



PAPER

The SHARP assessment cycle: from assessment design and real-time student voice to student learning development

Zeenat Soobedar de Villeneuve 

King's College London, UK

ABSTRACT

Assessment in higher education often suffers from a critical disconnect where student feedback benefits future cohorts rather than current participants. This study presents SHARP (Strategic, Holistic, Adaptive, Reflective, Process), a systematic framework for embedding real-time student voice into assessment design to develop students' assessment literacy and engagement. SHARP was piloted across three in-class Mathematics tests with 246 students on a foundation-level UK university programme using a mixed-methods quasi-experimental design with post-assessment surveys from 108 students administered before grade release. Quantitative analysis revealed statistically significant improvements in clarity ratings, whilst fairness perceptions rose from 70.3% to 96.0%. A quasi-experimental comparison with historical controls demonstrated a 73% reduction in assessment volatility, reflecting more consistent student performance across assessments. Qualitative evidence illustrated students' developing capacity to evaluate and articulate assessment design. This pilot demonstrates SHARP's potential as a transferable framework for student-informed assessment practice, though replication is needed to establish broader generalisability.

KEYWORDS: assessment design, assessment literacy, co-creation, inclusive assessment, student voice.

ARTICLE HISTORY: Received 13 March 2026. Accepted 5 August 2026.

Introduction

Traditional assessment models position students as passive recipients rather than active contributors to assessment design, with feedback mechanisms typically retrospective, informing future iterations rather than enhancing current experiences (Williams & Ingle, 2025; Winstone & Boud, 2018). This temporal gap is particularly pronounced in summative contexts where students voice concerns about fairness, clarity, or accessibility after the assessment experience has concluded, often too late to benefit their own learning journey.

This disconnect has consequences not only for assessment quality but for student development. When students cannot engage meaningfully with assessment design, a specific form of assessment literacy goes undeveloped: the capacity to notice, articulate, and communicate what makes an assessment clear, fair, and accessible, not evaluating their own performance, but evaluating the design conditions that shape everyone's experience (Carless & Boud, 2018; Nieminen, 2022). Students may well recognise that instructions are unclear, formatting is inconsistent, or difficulty feels disproportionate, yet without structured mechanisms to channel these observations within the current cycle, feedback reaches educators too late to benefit the students experiencing the problem.

This study examines systematic approaches to real-time student voice integration in assessment design. Student voice here refers to views about assessment design and experience that are collected, processed, and implemented during ongoing academic cycles, enabling current participants to see their input directly influence their assessment experience. The research focuses specifically on assessment of student learning, that is, the summative evaluation of academic progress, rather than student evaluation of teaching effectiveness.

The SHARP assessment cycle (Strategic, Holistic, Adaptive, Reflective, Process) was developed to address this challenge by creating a structured framework for embedding real-time student voice into assessment design and delivery. Unlike approaches where student input is retrospective, SHARP enables adaptations within the same cohort, positioning students as active contributors to assessment design and, through that process, as developing learners. The framework builds on established principles of feedback literacy (Carless & Boud, 2018), student partnership (Cook-Sather & Felten, 2017), and inclusive pedagogy (Waring & Evans, 2024), whilst offering a novel mechanism for operationalising these principles in summative contexts.



Using a mixed-methods quasi-experimental design, SHARP was piloted in a foundation-level Mathematics module involving three in-class Mathematics tests with 246 students. Grade performance data from all 246 students were analysed alongside perception surveys from 108 invited participants. The study provided evidence for enhanced clarity (Cohen's $d = 0.51$), improved fairness perceptions (from 70.3% to 96.0%), and substantially reduced assessment volatility (73% reduction) whilst maintaining academic rigour.

Literature review

Student voice, partnership, and learning development

The concept of students as partners in higher education has evolved significantly, with recent scholarship demonstrating both its potential and implementation challenges. Nieminen et al. (2025) provide a comprehensive recent analysis of student partnership in assessment (SaPA), reviewing 46 studies published between 1994 and 2022. Their work proposes three core propositions: (1) outcomes of SaPA are uncertain but can be optimised, (2) SaPA is fundamentally a relational practice, and (3) SaPA promotes democratic values. Critically for this study, their analysis positions student partnership as a mechanism for developing assessment literacy in both students and staff, understood here as a learning outcome in its own right.

Recent practice has moved beyond small-group partnerships toward whole-cohort approaches. Williams and Ingle (2025) demonstrate this through their co-creation initiative exploring generative AI in assessment, showing how entire cohorts can meaningfully contribute to assessment design. Their work exemplifies authentic co-creation rather than superficial consultation, aligning with SHARP's commitment to genuine partnership. Similarly, recent work on 'Assessment by Engagement' (Folwell & Brennan, 2025) combines continuous summative assessment with embedded dialogic feedback and co-creation of assessment tasks.

However, systematic frameworks for embedding real-time student voice specifically in summative assessment design remain under-developed. When such student voice is incorporated, it typically occurs through module evaluations completed after grades are finalised, creating the temporal disconnect that SHARP seeks to address.



A critical concern in student-assessment partnership involves whether students, as subject novices, can contribute meaningfully to assessment design without compromising standards. This ‘they don’t know what they don’t know’ critique is addressed by positioning students as consultants on their assessment experience rather than pedagogical designers. Literature demonstrates students’ sophisticated evaluative capabilities (Panadero et al., 2024) and unique experiential expertise about assessment conditions (instruction clarity, accessibility, and timing) distinct from content knowledge (Cook-Sather & Felten, 2017; Nieminen et al., 2025). This maintains appropriate boundaries between student experiential expertise and educator pedagogical responsibility. This positioning is itself a pedagogical act: students who reflect systematically on assessment conditions develop a more sophisticated understanding of academic processes (Nieminen, 2022).

Formative assessment and real-time responsiveness

Formative assessment is defined as ongoing evaluation conducted during learning to provide feedback for improvement (Black & Wiliam, 1998). Recent research has provided stronger evidence for the effectiveness of continuous, real-time assessment approaches. Bulut et al. (2025), in their learning analytics study, found that the frequency of formative assessments is particularly important because continuous feedback allows students to make incremental adjustments to their learning strategies. Morris et al. (2021) conducted a systematic review specifically examining causal evidence from trials of feedback and formative assessment in higher education. Although they found the evidence base currently limited, their results suggest that low-stakes quizzing is particularly powerful, and there are benefits for forms of peer and tutor feedback that depend on implementation factors. SHARP extends the principle by positioning students as both recipients and generators of feedback about assessment processes, developing their capacity to contribute meaningfully to assessment design whilst the assessment experience remains actionable. This dual role, as feedback recipients and as reflective evaluators of assessment design, supports the development of assessment and feedback literacies that extend beyond any single module (Carless & Boud, 2018).

Inclusive assessment and assessment literacy: equity and access

Contemporary inclusive assessment research has moved beyond accommodation models toward design-for-all approaches. Nieminen (2022) argues for ‘Assessment for Inclusion’ that



rethinks inclusive assessment in higher education, advocating for students as partners in both designing assessment accommodations and inclusive assessment design more broadly, demonstrating how partnership approaches can foster assessment and feedback literacies for both students and teachers. Graham et al. (2018) show that accessibility challenges are often embedded in assessment design from the outset, requiring inclusive design from inception rather than retrospective adaptations. Waring and Evans (2024) have developed frameworks for supporting both staff and student assessment literacy (particularly in generative AI) through the Equity, Agency and Transparency (EAT) Framework, building on research showing that the systematic student reflection enhances understanding of assessment purposes and quality indicators (Panadero & Alonso-Tapia, 2013). SHARP embeds this through continuous attention to fairness, accessibility, and diverse student experiences.

Research gaps and SHARP's contribution

Despite growing theoretical clarity about appropriate student expertise domains, systematic frameworks for operationalising real-time student voice in assessment design remain limited (Nieminen et al., 2025). Key gaps include: temporal integration, moving beyond retrospective feedback towards continuous, to real-time assessment approaches; systematic processes for structured assessment co-design; and institutional integration, creating policy frameworks that support assessment partnership whilst maintaining quality assurance.

SHARP aims to address some of these gaps by providing a structured framework for real-time assessment improvement that maintains academic rigour whilst enabling continuous adaptation and fostering students' development as reflective learners.

The SHARP assessment cycle

Framework overview

SHARP consists of five interconnected phases that operate cyclically across assessment periods rather than linearly within individual assessments (Table 1; Figure C1, see Appendix C).



Table 1. The SHARP assessment cycle: phases, purposes and pilot examples.

Phase	Core purpose	Key actions	Pilot examples
Strategic	Evidence-informed design drawing on prior feedback and research.	Review cohort data; equity and AI considerations.	Transitioned from online to on-campus testing; bolded action verbs; 30% increase in white space.
Holistic	Contextualise assessment within wider student experience.	Analyse workload; address anxiety; coordinate across modules.	Coordinated scheduling to avoid clustering; reduced topic coverage without compromising outcomes.
Adaptive	Real-time responsiveness to student voice within the same cohort.	Collect feedback within one week; implement targeted changes.	Simplified language; increased spacing; optimised question sequencing based on Test 1 responses.
Reflective	Systematic capture of learning for future improvement.	Debrief students and staff; document insights; plan next cycle.	Feedback sophistication grew from general complaints to specific design recommendations.
Process	Embed partnership practices within institutional systems.	Develop protocols; communicate transparently; integrate with Quality Assurance.	Students informed about which suggestions were and were not implemented and why.

A critical conceptual distinction is that the framework applies formative principles to assessment design, not to learning evaluation: the three Mathematics tests remain summative evaluations, whilst the design process itself operates iteratively. The Strategic phase grounds decisions in accumulated evidence rather than convenience alone (Xia et al., 2024). The Holistic phase asks educators to step back and consider the full landscape of student experience surrounding each assessment, recognising that no assessment sits in isolation from the wider demands on students' time and attention (Advance HE, 2024). The Adaptive phase is where student voice translates directly into action: rapid systematic analysis of post-assessment feedback produces concrete changes for the next assessment within the same cohort, addressing a key gap in current partnership practice (Nieminen et al., 2025). The Reflective phase documents what the process has taught both staff and students, fostering the evaluative capacity the literature associates with developing assessment literacy (Carless & Boud, 2018). The Process phase embeds these practices within institutional structures,



making the approach sustainable beyond any individual educator's effort (Williams & Ingle, 2025).

SHARP in action: the Mathematics pilot

The pilot illustrated how each phase operates under real institutional constraints. Strategic decisions were shaped not only by prior student voice but by the practical shift from online to on-campus testing, requiring reconsidering layout, timing, and accessibility from the ground up. The Holistic phase required navigating conflicting student priorities: requests for clearer instructions sat alongside requests for maintained difficulty, and responses had to find solutions that served both without compromising academic standards. The Adaptive phase demonstrated that even small, targeted changes derived from student voice produced measurable shifts in student experience (Appendix A). For example, cramped layouts were addressed by increasing white space by 30% and separating multi-part questions into labelled sections: 'there was enough answer space on the test paper' (Student 27).

The Reflective phase revealed perhaps the most significant development: as students were invited to evaluate successive assessments, the quality and specificity of their feedback grew markedly, suggesting that structured reflection was itself building their capacity to engage with assessment design as informed participants rather than passive recipients, a process central to developing assessment literacy. The Process phase ensured these adaptations were documented and communicated transparently, maintaining academic integrity whilst demonstrating to students that their voice had genuine influence (Cook-Sather & Felten, 2017).

SHARP implementation insights

SHARP requires systematic but manageable time investment: approximately three to four hours per assessment cycle initially, reducing to one and a half to two hours with experience. This places SHARP within the reach of individual educators without requiring large-scale institutional infrastructure. Sustainability depends on embedding SHARP within existing institutional processes rather than treating it as an additional requirement (Appendix B).

The most significant staff development need is learning to distinguish between legitimate improvement opportunities and requests that would compromise standards or fairness. Transparent communication about which suggestions were and were not implemented is



itself a learning development intervention, modelling how academic decisions are made. Where course outlines are viewed as legally binding contracts restricting mid-stream modifications, the Adaptive phase can instead adjust surrounding support structures, whilst the Strategic and Reflective phases continue to operate as planned.

Methods

Methodological approach

This study employed design-based research (DBR) methodology, recognised as particularly appropriate for educational innovation research (McKenney & Reeves, 2019). DBR is characterised by iterative cycles of design, enactment, analysis, and redesign in authentic educational settings, making it particularly suited to studies that aim both to develop and to evaluate educational innovations, as each cycle generates evidence that informs subsequent refinements. This creates methodological coherence between approach and theoretical framework here, as SHARP itself operates through the same iterative logic.

The study employed a quasi-experimental before-after treatment design within the DBR framework, with Test 1 serving as baseline and Tests 2 and 3 representing SHARP implementation phases. Randomised controlled trials are often impractical in authentic assessment settings, whilst ensuring all students benefit from demonstrated improvements. Statistical analysis and categorisation of open-ended survey responses were employed within this design, consistent with DBR's commitment to triangulating evidence from multiple sources. Statistical methods were therefore not a separate strand but one source of evidence within the iterative DBR design, used to assess whether each implementation cycle produced measurable change. Statistical analysis used Welch's t-tests, one-way ANOVA, and effect size calculations (Cohen's d ; conventional guidelines: 0.2 small effect, 0.5 medium effect, 0.8 large effect) and η^2 for overall intervention impact.

Research context and participants

The study was conducted within a foundation-level Mathematics module at a UK university during the 2024–25 academic year. This module serves students preparing for undergraduate study, creating a diverse cohort in terms of mathematical backgrounds, cultural contexts, and educational experiences. Three in-class tests (December 2024, February 2025, March 2025)



provided multiple opportunities to implement and refine SHARP. The module enrolled 246 students across multiple tutorial groups; 236 completed all three assessments and are included in grade performance comparisons. For grade performance analysis, this cohort was compared with the previous year's cohort (n=227) using anonymised institutional data. A convenience sample of 108 students from selected tutorial groups participated in post-assessment surveys. To address evaluative biases, four tutorial groups (n=64) completed surveys before grade release and three tutorial groups (n=44) completed surveys after. Survey response rates from pre-grade-release groups were: Test 1: 40 participants (62.5%); Test 2: 28 participants (43.8%); and Test 3: 38 participants (59.4%). Analysis revealed no meaningful differences between pre- and post-grade release groups, so all perception data reported are from pre-grade-release groups only. The 2023–24 and 2024–25 cohorts showed similar demographics (Table 2), with changes occurring within a stable academic context of identical curriculum, teaching staff, and institutional policies.

Table 2. Student characteristics (2023–2025).

Characteristic	2023–24		2024–25	
	Mean	Std. dev.	Mean	Std. dev.
Age	18.9	1.05	19.2	1.08
IELTS Overall	6.7	0.69	6.8	0.70
Gender:	%		%	
Male	56.4		47.9	
Female	43.6		51.7	
Academic Pathway:	%		%	
Economics & Mathematics	58.7		50.4	
Mathematics & Business Mgmt	41.3		48.8	
Top 3 Nationalities:	%		%	
Chinese	45.8		37.6	
Saudi	17.8		13.6	
Turkish	6.7		12.4	
Total countries	33		44	
Sample size	227		246	



Historical data from 2020–2024 showed declining performance across all assessments, coinciding with the transition from online to on-campus delivery from 2023–24. Baseline qualitative evidence collected in 2023–24 identified specific assessment quality concerns: unclear Test 1 wording, disproportionately difficult Test 3 content, and notably easier Test 2 assessments that students acknowledged influenced their preparation strategies.

Mathematics was selected for its structured assessments, clear topic progression, and high student anxiety; the foundation context (including international students adjusting to UK expectations, first-generation university students, and those with educational gaps) provided insights into SHARP's applicability for students with varied academic preparation and cultural backgrounds.

Comparative analysis with historical control group

The quasi-experimental design compared the 2024–25 academic year (SHARP intervention) with the 2023–24 academic year (historical control). Key contextual factors were held constant: unchanged curriculum and learning objectives, identical teaching staff, and stable institutional policies. Sample characteristics demonstrate comparable cohort sizes (2024: $n=227$; 2025: $n=236$, reflecting 246 enrolled with some attrition across the three assessment points) and similar cohort demographics (Table 3), following identical assessment frameworks. Changes in test format between years were direct outcomes of SHARP implementation, representing evidence of effectiveness rather than confounding factors. The four-year baseline trend (2020–2024) provides stronger pre-intervention context than single-year historical controls.

Data collection and analysis

Brief post-assessment surveys were administered immediately after each test, before grade release, to capture authentic experiences without performance bias (Morris et al., 2021). Survey questions focused on assessment design elements (instruction clarity, layout, difficulty appropriateness, and time allocation), rather than content mastery. Quantitative measures covered fairness, clarity, difficulty, timing adequacy, and perceived responsiveness using 5-point Likert scales. Fairness refers specifically to procedural conditions that minimise barriers unrelated to learning objectives such as unclear instructions, inconsistent formatting, or unpredictable difficulty levels that may disproportionately affect certain student groups



(Messick, 1989; Xi, 2010). Open-ended survey questions encouraged students to reflect on experiences and suggest changes as assessment design partners. Given the limited scope of qualitative data collected through brief post-assessment surveys, this component provided supportive commentary rather than comprehensive qualitative analysis. Two researchers independently reviewed responses to identify recurring patterns and illustrative examples.

Limitations

The before-after design limits causal inferences, and unequal sample sizes across tests may have affected statistical power. The smaller Test 2 sample could reflect selection effects, as voluntary participation may have attracted more engaged students. The fairness question wording differed across tests: Test 1 asked whether it was fair; Tests 2 and 3 asked whether it was fairer than the previous test. This shift in question framing means the improvement from 70.3% to 96.0% should be interpreted cautiously, as part of the difference may reflect the changed question rather than changed assessment quality. Replication across diverse institutional contexts would strengthen confidence in generalisability.

Ethical considerations

This study received ethics approval from the university's Research Ethics Committee (LRS/DP-24/25-46257). Informed consent was obtained from all student participants prior to their involvement in the survey component of the study. All assessment scores were fully anonymised by the Student Experience Team before provision to researchers. Students were informed through standard institutional mechanisms that such data may be used for educational research and quality improvement. A convenience sample of 108 students participated in post-assessment surveys with explicit informed consent and they received £20 vouchers as compensation. The dual role of the research team as instructors and researchers was addressed through independent third-party collection and anonymisation.

Results

Assessment quality and student experience

SHARP implementation across three assessment cycles demonstrated patterns of change in key measures of assessment quality and student experience (Table 3), examined here in



terms of both statistical significance and what they suggest about students' developing engagement with assessment design.

Table 3. Descriptive statistics by test.

Variable	Test 1	Test 2	Test 3	Statistics	p-value
Clarity rating (1–5)	3.98 ± 1.00 (40)	4.41 ± 0.50 (22)	4.29 ± 0.76 (34)	F = 2.40	0.096
Fairness (%)	70.3% (26/37)	96.0% (24/25)	77.4% (24/31)	$\chi^2 = 6.21$	0.045
Timing adequacy (%)	75.0% (30/40)	96.2% (25/26)	100% (33/33)	$\chi^2 = 13.33$	0.001
Difficulty rating (1–5)	3.25 ± 1.03 (36)	3.20 ± 0.82 (25)	3.03 ± 1.07 (30)	F = 0.42	0.661
Student voice (1–5)	N/A	4.20 ± 0.82 (25)	4.13 ± 0.76 (31)	t = 0.34	0.739

Note: Responses in brackets from 4 pre-grade-release groups (n=64). Post-grade-release groups (n=44) showed equivalent patterns.

Clarity survey question: How would you rate the quality of questions on Test 1/2/3 in terms of clarity and relevance? 1 (Very poor)–5 (Excellent).

Fairness survey question: Do you feel Test 1 was fair to you, regardless of background and learning style? Then for Tests 2 and 3: Do you feel Test 2 (3) was fairer than Test 1(2) to you, regardless of background and learning style? Yes/Somewhat/No.

Timing adequacy question: Was the time given for Test 1 sufficient? Yes/No.

Student voice question: On a scale of 1 to 5, how well do you think student feedback from Test 1 was considered in designing Test 2? 1 (Not at all)–5 (Completely).

Difficulty question: How would you rate the difficulty of questions on Test 3 compared to Test 2? 1 (Very difficult)–5 (Very easy).

5% level of significance used for analysis.

Clarity and comprehensibility

Student ratings of assessment clarity improved across the test sequence (Table 3; Figure C2): from 3.98 in Test 1 to 4.41 in Test 2, maintained at 4.29 for Test 3. Adaptive phase modifications included increased spacing between questions, bolded instruction verbs, and simplified question language. The spread of ratings also narrowed considerably between Tests 1 and 2, suggesting that improvements were experienced more consistently across the cohort.



Qualitative data illustrated a parallel development in how students engaged with assessment design. Early student voice responses featured general criticisms such as, 'I cannot understand what some question[s] ask about' (Student 7, Test 1 survey), whilst later contributions included specific, actionable recommendations such as bolding 'the key instruction words' and providing 'more spaces between each small question' (Student 5, Test 2 survey). By Test 3, feedback had moved further still, towards strategic insights about the assessment as a whole, such as calls for consistent formatting and a more logical question sequence. This progression from basic complaint to specific suggestion to strategic insight reflects developing assessment literacy (Carless & Boud, 2018), discussed further in Section 6.1.

Fairness perceptions

Fairness perceptions improved substantially from 70.3% in Test 1 to 96.0% in Test 2, before dropping to 77.4% in Test 3. A statistically significant association was found between test condition and fairness perceptions (chi-square test, $p=0.045$). The Test 3 drop likely reflects the introduction of new question types rather than a deterioration in design quality, a point students themselves acknowledged in survey comments. Note that the fairness question changed between tests (absolute fairness for Test 1; comparative fairness for Tests 2 and 3), which should be borne in mind when interpreting the magnitude of change.

Student voice and involvement

Students rated perceived responsiveness highly: mean ratings of 4.20 (Test 2) and 4.13 (Test 3) on a five-point scale, with no significant difference between tests. Comments such as 'Yes, so you can know what works best for students' (Student 6, Test 3 survey) and 'I think this series of surveys has somewhat increased my sense of engagement' (Student 27, Test 3 survey) reflected positive perceptions of the feedback process, though this may reflect self-selection among survey respondents. Time adequacy increased progressively from 75.0% (Test 1) to 96.2% (Test 2), reaching 100% (Test 3), a statistically significant improvement (chi-square test, $p=0.001$). Difficulty ratings remained stable across all three tests ($F = 0.42$, $p = 0.661$ in Table 3), indicating that clarity and fairness improvements did not reduce academic challenge, a critical concern in assessment innovation (Baartman & Quinlan, 2024).

In practical terms, students were experiencing a fairer and clearer assessment process without being given an easier one.



Sequential intervention effects

Pairwise comparisons revealed that the primary improvement occurred between Test 1 and Test 2 (mean difference of 0.43 points, Cohen's $d = 0.51$, $p = 0.03$, a medium effect size). In practical terms this means students' experience of assessment clarity shifted by approximately half a standard deviation between the baseline and the first SHARP-informed test, a change that is both statistically detectable and educationally meaningful. The Test 2-to-Test 3 comparison yielded a negligible effect size, confirming that gains were maintained rather than reversed. This improvement-then-stability pattern is consistent with genuine initial impact followed by sustained effects (Table 4, Table 5; Figure C4). Historical data (Table 6) revealed consistent declining trends between 2020 and 2024.

Table 4. One-way ANOVA for assesement clarity ratings.

Source	SS	Df	MS	F	p-value	η^2
Between groups	3.27	2	1.64	2.40	0.096	0.049
Within groups	63.35	93	0.68			
Total	66.63	95				

Table 5. Pairwise comparisons (Welch's tests) for assessment clarity ratings.

Comparison	Mean diff.	T	p-value	Cohen's d	CI lower	CI upper
Test 1 vs Test 2	0.43	2.27	0.03	0.51	-0.02	1.03
Test 2 vs Test 3	-0.11	0.68	0.50	-0.17	-0.71	0.36

Table 6. Student grade performance (2020–2024).

	2020–21	2021–22	2022–23	2023–24
Test 1:				
Mean	86.4	80.3	65.8	69.5
Standard deviation	13.8	14.5	16.7	18.2
Median	90	83	66	71
Sample size	166	203	211	227
Test 2:				
Mean	91	88.4	80.6	82.8
Standard deviation	8.2	10.6	13	13.3
Median	93	90	83	85



Sample size	166	203	211	227
Test 3:				
Mean	90.6	79.8	77.8	72.2
Standard deviation	10	13.7	15.1	15.9
Median	93.5	83	82	73
Sample size	166	203	211	227

The 2023–24 control group demonstrated considerable volatility across their assessment journey: a moderate Test 1 (69.43%) was followed by an elevated Test 2 (82.63%) and then a declining Test 3 (72.14%). The 2024-25 cohort showed more consistent progression: Test 1 (66.07%), Test 2 (74.23%), and Test 3 (74.92%), differing from all four baseline years (Table 7).

Table 7. Descriptive statistics by year and test.

Test	2024 control (No SHARP)			2025 (SHARP)			Year diff.
	Sample	Mean	Std. dev.	Sample	Mean	Std. dev.	
1	230	69.43	22.45	246	66.07	23.40	-3.36
2	229	82.63	15.89	240	74.23	19.56	-8.40
3	229	72.14	17.82	236	74.92	21.78	2.78

The Assessment Volatility Index (sum of absolute effect sizes) was 1.54 for 2023–24 and 0.41 for 2024-25, representing a 73% reduction and the lowest volatility across the five-year series (Figure C6). Welch's t-tests comparing the two years found a significant difference only for Test 2, where the 2024-25 cohort scored lower than 2023-24 (Cohen's $d = -0.50$, $p < 0.001$, medium effect). Tests 1 and 3 showed no significant difference. This is consistent with the steadier trajectory reported above. The 2023-24 cohort's performance was considerably more volatile: it rose from Test 1 to Test 2 (a large effect) before falling from Test 2 to Test 3 (a medium-to-large effect), whereas the 2024–25 cohort's test-to-test changes stayed small to negligible.

The lower Test 2 performance in the SHARP year likely reflects the deliberate moderation of the Test 2 difficulty spike that students had identified as distorting their preparation strategies in 2023–24, rather than a negative effect of the intervention.



Causal attribution nonetheless requires caution. Alternative explanations include cohort-specific characteristics, regression to the mean and concurrent teaching or support changes. As this is the first SHARP implementation year, replication across multiple subsequent cohorts is essential before drawing conclusions about effectiveness.

Discussion

Theoretical contributions

SHARP's contribution is twofold: it offers a structured approach to improving assessment design quality, and it initiates learning development processes in students through that structure. Students who engaged systematically with assessment feedback developed a more sophisticated understanding of what constitutes clear, fair, and accessible assessment, a form of assessment literacy that extends beyond the module itself (Carless & Boud, 2018; Nieminen, 2022). The progression from general complaints to specific, design-oriented recommendations is consistent with developing evaluative judgement (Panadero & Broadbent, 2018). One student's comment is particularly instructive: 'I gave great feedback for Test 2, which made Test 3 similar to the second' (Student 28, Test 3 survey). This reflects not only satisfaction with the process but an emerging understanding of how assessment design works. This developmental trajectory suggests that single-cycle implementations may underestimate SHARP's potential, as feedback sophistication develops through repeated, structured engagement. Educators considering similar approaches should expect early feedback to be general and should not judge the overall process by its initial returns.

The substantial improvement from baseline to first implementation ($d = 0.51$, $p = 0.03$), followed by stable maintenance ($d = -0.17$, $p = 0.50$) is consistent with genuine enhancement rather than temporary effects. However, this pilot cannot definitively distinguish between genuine quality improvement and the placebo effects of consultation. Future research with independent assessment quality measures and control groups receiving equivalent modifications without student involvement would help establish this distinction. The mixed-methods findings demonstrate that whole-cohort partnership approaches are feasible in this context, addressing a gap identified in recent systematic reviews (Nieminen et al., 2025; Sun et al., 2022), though these findings require replication before broader theoretical contributions can be established.



Practical implications and applications

Three insights emerge for educators pursuing similar approaches. First, performance stability may matter as much as average achievement: The Assessment Volatility Index offers a concrete metric that complements traditional quality assurance measures. Second, the improvement-then-stability pattern represents the desired outcome of sustainable enhancement; practitioners should not interpret stability as evidence of intervention failure. Third, student voice sophistication develops through participation rather than being a precondition for it. Educators should design for iterative engagement and not judge the overall process by its initial returns. Resource requirements proved manageable (2–3 hours per assessment cycle) for individual educators, suggesting institutional infrastructure is not prerequisite for trial implementation.

Limitations

Institutional barriers may limit SHARP's implementation, particularly where course outlines are viewed as legally binding contracts, restricting mid-stream modifications (Nieminen et al., 2025). This study was conducted at one institution within a single Mathematics module, which limits generalisation. The quasi-experimental design cannot establish causation, and the foundation-level setting may have produced stronger results than would occur in other contexts. Response rates of 26 to 37% mean findings may not fully represent all students' experiences. The research team designed and implemented SHARP, potentially introducing enthusiasm that is difficult to replicate. The absence of systematic baseline perception data from previous years makes it difficult to isolate SHARP's contribution from other factors.

Future directions

Future research needs to examine whether benefits replicate across different disciplines and institutional contexts, and what mechanisms link student voice integration to observed changes. The 2025–26 academic year will be particularly informative: similar patterns would strengthen confidence in SHARP's contribution; reversion to pre-SHARP characteristics would suggest 2024–25 represented cohort-specific factors. Honest reporting regardless of outcome—including null or negative results—would advance understanding as meaningfully as successful replications, and help to address the field's publication bias towards positive results.



Conclusion

This pilot study demonstrates that systematic student voice integration during ongoing assessment cycles is feasible, resource-efficient, and corresponds with observable changes in both perception and performance patterns. Importantly, it suggests that the process of engaging students in assessment design is itself a learning development intervention: students who reflect systematically on assessment conditions develop more sophisticated evaluative capacities over time. This is perhaps SHARP's most significant contribution to learning development practice—not that it improved assessment scores, but that it engaged students as active participants in understanding and shaping the academic processes that affect them. Whether this initial evidence translates to sustained benefits across diverse contexts remains an empirical question requiring collaborative investigation across institutions and educational levels.

Acknowledgements

The authors would like to thank the institutional foundation programme for approving this project and Mr Stephen McShane, the Mathematics module leader, for their input and support throughout the project. Special thanks to the student experience team, especially Ms Chloe Humphreys, for support with data collection and anonymising the dataset, and to business support staff for arranging student incentives. Particular appreciation goes to Ms Sophie Bennett for excellent research assistant support in performing thematic analysis and creating summary reports within tight deadlines. Our gratitude extends to the anonymous reviewers whose detailed and insightful comments greatly improved the quality of this work. Additional thanks to Dr Maria Chiara Iannino for insightful review comments, meticulous proofreading and editorial support.

Disclosure statement

The author used the following generative AI tool in the preparation of this manuscript: Claude Sonnet (Anthropic). The tasks performed were limited to supporting the flow of ideas, structural organisation, and grammatical review. Additionally, Claude Sonnet was consulted to suggest potentially relevant sources, all of which were independently verified and assessed by the author before inclusion. All intellectual content, including the literature review, research



design, data collection, analysis, interpretation, and conclusions, remains solely the work of the human author.

Funding

This project was funded through King's Academy under the College Teaching Fund in 2024.

The author reports there are no competing interests to declare.

References

- Advance HE. (2024). *Student partnerships in assessment*. <https://www.advance-he.ac.uk/membership/all-member-benefit-projects/student-partnerships-in-assessment>
- Baartman, L. K. J., & Quinlan, K. M. (2024). Assessment and feedback in higher education reimagined: Using programmatic assessment to transform higher education. *Perspectives: Policy and Practice in Higher Education*, 28(2), 57–67. <https://doi.org/10.1080/13603108.2023.2283118>
- 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>
- Bulut, O., Gorgun, G., & Yildirim-Erbasli, S. N. (2025). The impact of frequency and stakes of formative assessment on student achievement in higher education: A learning analytics study. *Journal of Computer Assisted Learning*, 41(1), Article e13087. <https://doi.org/10.1111/jcal.13087>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Cook-Sather, A., & Felten, P. (2017). Where student engagement meets faculty development: how student-faculty pedagogical partnership fosters a sense of belonging. *Student Engagement in Higher Education Journal*, 1(2), 3–11. <https://sehej.raise-network.com/raise/article/view/cook>
- Folwell, E. J., & Brennan, J. D. (2025). Assessment by engagement: building confidence and autonomy in the first year. *Assessment & Evaluation in Higher Education*, 50(6), 819–932. <https://doi.org/10.1080/02602938.2025.2483268>



- Graham, L. J., Tancredi, H., Willis, J., & McGraw, K. (2018). Designing out barriers to student access and participation in secondary school assessment. *The Australian Educational Researcher*, 45(1), 103–124. <https://doi.org/10.1007/s13384-018-0266-y>
- McKenney, S., & Reeves, T. C. (2019). *Conducting educational design research* (2nd ed.). Routledge.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education.
- Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, 9(3), Article e3292. <https://doi.org/10.1002/rev3.3292>
- Nieminen, J. H. (2022). Assessment for Inclusion: rethinking inclusive assessment in higher education. *Teaching in Higher Education*, 29(4), 841–859. <https://doi.org/10.1080/13562517.2021.2021395>
- Nieminen, J. H., Dollinger, M., & Zou, T. X. P. (2025). Student partnership in assessment: What works, for whom, why, and under what circumstances? *Review of Education*, 13(1), Article e70059. <https://doi.org/10.1002/rev3.70059>
- Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology*, 11(2), 551–576. <https://doi.org/10.14204/ejrep.30.12200>
- Panadero, E., & Broadbent, J. (2018). Developing evaluative judgement: A self-regulated learning perspective. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing evaluative judgement* (pp. 89–99). Routledge.
- Panadero, E., Fernández, J., Pinedo, L., Sánchez, I., & García-Pérez, D. (2024). A self-feedback model (SEFEMO): Secondary and higher education students' self-assessment profiles. *Assessment in Education: Principles, Policy & Practice*, 31(3–4), 221–253. <https://doi.org/10.1080/0969594X.2024.2367027>
- Sun, S. (Alice), Gao, X. (Andy), Rahmani, B. D., Bose, P., & Davison, C. (2022). Student voice in assessment and feedback (2011–2022): A systematic review. *Assessment & Evaluation in Higher Education*, 48(7), 1009–1024. <https://doi.org/10.1080/02602938.2022.2156478>
- Waring, M., & Evans, C. (2024). Facilitating students' development of assessment and feedback skills through critical engagement with generative artificial intelligence. In C. Evans, & M.



Waring (Eds.), *Research Handbook on Innovations in Assessment and Feedback in Higher Education* (pp. 330–354). Edward Elgar Publishing.

<https://doi.org/10.4337/9781800881600.00027>

Williams, A., & Ingle, E. (2025). Assessment design through co-creation: Student-staff partnership in evaluating the impact of artificial intelligence. *International Journal for Students as Partners*, 9(1), 214–226. <https://doi.org/10.15173/ijasp.v9i1.5802>

Winstone, N., & Boud, D. (2018). Exploring cultures of feedback practice: the adoption of learning-focused feedback practices in the UK and Australia. *Higher Education Research & Development*, 38(2), 411–425. <https://doi.org/10.1080/07294360.2018.1532985>

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, 27(2), 147–170. <https://doi.org/10.1177/0265532209349465>

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21(1), Article 40. <https://doi.org/10.1186/s41239-024-00468-z>

Author details

Zeenat Soobedar de Villeneuve is a Lecturer in Maths Education at the King's Foundation, King's College London. With a PhD in Economics from Queen Mary University of London and over 20 years of teaching experience, she developed the SHARP assessment cycle in response to the shift to online learning during the Covid-19 pandemic and the rise of AI. Her work is driven by the view that staff and students working together, within boundaries that maintain academic rigour, leads to better learning experiences and a stronger sense of belonging.

ORCID

Zeenat Soobedar de Villeneuve  <https://orcid.org/0009-0003-4924-230X>



Appendix A: Full list of adaptations following student responses

Issue identified	Adaptation made	Student response evidence
Layout cramped (Test 1).	Increased white space by 30%.	'There was enough answer space on the test paper' (Student 27, Test 3 survey).
Instructions unclear (Test 1).	Bolded action verbs, simplified language.	'Yes, the teacher gave us clear instructions' (Student 31, Test 3 survey).
Page flipping confusion (Test 1).	Repeated diagrams on multiple pages.	'We don't have to keep flipping' (Student 18, Test 2 survey).
Technical terms (Test 2).	Glossary provided, simplified contexts.	'It was all clear. At least for me' (Student 5, Test 3 survey).
Question sequencing (Test 2).	Logical progression from easy to complex.	'Test 3's question types were not so different from test 2' (Student 17, Test 3 survey).

Appendix B: SHARP implementation framework

Phase	Key actions	Guiding questions	Evidence sources	Timing
Strategic	Review prior feedback.	What do previous students tell us?	Previous cohort evaluations, literature review, institutional priorities.	Pre-semester.
	Align with learning outcomes.	How do current contexts influence design?		
	Consider emerging challenges.	What evidence should inform decisions?		
Holistic	Assess workload context.	How does this assessment fit student experience?	Student workload data, wellbeing indicators, programme mapping.	Pre-semester + week 1.
	Consider emotional impacts.	What competing demands exist?		
	Align with programme goals.	What support is needed?		
Adaptive	Collect rapid feedback.	What is working well/poorly?	Post assessment surveys, focus groups, performance data.	Week when assignment brief is released.
	Analyse systematically.	Which changes maintain standards?		Week after assignment has been submitted.
	Implement appropriate changes.	How can we respond quickly?		Repeat for assignment 2.
Reflective	Document learning.	What have we learned?	Implementation logs, stakeholder reflections, outcome analysis.	Between brief and assignment.
	Analyse patterns.	What should be sustained/changed?		
	Plan improvements.	How can insights be shared?		
Process	Embed in system.	How can this become routine?	Policy alignment, training evaluations, adoption metrics.	End of summer + pre-next year.
	Train stakeholders.	What support is needed?		
	Plan scaling.	How can practices spread?		

Note: Two assessments within a semester—can be easily adapted for two assessments over the course of one academic year.



Appendix C

Figure C1. The SHARP assessment cycle.

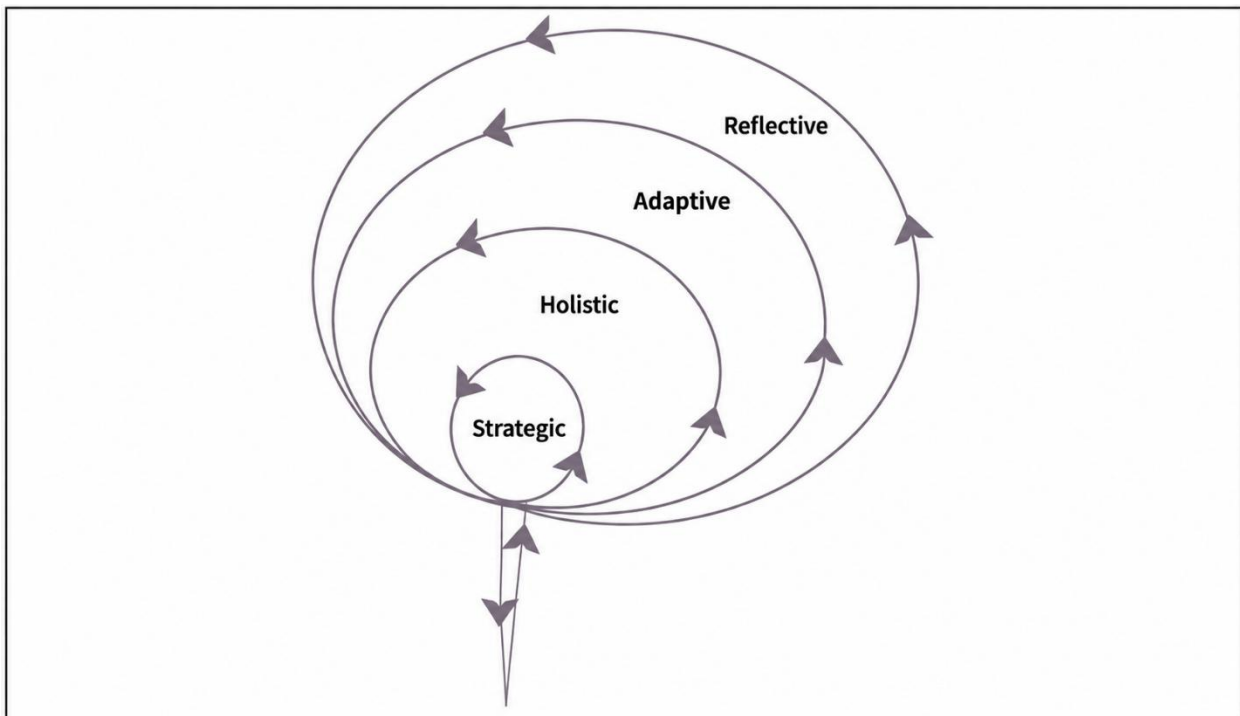


Figure C2. SHARP intervention: assessment quality and student experience measures.

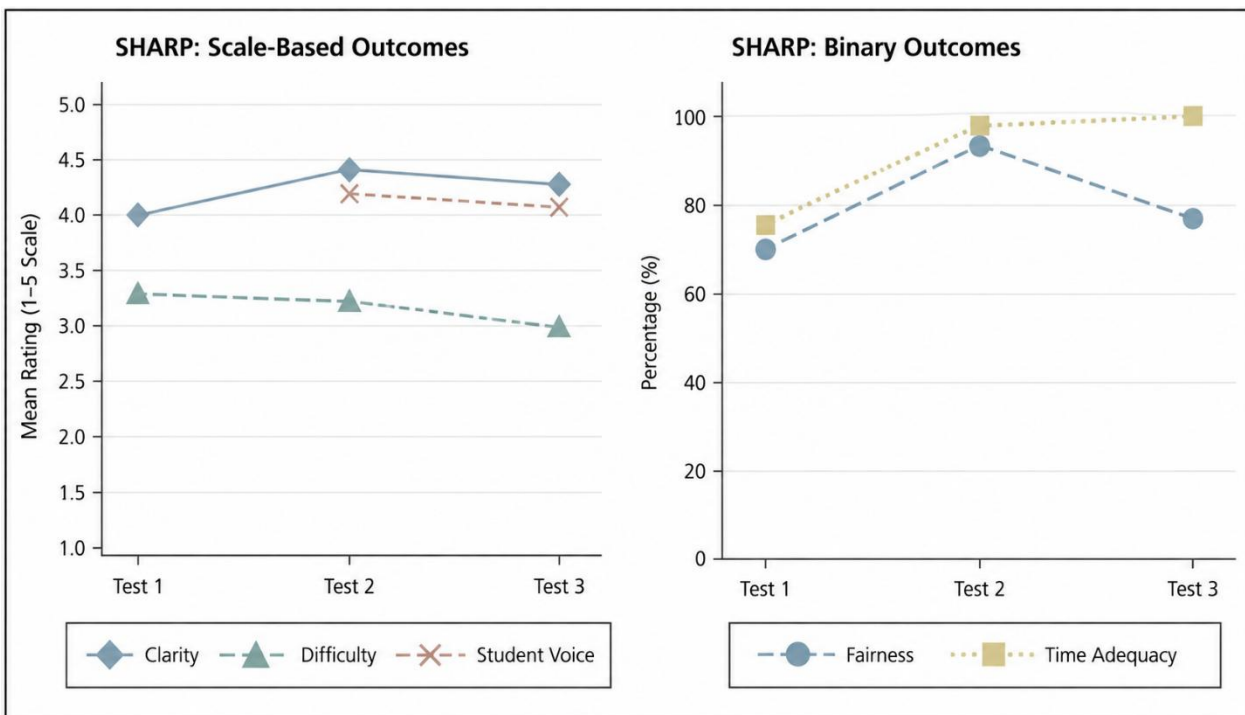


Figure C3. SHARP in action.

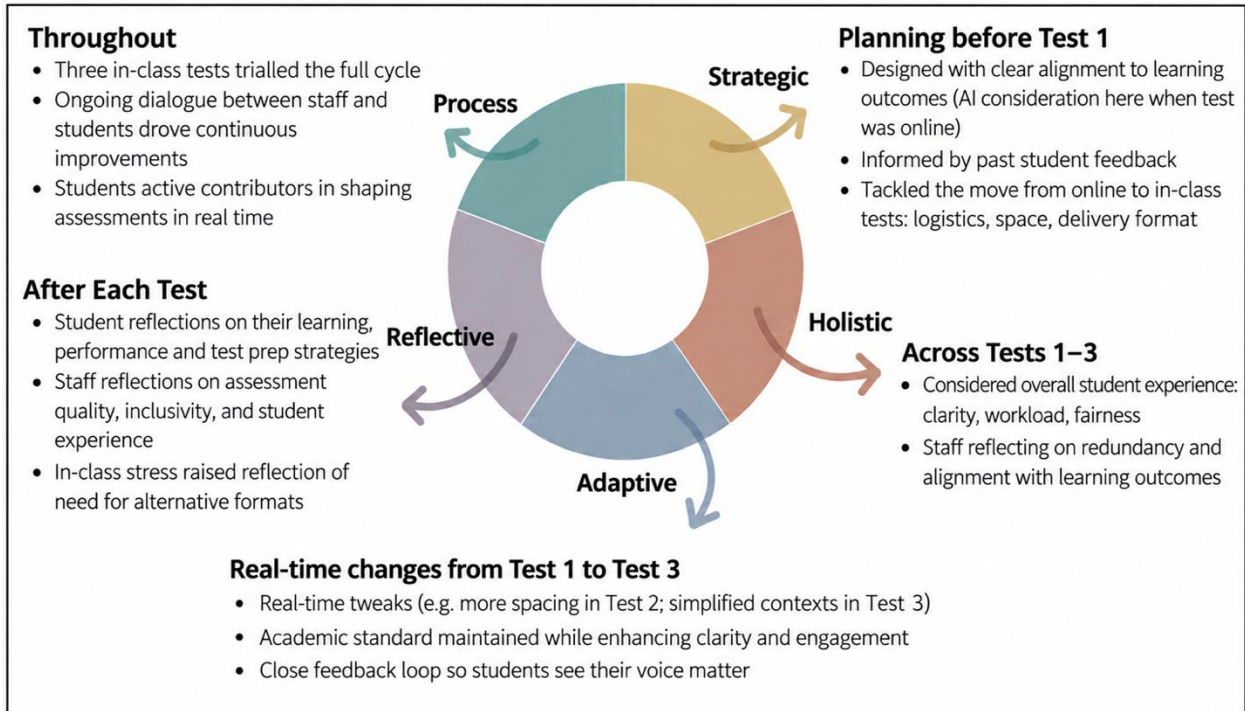


Figure C4. Sequential SHARP implementation effect sizes (95% confidence intervals).

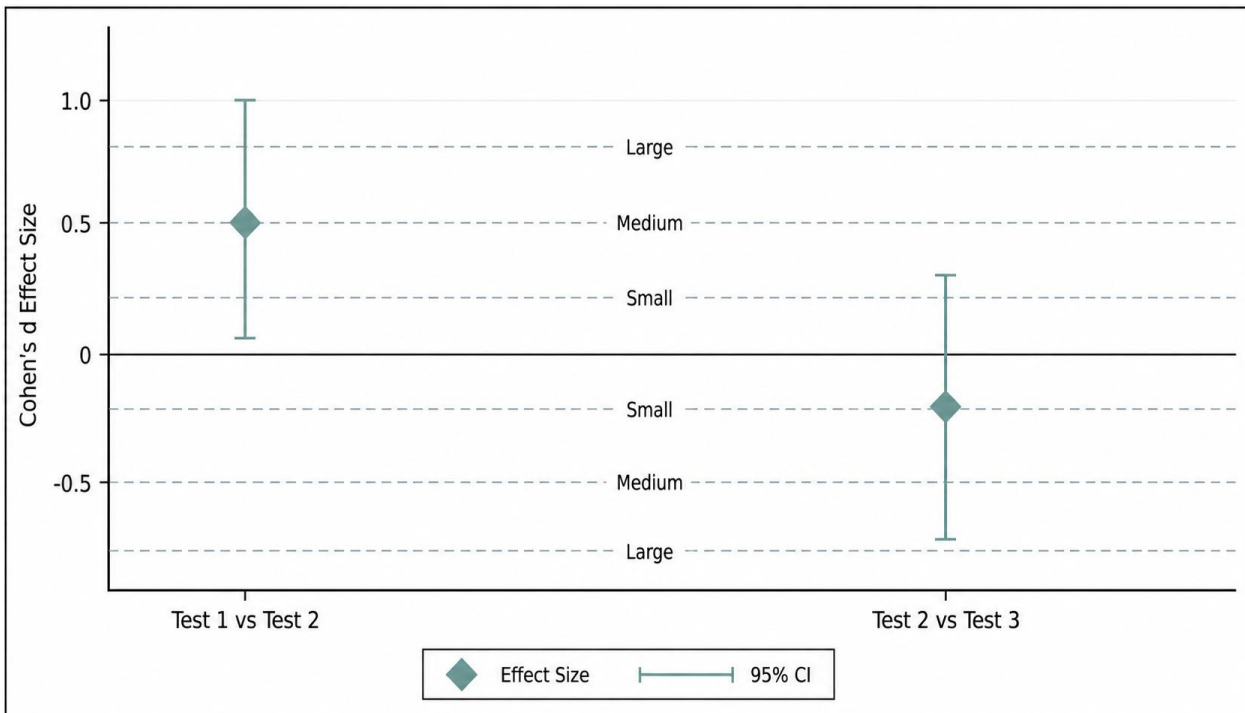


Figure C5. Score distributions by year and test.

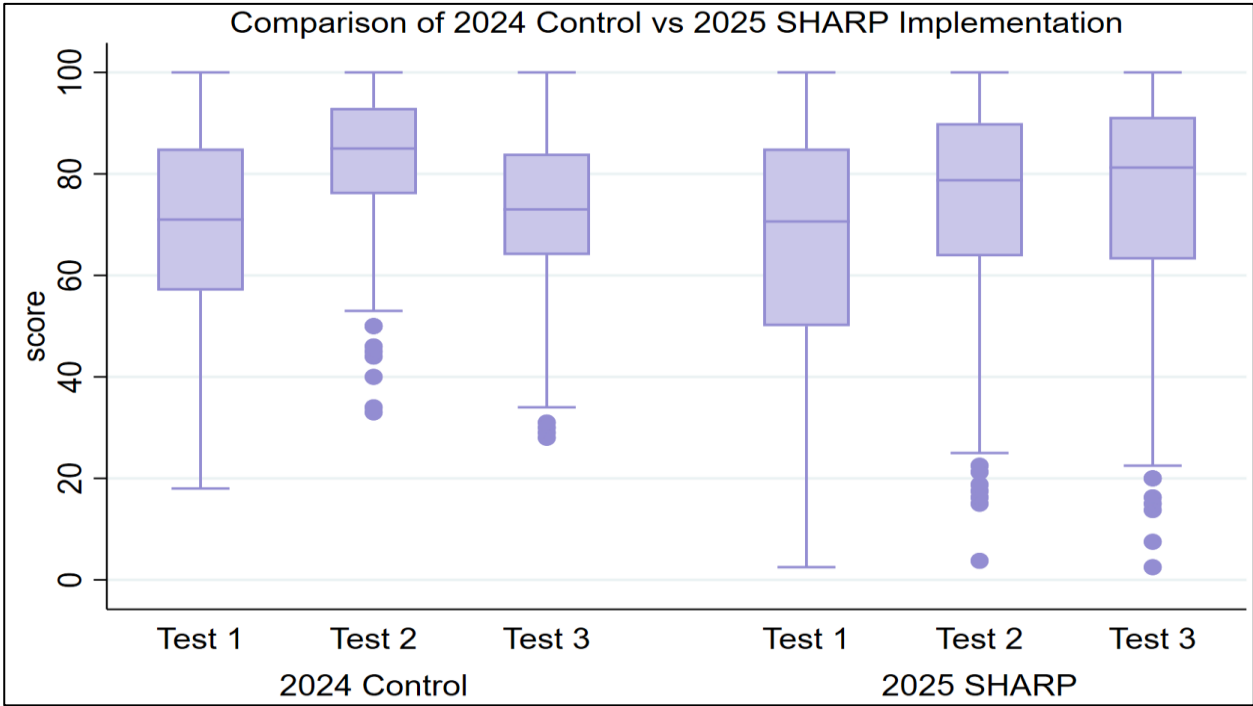


Figure C6. Assessment volatility index comparison.

