differences in dimension-level alpha partly reflect the number of items, and reliability estimates for a four-item group cannot be directly compared with those for a six-item group. Before setting dimension-specific targets, the school should test whether each group is stable in a larger sample and whether students at different grade levels interpret the items similarly. Cognitive interviews would add evidence about response processes to the statistical evidence.

Unequal rater counts are a major issue for score use. With 1–34 raters per teacher and 15 teachers having fewer than 10 raters, individual scores would have markedly different precision. Ranking could transform random differences or class composition into apparent performance differences. Research on teaching evaluation recommends triangulating evidence and using results for development, particularly when consequences are substantial [9], [12], [13]. For internal reporting, the school could set a minimum rater count, pool several periods, report uncertainty intervals, and protect teacher confidentiality.

The data-collection process should also be strengthened. Surveys should ideally be administered at comparable times across classes, explained as an improvement activity, and completed without direct supervision by the teacher being evaluated. Student identities should not be collected unless strictly necessary; anonymous codes can prevent duplicate submissions. Instructions should state that responses will not affect grades. These steps reduce social pressure and increase the likelihood of candid responses. The school should also document the target population, number invited, response rate, and reasons for nonresponse so that representativeness can be evaluated.

The practical implication is to establish a short improvement cycle: discuss aggregate results with teachers; select one or two indicators; test a small strategy over several lessons; collect classroom evidence; and review the change. This is consistent with the function of feedback as information for action [26]–[29]. Repeated surveys should retain core items for trend comparisons while adding brief open-ended questions about helpful practices and desired changes. If the school wishes to use Importance–Performance Analysis, future surveys must measure importance and performance separately.

In the next cycle, the instrument can be shortened after stronger psychometric analysis. Redundant items may be reviewed substantively, while uncovered areas—such as psychological safety, feedback clarity, accessibility, and support during difficulty—can be considered. Revisions should undergo expert review, cognitive interviews with students, pilot testing, and version

documentation. Scores should be compared across years only when wording, response scales, target populations, and administration procedures remain reasonably consistent. With this governance, the survey can evolve from an administrative exercise into a traceable organizational-learning system.

Limitations

This study has several limitations. First, the sample comprised actual respondents and lacked information on class, grade, gender, or subject; representativeness and group differences could not be examined. Second, each student evaluated one teacher and rater counts were unequal, preventing full separation of teacher, class, and student effects. Third, all items were positively worded, so acquiescence and socially desirable responding may have inflated scores. Fourth, no separate expectations or importance data, open comments, learning outcomes, or comparison observations were available. Fifth, the factor structure was not tested in an independent sample. The strongest use of these findings is therefore aggregate diagnosis and formative reflection, not causal evaluation or high-stakes decisions.

Conclusion

The 233 responses indicated high student satisfaction with instructional delivery, with a mean of 4.26/5, an index of 81.57/100, and 91.9% of answers in the satisfied or very satisfied categories. The instrument showed high internal consistency, but its five conceptual dimensions were not confirmed; parallel analysis indicated one factor. Improvement should focus on systematic and progressive content presentation, adaptation to student needs, connections with science, technology, and real life, holistic assessment, and end-of-lesson reflection. Results should be used as school-level input for professional development and triangulated with classroom observations and evidence of learning. Teacher ranking is not recommended because rater counts were highly unequal. Future surveys should strengthen anonymity, respondent characteristics, open-ended feedback, balance in rater counts, and psychometric validation.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] OECD, TALIS 2018 Results Volume I: Teachers and School Leaders as Lifelong Learners. Paris: OECD Publishing, 2019.
[2] OECD, TALIS 2018 Results Volume II: Teachers and School Leaders as Valued Professionals. Paris: OECD Publishing, 2020.
[3] R. J. Marzano, D. J. Pickering, and J. E. Pollock, Classroom Instruction That Works. Alexandria, VA: ASCD, 2001.
[4] J. Hattie, Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement. London: Routledge, 2009.
[5] UNESCO, Global Education Monitoring Report 2023: Technology in Education. Paris: UNESCO, 2023.

[6] Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, Guidelines for Learning and Assessment in Early Childhood, Primary, and Secondary Education. Jakarta, 2022.
[7] E. A. van der Scheer, H. J. Glas, and T. Visscher, "Validity and reliability of student perceptions of teaching quality in primary education," School Effectiveness and School Improvement, vol. 30, no. 1, pp. 30–50, 2019.
[8] B. Fauth, J. Decristan, A.-T. Decker, G. Büttner, I. Hardy, E. Klieme, and M. Kunter, "The effects of teacher competence on student outcomes in elementary science education," Teaching and Teacher Education, vol. 36, pp. 1–12, 2013.
[9] B. Fauth, J. Decristan, S. Rieser, M. Klieme, and G. Büttner, "Student ratings of teaching quality in primary school," Learning and Instruction, vol. 29, pp. 1–9, 2014.
[10] B. Herbert, J. Fischer, and E. Klieme, "How valid are student perceptions of teaching quality across education systems?," Learning and Instruction, vol. 82, Art. no. 101652, 2022, doi: 10.1016/j.learninstruc.2022.101652.
[11] T. J. Kane and D. O. Staiger, Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Seattle, WA: Bill & Melinda Gates Foundation, 2012.
[12] R. A. Berk, "Survey of 12 strategies to measure teaching effectiveness," International Journal of Teaching and Learning in Higher Education, vol. 17, no. 1, pp. 48–62, 2005.
[13] B. Spooren, B. Brockx, and D. Mortelmans, "On the validity of student evaluation of teaching," Review of Educational Research, vol. 83, no. 4, pp. 598–642, 2013.
[14] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement," Review of Educational Research, vol. 74, no. 1, pp. 59–109, 2004.
[15] E. A. Skinner, T. A. Kindermann, and C. J. Furrer, "A motivational perspective on engagement and disaffection," Educational and Psychological Measurement, vol. 69, no. 3, pp. 493–525, 2009.
[16] J. Reeve and C.-M. Tseng, "Agency as a fourth aspect of students' engagement," Contemporary Educational Psychology, vol. 36, no. 4, pp. 257–267, 2011.
[17] D. L. Roorda, H. M. Y. Koomen, J. L. Spilt, and F. J. Oort, "The influence of affective teacher–student relationships," Review of Educational Research, vol. 81, no. 4, pp. 493–529, 2011.
[18] J. Cornelius-White, "Learner-centered teacher-student relationships are effective," Review of Educational Research, vol. 77, no. 1, pp. 113–143, 2007.
[19] J. Biggs and C. Tang, Teaching for Quality Learning at University, 4th ed. Maidenhead: Open University Press, 2011.
[20] C. A. Tomlinson, The Differentiated Classroom, 2nd ed. Alexandria, VA: ASCD, 2014.
[21] M. I. Deunk, A. E. Smale-Jacobse, H. de Boer, S. Doolaard, and R. J. Bosker, "Effective differentiation practices," Educational Research Review, vol. 24, pp. 31–54, 2018.
[22] R. E. Mayer, Multimedia Learning, 2nd ed. New York: Cambridge University Press, 2009.
[23] R. M. Tamim, R. M. Bernard, E. Borokhovski, P. C. Abrami, and R. F. Schmid, "What forty years of research says about technology," Review of Educational Research, vol. 81, no. 1, pp. 4–28, 2011.
[24] P. Black and D. Wiliam, "Assessment and classroom learning," Assessment in Education, vol. 5, no. 1, pp. 7–74, 1998.
[25] D. R. Sadler, "Formative assessment and the design of instructional systems," Instructional Science, vol. 18, pp. 119–144, 1989.
[26] J. Hattie and H. Timperley, "The power of feedback," Review of Educational Research, vol. 77, no. 1, pp. 81–112, 2007.
[27] D. J. Nicol and D. Macfarlane-Dick, "Formative assessment and self-regulated learning," Studies in Higher Education, vol. 31, no. 2, pp. 199–218, 2006.
[28] D. Carless and D. Boud, "The development of student feedback literacy," Assessment & Evaluation in Higher Education, vol. 43, no. 8, pp. 1315–1325, 2018.
[29] F. M. van der Kleij, R. C. W. Feskens, and T. J. H. M. Eggen, "Effects of feedback in a computer-based learning environment," Review of Educational Research, vol. 85, no. 4, pp. 475–511, 2015.
[30] AERA, APA, and NCME, Standards for Educational and Psychological Testing. Washington, DC: AERA, 2014.
[31] G. Norman, "Likert scales, levels of measurement and the laws of statistics," Advances in Health Sciences Education, vol. 15, pp. 625–632, 2010.
[32] G. M. Sullivan and A. R. Artino, "Analyzing and interpreting data from Likert-type scales," Journal of Graduate Medical Education, vol. 5, no. 4, pp. 541–542, 2013.
[33] L. Kish, Survey Sampling. New York: Wiley, 1965.
[34] B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap. New York: Chapman & Hall, 1993.

[35] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," Psychometrika, vol. 16, pp. 297–334, 1951.
[36] R. P. McDonald, Test Theory: A Unified Treatment. Mahwah, NJ: Erlbaum, 1999.
[37] T. J. Dunn, T. Baguley, and V. Brunsden, "From alpha to omega," British Journal of Psychology, vol. 105, no. 3, pp. 399–412, 2014.
[38] A. M. Gadermann, M. Guhn, and B. D. Zumbo, "Estimating ordinal reliability for Likert-type data," Practical Assessment, Research & Evaluation, vol. 17, no. 3, pp. 1–13, 2012.
[39] H. F. Kaiser, "An index of factorial simplicity," Psychometrika, vol. 39, pp. 31–36, 1974.
[40] M. S. Bartlett, "A note on the multiplying factors for various chi square approximations," Journal of the Royal Statistical Society Series B, vol. 16, no. 2, pp. 296–298, 1954.
[41] J. L. Horn, "A rationale and test for the number of factors in factor analysis," Psychometrika, vol. 30, pp. 179–185, 1965.
[42] A. B. Costello and J. W. Osborne, "Best practices in exploratory factor analysis," Practical Assessment, Research & Evaluation, vol. 10, no. 7, pp. 1–9, 2005.
[43] S. Messick, "Validity of psychological assessment," American Psychologist, vol. 50, no. 9, pp. 741–749, 1995.
[44] F. P. Holgado-Tello, S. Chacón-Moscoso, I. Barbero-García, and E. Vila-Abad, "Polychoric versus Pearson correlations in factor analysis of ordinal variables," Quality & Quantity, vol. 44, pp. 153–166, 2010.

Authors



Robertus Adi Sarjono Owon was born in Nitakloang on September 20, 1969. He serves as a teacher in SMP Katolik Virgo Fidelis Maumere, Indonesia. He has authored numerous books and academic publications. His research interests primarily focus on local wisdom, cultural values, and local policies across diverse ethnic communities in Indonesia. (email: robertusadi99@gmail.com).



Asst. Prof. Dr. Marini Abd Razak is an Assistant Professor at the Tun Razak Graduate School, Universiti Tun Abdul Razak (UNIRAZAK), Malaysia. Her academic work focuses on graduate education, research methodology, and interdisciplinary studies in education and social sciences. She is actively involved in postgraduate teaching, academic supervision, and research collaboration. Her scholarly interests include educational development, higher education, and evidence-based research practices that contribute to academic excellence and institutional capacity building. (email: marini@unirazak.edu.my).