ISSN: 2181-3337 | SCIENTISTS.UZ VOLUME 5 ISSUE 7 JULY 2026

FORMATIVE ASSESSMENT IN SPECIALIZED SCHOOLS IN UZBEKISTAN: A PILOT STUDY OF PRACTICES, BARRIERS, AND CONDITIONS FOR SCALING

Mirzazhanova Nargiza
Tashkent State University of Economics
shyuldashev@tsue.uz

DOI: 10.67620/sai.57b0015

ABSTRACT

This article examines formative assessment practices in one specialized school and two Presidential Schools in Uzbekistan, as well as the organizational conditions that should be considered when transferring selected practices to the broader general secondary education system. The study employs a pilot, multi-source, predominantly qualitative design supported by descriptive quantitative evidence. The empirical material comprises 12 structured lesson observations, 15 semi-structured interviews with teachers, six focus groups with students in Grades 9–11 (n = 24), and 45 samples of written feedback. The mean descriptive score on the author-developed observation scale was 4.7 out of 7; systematic student involvement in self-assessment emerged as the weakest domain. Explicit self-assessment or peer-assessment procedures based on criteria were observed in 4 of the 12 lessons. Of the 45 written comments, 18 (40.0%) focused on the performance process or self-regulation, whereas 11 (24.4%) consisted of general evaluations of the learner or effort without specific instructional information. Students valued specific comments while simultaneously maintaining a strong orientation toward marks. Through comparison across data sources, a conceptual four-phase scaling model is proposed: diagnostic piloting, contextual adaptation, controlled expansion, and institutionalization. The findings do not permit an assessment of the causal effect of formative assessment on human-potential development, but they identify practical conditions for its more consistent implementation.

ACCEPTED

05.08.26

Keywords: *formative assessment, pedagogical feedback, self-assessment, assessment literacy, specialized schools, education policy, Uzbekistan.*

CC BY

Introduction

Amid digitalization, changes in the structure of employment, and growing expectations for independent learning, education systems increasingly associate the quality of schooling not only with the acquisition of subject content but also with students' capacity to plan, monitor, and adjust their own learning. The Education Policy Outlook 2025 describes this capacity through a combination of will, skills, and means that enable individuals to continue learning throughout life (OECD, 2025a). Formative assessment can support such learner agency when information about students' current understanding is used to modify the subsequent actions of teachers, students themselves, or their peers.

The classic definition offered by Black and Wiliam (1998) emphasizes the functional rather than instrumental nature of formative assessment: the central issue is not whether a separate test, rubric, or

comment is present, but whether the information obtained is used to improve subsequent learning. The authors later connected formative assessment to a broader theory of pedagogy and to interactions among the teacher, the learner, and the subject matter (Black & Wiliam, 2018). Formative quality, therefore, does not arise automatically from an international curriculum, a digital platform, or a new reporting system.

This issue is particularly relevant to Uzbekistan because of the rapid expansion of Presidential and specialized schools, the introduction of international curricula, and the simultaneous objective of improving the quality of mainstream general education. Government documents report a substantial increase in funding for the school sector, the creation of additional school places, and greater attention to teacher development and quality assessment (President of the Republic of Uzbekistan, 2025). Nevertheless, the availability of infrastructure and curricula does not guarantee an equivalent depth of change in classroom practice.

The original version of the study combined overly broad claims concerning human-potential development, national policy, and the effectiveness of formative assessment. The present revised version retains the central idea but restricts its conclusions to what is directly supported by the data presented: feedback practices, student participation, perceptions of assessment, and implementation conditions in the three selected institutions.

The purpose of the study is to describe how formative assessment is implemented in selected specialized schools in Uzbekistan, identify recurring barriers and enabling conditions, and use this evidence to propose a cautious conceptual model for further scaling.

The study addresses three research questions:

1. Which formative assessment practices are observed during lessons and in written feedback?
2. How do teachers and students perceive the strengths and limitations of these practices?
3. Which conditions should be considered when transferring formative assessment practices beyond the pilot institutions?

2. Theoretical Foundations

2.1. Formative Assessment as a Process of Using Evidence of Learning

Formative assessment should be distinguished from frequent testing and from the mere provision of comments. Assessment becomes formative only when evidence of a student's current understanding leads to a justified next step. Such a step may involve modifying an explanation, regrouping students, repeating a task, setting an individual goal, conducting criterion-referenced self-assessment, or adjusting a problem-solving strategy.

Bennett (2011) cautions that generalized claims about the effectiveness of formative assessment often conceal differences among specific techniques, the quality of their implementation, and the conditions in which they are applied. Accordingly, analysis of practice should focus not on the presence of an isolated technique but on the complete sequence: learning goal, evidence elicitation, interpretation, and action. A break between information collection and subsequent action frequently reduces formative assessment to a procedural formality.

Cambridge International defines assessment for learning as an approach embedded in teaching and learning that enables teachers and students to answer three questions: where the learner is going, where the learner is now, and how the gap between current and expected performance can be closed (Cambridge University Press & Assessment, 2025). This approach requires clear goals and criteria, effective questioning, the use of evidence of learning, and the development of learner autonomy.

2.2. Feedback, Self-Assessment, and Self-Regulation

The model developed by Hattie and Timperley (2007) distinguishes feedback directed at the task, the process, self-regulation, and the self. Task-level feedback can help correct a specific error, but its transfer to new situations is limited. Process-level feedback explains the strategy or reasoning underlying performance, whereas self-regulation feedback supports planning, self-checking, and the adjustment of

ISSN: 2181-3337 | SCIENTISTS.UZ

actions. General praise or criticism directed at the person, without reference to a criterion, usually provides students with little information about what to do next.

A meta-analysis of educational feedback demonstrates that its effects are heterogeneous and depend on its content, level, and mode of use (Wisniewski, Zierer, & Hattie, 2020). Similarly, self-assessment has its most consistent influence not when students are simply asked to "assess themselves," but when it is connected to criteria, reflection, and the subsequent revision of work (Panadero, Jönsson, & Botella, 2017). The present study therefore considers not only whether a comment is provided, but also whether it can support the learner's next instructional action.

2.3. Scaling and the Context of Implementation

Scaling a pedagogical practice is not a matter of mechanically copying an instrument from one school to another. Research indicates that change depends on teachers' assessment literacy, the quality of materials, local marking policies, opportunities for professional collaboration, and leadership support (Evans, 2022). The OECD (2025b) likewise treats formative assessment and feedback as practices whose quality depends on the school environment and professional support. In less-resourced or examination-oriented contexts, international ideas may be reinterpreted, reduced to their external form, or brought into conflict with established expectations among teachers and students (Chen, Zhang, & Li, 2021).

Improvement in formative practice generally results from a combination of strategies rather than the introduction of a single technique. Andersson and Palm (2017) show that meaningful changes include clarifying goals, systematically eliciting evidence of learning, using criteria, activating students, and adapting teaching. Scaling should therefore assess not the number of training sessions delivered or rubrics distributed, but the quality of changes in everyday classroom interactions.

Table 1

Key Conditions for Implementing Formative Assessment Identified in the Literature

Condition	Description	Risk if Absent
Assessment literacy	Understanding of goals, criteria, evidence of learning, and forms of subsequent action.	Reduction of formative assessment to a test, mark, or general praise.
Alignment of materials	Rubrics, tasks, and criteria are connected to learning objectives and the actual course content.	Fragmented techniques that do not influence teaching.
Time and collaboration	Teachers have opportunities to discuss student work and plan instructional adjustments.	Collection of data that cannot be used because of workload pressures.
Leadership support	School leaders encourage safe experimentation and do not use formative data as a punitive control mechanism.	Compliance-oriented implementation and resistance.
Active student role	Students understand criteria, use feedback, and participate in self-assessment.	Passive receipt of comments without any change in action.

3. National and Institutional Context

Since 2017, school education in Uzbekistan has undergone an intensive period of institutional renewal. According to official data, by 2025 funding for preschool and school education had increased sevenfold, approximately one million additional school places had been created, and the system of supplementary teacher payments had begun to reflect qualifications, language proficiency, and the results of professional appraisal (President of the Republic of Uzbekistan, 2025). The 2023–2026 Partnership

ISSN: 2181-3337 | SCIENTISTS.UZ

Compact for Education Reform similarly emphasizes the need to expand infrastructure, develop the teaching workforce, and improve the assessment of learning (UNICEF, 2023).

Presidential and specialized schools serve as sites for implementing advanced curricula, contemporary teaching methods, and international approaches to assessment. Part of the network has received Cambridge International accreditation and delivers IGCSE and A Level programmes (Agency for the Development of Presidential, Creative, and Specialized Schools, n.d.). However, institutions within this network differ in their histories, staffing profiles, curricula, and access to external professional support. They should therefore not be treated as a homogeneous model.

The transferability of these practices is particularly important for the mainstream system. Specialized schools use selective admission, have comparatively more favourable resources, and operate within a different professional environment. Their experience may therefore provide hypotheses and prototypes, but it does not constitute direct evidence that the same results will be obtained in ordinary schools without contextual adaptation.

4. Methodology

4.1. Research Design

The study was conducted as a pilot, multi-source investigation with a predominantly qualitative design and descriptive quantitative support. It is neither an experiment nor a national monitoring exercise or a representative comparison of schools. Quantitative indicators are used to summarize the observed cases rather than to support causal conclusions or statistical generalization to the education system as a whole.

Data were collected between February and May 2026 in three institutions: one specialized institution in Tashkent and two Presidential Schools in Samarkand and Namangan. The institutions were selected purposively to provide variation in geographical and institutional context. This form of sampling does not provide statistical representativeness.

4.2. Participants and Data Sources

Table 2

Composition of the Empirical Material

Source	Volume	Distribution	Purpose
Lesson observations	12 lessons	Four in each institution; mathematics, science, and English	Description of feedback, self-assessment, and instructional adaptation practices
Teacher interviews	15 interviews	Five in each institution	Identification of perceptions, barriers, and enabling conditions
Focus groups	6 groups; n = 24	Two groups in each institution; Grades 9–11; 13 girls	Analysis of perceptions of marks, comments, self-assessment, and motivation
Written feedback	45 samples	15 from each institution	Classification of the level and instructional orientation of comments

The interviewed teachers had a mean teaching experience of 8.4 years (SD = 3.2). According to the source dataset, 11 of the 15 teachers had completed professional learning related to Cambridge programmes, while four had experience working in international schools outside Uzbekistan.

4.3. Instruments

Lesson observations employed an author-developed pilot scale comprising eight indicators scored from 1 (“not observed”) to 7 (“consistent, high-quality practice”). The indicators were conceptually linked

ISSN: 2181-3337 | SCIENTISTS.UZ

to the literature on learning goals and criteria, feedback quality, self-assessment, peer assessment, instructional adaptation, self-regulation, and emotional safety (Black & Wiliam, 2018; Hattie & Timperley, 2007; Cambridge University Press & Assessment, 2025). The scale is not an official CLASS instrument and has no externally validated thresholds. Scores are therefore interpreted descriptively only.

The semi-structured teacher interviews addressed understandings of formative assessment, typical techniques, the use of criteria, time constraints, the influence of school policy, and professional support. Student focus groups explored how students received information about errors, the comparative value of marks and comments, experiences of self-assessment, and the motivational consequences of assessment.

Written comments were classified according to the four levels in Hattie and Timperley's (2007) model: task, process, self-regulation, and self. Process- and self-regulation-level comments were treated as the most transferable to subsequent action; task-level comments were regarded as useful but limited to a specific error; and general praise or criticism without instructional guidance was classified as low in informational value.

4.4. Data Analysis

The interviews and focus groups were transcribed and analyzed thematically in NVivo 14 using Braun and Clarke's (2006) six-phase procedure. Quantitative characteristics of the observations and artefacts were summarized in SPSS 29 using frequencies, proportions, means, and standard deviations. Because of the small sample and its nested structure, inferential statistics were not used in the revised version.

Data sources were integrated during interpretation: observed practices were compared with teacher and student accounts and with the content of written feedback. Convergence across multiple sources was treated as strengthening an interpretation, but not as evidence of causality.

4.5. Ethical and Methodological Limitations of the Procedure

The source documentation reports ethical approval under Protocol No. 045 dated 15 March 2026. The findings are presented in aggregate form, without identifying students or teachers.

The observations and coding were conducted as part of a pilot project; independent double coding and an assessment of inter-rater reliability were not reported. This limits the reproducibility of categorical decisions. In addition, observations of 12 lessons cannot distinguish stable institutional differences from the characteristics of particular teachers, subjects, or observation days.

5. Results

5.1. Observed Formative Assessment Practices

The mean descriptive score on the author-developed observation scale was 4.7 out of 7. The highest mean score was obtained for feedback quality (4.9), followed by the adaptability of instructional decisions (4.8). Student involvement in self-assessment received the lowest mean score (4.4). These values should not be interpreted as normative levels of quality because the scale has no externally validated thresholds.

Mean scores differed across institutions: Tashkent, 5.4; Samarkand, 4.5; and Namangan, 4.2. With only four observations per institution, these differences are descriptive. They may reflect a combination of teacher characteristics, subjects, class composition, and access to professional support, but they do not permit the effect to be attributed to the institution itself.

Structured self-assessment or peer assessment using explicit criteria was observed in 4 of the 12 lessons (33.3%). In the remaining cases, students were more often recipients of comments than participants in interpreting criteria or setting a subsequent goal. This distinction is important: the presence of oral feedback alone does not demonstrate the development of students' capacity to use feedback independently.

Teachers used rubrics and specified errors in tasks most consistently. A complete formative cycle—in which the information obtained directly changed the subsequent task, the pace of the lesson, or an individual learner strategy—was observed less frequently.

ISSN: 2181-3337 | SCIENTISTS.UZ

Table 3

Descriptive Observation Results by Formative Assessment Domain (N = 12)

Domain	Tashkent (n = 4)	Samarkand (n = 4)	Namangan (n = 4)	Overall (N = 12)
Feedback quality	5.6	4.8	4.3	4.9
Involvement in self-assessment	5.2	4.2	3.9	4.4
Adaptability of decisions	5.5	4.5	4.4	4.8
Overall mean score	5.4	4.5	4.2	4.7

5.2. Student Perceptions

The focus groups revealed a dual orientation among students. Nineteen of the 24 participants (79.2%) stated that they would prefer to receive specific comments more frequently instead of a mark alone. At the same time, 14 students (58.3%) reported that a mark motivated them more strongly than a comment. These proportions describe the statements of focus-group participants and are not results from a representative survey.

This duality may be related to the competitive educational environment: a comment helps students understand what should be improved, whereas a mark remains a readily understood symbol of status and progress. A transition toward a formative culture therefore requires not the abolition of summative assessment, but a clear separation of functions: a mark records performance at a particular stage, whereas a comment directs subsequent action.

5.3. Quality of Written Feedback

Table 4

Classification of Written Feedback by Level (N = 45)

Level	n	%	Example	Interpretation
Task	16	35.6	“Incorrect; review the formula on p. 24.”	Specific, but limited to the current error.
Process	8	17.8	“The method was appropriate, but the error occurred during the calculation stage.”	Helps the student understand the performance strategy.
Self-regulation	10	22.2	“Before submission, check the logical sequence of your arguments.”	Supports planning and self-checking.
Self	11	24.4	“Well done!” / “You need to try harder.”	Does not indicate exactly what or how to improve.
Total	45	100.0	—	—

Process- and self-regulation-level comments together accounted for 18 of the 45 samples (40.0%). This combined category was identified in the source materials as having the greatest formative value. A further 16 comments (35.6%) indicated correctness or error in a specific task but did not always explain how the guidance could be transferred to subsequent work. Eleven comments (24.4%) consisted of general praise or an exhortation to make greater effort.

The artefacts confirm the observed pattern: teachers frequently provide information about an error, but less often formulate comments in a way that enables the learner to plan the next step independently.

ISSN: 2181-3337 | SCIENTISTS.UZ

5.4. Comparison Across Data Sources

Comparison of the observations, focus groups, and written materials revealed three recurring patterns. First, formative techniques were present, but their completeness varied: criteria and rubrics were used more frequently than procedures supporting student self-regulation. Second, students understood the value of explanatory comments, yet the institutional importance of marks continued to exert a strong influence on motivation. Third, the quality of practice was related not only to the school's formal curriculum but also to the professional decisions of individual teachers and the support available to them.

These findings do not demonstrate that specialized schools are superior to general schools because no common comparative instrument or primary data from the mainstream system were available. They show that even in a relatively favourable environment, formative assessment remains a complex practice requiring pedagogical, organizational, and cultural support.

6. Discussion

6.1. From the Presence of Instruments to the Use of Information

The first finding is that the use of rubrics and comments does not in itself constitute a complete formative cycle. The results are consistent with Bennett's (2011) critique: effects depend on how evidence of learning is interpreted and translated into action. In the observed lessons, information about the quality of current work was more prominent than students' active use of that information.

The second finding concerns the level of feedback. Comments directed at the process and self-regulation accounted for 40.0% of the sample. This does not imply that the remaining comments were useless; identifying an error may be necessary. Without an explanation of the strategy or the next step, however, the learner remains dependent on the teacher. Such dependence limits the connection between formative assessment and the broader objective of developing learner autonomy.

The third finding is the continued strong role of marks. Students' desire for substantive comments coexisted with their recognition of the stronger motivational force of numerical or categorical marks. This duality demonstrates that formative and summative assessment should not be presented as a simple opposition. The task is to ensure functional clarity and coherence: summative assessment may record achievement, but it should not displace the information required for improvement.

6.2. Conditions for Scaling

Five critical conditions were identified on the basis of the literature and the pilot findings. They constitute an analytical proposal rather than an empirically validated national model.

1. Practice-oriented teacher assessment literacy. Professional learning should include collaborative analysis of authentic student work, formulation of criteria, design of feedback comments, and planning of the next step, rather than only familiarization with terminology.
2. Coherence between formative and summative requirements. National marking practices, school reporting requirements, and curriculum pacing should not encourage teachers to collect information that cannot be used to modify learning.
3. Organizational time and professional collaboration. Teachers need regular opportunities to discuss criteria, compare samples, and revise tasks. Without such opportunities, formative assessment increases workload but does not necessarily improve instructional decisions.
4. Student assessment literacy. Students should be taught to understand criteria, distinguish between the quality of their work and judgments of themselves as persons, use feedback, and set verifiable goals.
5. Context-sensitive monitoring. Monitoring should assess not the number of rubrics or training sessions, but the quality of feedback, student participation, and actual changes in subsequent teaching.

ISSN: 2181-3337 | SCIENTISTS.UZ

6.3. A Conceptual Model for Phased Expansion

Table 5

A Four-Phase Conceptual Model for Scaling

Phase	Purpose	Key Actions	Indicators of Readiness
1. Diagnostic piloting	Understand existing practice and constraints.	Lesson observations; analysis of student work; interviews; agreement on a minimum set of indicators.	A baseline picture of practices and local barriers is available.
2. Contextual adaptation	Test instruments under authentic conditions.	Co-development of criteria; small cycles of trial implementation; workload analysis; student training in self-assessment.	Teachers and students can use criteria without continuous external support.
3. Controlled expansion	Transfer the practice to diverse schools.	Cluster-based implementation; mentors; comparable observations; adjustment of materials by school type.	Quality is maintained across several contexts rather than only in the original institutions.
4. Institutionalization	Embed the practice in sustainable systems.	Alignment with appraisal and curricula; preparation of new teachers; regular moderation; monitoring of feedback use.	The practice is supported by rules, resources, and professional communities.

The model operates simultaneously at four levels: policy establishes coherent requirements; agencies and institutions provide materials and professional learning; the school creates time, leadership, and a culture of collaboration; and the classroom translates criteria and feedback into specific actions by teachers and students. Failure at any level may produce formal implementation without a change in learning.

For Uzbekistan, the most rational next step is not the immediate nationwide prescription of a single instrument, but a series of comparable pilots in urban, rural, single-shift, and double-shift schools. Specialized institutions may serve as centres of professional support, but materials should be co-developed and tested with teachers from the mainstream system.

6.4. Study Limitations

The study has five principal limitations. First, the sample was small and purposively selected; the findings do not represent all Presidential, specialized, or general schools. Second, four observations per institution do not permit reliable assessment of stable differences among schools. Third, the author-developed observation scale has not undergone independent psychometric validation, and inter-rater reliability was not reported. Fourth, the focus-group proportions are based on a small number of participants and are influenced by group dynamics. Fifth, the study did not measure changes in students' academic, metacognitive, or socio-emotional outcomes.

The findings should therefore be understood as a description of practices and implementation conditions, rather than as evidence of effectiveness or a causal influence on human potential. Testing such effects would require validated baseline and outcome measures, comparable groups, and a longer period of observation.

7. Conclusion

The pilot study indicates that formative assessment is already represented in the selected specialized institutions in Uzbekistan through rubrics, the specification of errors, and selected forms of instructional adaptation. Nevertheless, students' active use of criteria and feedback remains the most vulnerable component. Only 4 of the 12 observed lessons included explicit self-assessment or peer assessment based on criteria, while 40.0% of written comments focused on the process or self-regulation.

The principal practical conclusion is that scaling should concern not a set of external techniques but schools' capacity to transform evidence of learning into subsequent action. This requires assessment

ISSN: 2181-3337 | SCIENTISTS.UZ

literacy among teachers and students, coherence with the summative assessment system, time for collaborative work, leadership support, and monitoring of the actual use of feedback.

The proposed four-phase model—diagnostic piloting, contextual adaptation, controlled expansion, and institutionalization—provides a sequence for further testing. Future studies should include a more diverse range of schools, independent coding of observations, validated instruments, and longitudinal indicators of learning outcomes. Only then will it be possible to evaluate with sufficient justification the contribution of formative assessment to the broader development of students' human potential.

REFERENCES

1. Agency for the Development of Presidential, Creative, and Specialized Schools. (n.d.). Presidential schools successfully passed Cambridge Assessment International Education accreditation [in Uzbek]. https://portal.piima.uz/uz/news/prezident-maktablari-cambridge-assessment-international-education-akkreditatsiyasidan-muvaffaqiyatli-otdi
2. President of the Republic of Uzbekistan. (2025, May 15). A new stage of reforms in school education [in Russian]. https://president.uz/ru/lists/view/8127
3. Andersson, C., & Palm, T. (2017). Characteristics of improved formative assessment practice. Education Inquiry, 8(2), 104–122. https://doi.org/10.1080/20004508.2016.1275185
4. Bennett, R. E. (2011). Formative assessment: A critical review. Assessment in Education: Principles, Policy & Practice, 18(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678
5. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74. https://doi.org/10.1080/0969595980050102
6. Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. Assessment in Education: Principles, Policy & Practice, 25(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807
7. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
8. Cambridge University Press & Assessment. (2025). Assessment for learning: Education brief. https://www.cambridgeinternational.org/Images/271179-assessment-for-learning.pdf
9. Chen, Q., Zhang, J., & Li, L. (2021). Problematising formative assessment in an undeveloped region of China: Voices from practitioners. Educational Assessment, Evaluation and Accountability, 33(4), 649–673. https://doi.org/10.1007/s11092-021-09369-5
10. Evans, C. (2022, August 24). Supporting the implementation of formative assessment. Center for Assessment. https://www.nciea.org/blog/supporting-the-implementation-of-formative-assessment/
11. Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487
12. OECD. (2025a). Education policy outlook 2025: Nurturing engaged lifelong learners. OECD Publishing. https://doi.org/10.1787/c3f402ba-en
13. OECD. (2025b). Unlocking high-quality teaching. OECD Publishing. https://doi.org/10.1787/f5b82176-en
14. Panadero, E., Jönsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. Educational Research Review, 22, 74–98. https://doi.org/10.1016/j.edurev.2017.08.004
15. UNICEF. (2023). Uzbekistan partnership compact for education reform, 2023–2026. https://www.unicef.org/uzbekistan/media/5671/file/Partnership%20Compact%20for%20Education%20reform.pdf
16. Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. Frontiers in Psychology, 10, Article 3087. https://doi.org/10.3389/fpsyg.2019.03087

ISSN: 2181-3337 | SCIENTISTS.UZ

Appendix A. Pilot Study Instruments

A.1. Student Focus-Group Protocol

1. How do you usually learn about your mistakes and successes in your studies: through a mark, a teacher's comment, self-checking, or a classmate's assessment?
2. Which is more important to you: receiving a mark or receiving an explanation of what and how to improve? Why?
3. Give an example of a situation in which a teacher's comment helped you understand a topic better or improve your work.
4. Do you ever receive a mark without understanding exactly why it was awarded?
5. Do you use criteria or self-assessment before submitting your work? If so, how?
6. Does assessment in your school help you learn, or does it mainly indicate comparative performance? Explain your answer.

A.2. Author-Developed Pilot Observation Scale

Scale: 1—not observed; 2—isolated occurrence; 3—sometimes; 4—regularly; 5—consistently; 6—high quality; 7—consistent exemplary practice. The scores have no externally validated normative thresholds.

1. The teacher provides specific feedback on the content of the work rather than merely stating "correct/incorrect."
2. Feedback is aligned with the learning objective and success criterion.
3. The teacher uses a rubric, model, or other clear criteria.
4. Students use criteria for self-assessment.
5. Students participate in peer assessment under clear procedures.
6. The teacher modifies the next stage of instruction on the basis of the evidence of learning obtained.
7. Feedback includes guidance on planning, self-checking, or self-regulation.
8. Errors are discussed in an emotionally safe environment and used as a resource for learning.

A.3. Written Feedback Analysis Scale

- Task level: identifies correctness, an error, or a specific element of the current task.
- Process level: explains the strategy, method, reasoning, or cause of an error.
- Self-regulation level: guides planning, monitoring, self-checking, or the revision of subsequent work.
- Self level: contains general praise or an evaluation of the student without reference to a learning objective, criterion, or next action.

ISSN: 2181-3337 | SCIENTISTS.UZ