INTERNATIONAL JOURNAL OF TEACHER EDUCATION RESEARCH STUDIES (IJTERS)
(Open Access, Double-Blind Peer Reviewed Journal)
ISSN Online: 3049-1614

Formative Versus Summative Assessment Practices and Their Effects on Student Achievement, Self-Efficacy, and Motivation in Secondary Schools

Anitha N V

Retired HM, G. L P S Karikkinmedu, Kerala, India.

Article information

Received: 8th March 2026
Received in revised form: 10th April 2026
Accepted: 12th May 2026
Available online: 18th June 2026

Volume: 3
Issue: 2
DOI: https://doi.org/10.63090/IJTERS/3049.1614.0036

Abstract

This quantitative study examined the differential effects of formative and summative assessment practices on secondary school students' academic achievement, academic self-efficacy, test anxiety, student engagement, and intrinsic motivation. Using a quasi-experimental, nonequivalent comparison group design, data were collected from 364 students (Grade 9-12) enrolled in eight secondary schools across two school districts in a mid-sized metropolitan region. Students were assigned to one of two conditions: classrooms employing predominantly formative assessment strategies (n = 186) or classrooms employing predominantly summative assessment strategies (n = 178). A one-semester intervention period was followed by collection of outcome data through standardized achievement tests, validated self-report scales, and administrative records. Multivariate analysis of variance (MANOVA) revealed statistically significant differences across all five outcome variables, with students in formative assessment classrooms demonstrating higher academic achievement (M = 74.82 vs. 68.37, d = 0.65), higher self-efficacy (d = 0.62), lower test anxiety (d = 0.93), greater engagement (d = 0.62), and higher intrinsic motivation (d = 0.71) compared to their peers in summative assessment classrooms. Moderation analyses indicated that the benefits of formative assessment were particularly pronounced for students from low-income backgrounds. Findings have significant implications for assessment policy, teacher professional development, and instructional design in secondary education contexts.

Keywords: - Formative Assessment, Summative Assessment, Academic Achievement, Self-Efficacy, Test Anxiety, Student Engagement, Intrinsic Motivation, Quasi-Experimental

I. INTRODUCTION

Assessment is widely recognized as one of the most powerful levers available to educators for shaping student learning, motivation, and achievement. The question of how assessment is designed and deployed, and for what purposes, lies at the heart of enduring debates in educational policy and practice. Two broad paradigms have dominated secondary school assessment discourse: formative assessment, which is embedded in ongoing instruction and designed to provide feedback that guides learning; and summative assessment, which is administered at the conclusion of instructional units to evaluate and certify student achievement (Black & Wiliam, 1998; Harlen, 2005).

The distinction between these two assessment paradigms is not merely technical but reflects fundamentally different assumptions about the purposes of assessment, the nature of learning, and the role of feedback in educational processes. Proponents of formative assessment argue that its emphasis on ongoing feedback, student self-regulation, and instructional adaptation produces superior learning outcomes and more positive motivational profiles compared to predominantly summative approaches (Black & Wiliam, 1998; Hattie & Timperley, 2007). Critics of summative-heavy assessment cultures contend that high-stakes testing environments elevate test anxiety, undermine intrinsic motivation, and create conditions more conducive to performance-avoidance goal orientations than to genuine intellectual engagement (Harlen & Deakin Crick, 2003; Ryan & Brown, 2005).

Despite a robust theoretical literature and a substantial body of small-scale qualitative and mixed-methods research supporting the benefits of formative assessment, large-scale quantitative studies comparing the two paradigms in secondary school contexts remain relatively scarce. The majority of existing quantitative research has been conducted at the primary

school level or in higher education, leaving a significant empirical gap at the secondary level, where assessment pressures are acutely felt and where assessment practices have the most direct bearing on students' academic trajectories and postsecondary opportunities (Kingston & Nash, 2011; Wiliam, 2011).

The present study addresses this gap by examining the differential effects of formative and summative assessment practices on five key outcome variables among secondary school students: academic achievement, academic self-efficacy, test anxiety, student engagement, and intrinsic motivation. The study employs a quasi-experimental, nonequivalent comparison group design across eight secondary schools in two school districts, with a one-semester intervention period.

1.1. Research Questions and Hypotheses

The study was guided by the following research questions:

- RQ1: Are there statistically significant differences in academic achievement between secondary school students in formative assessment classrooms and those in summative assessment classrooms?
- RQ2: Are there statistically significant differences in academic self-efficacy, test anxiety, student engagement, and intrinsic motivation between the two groups?
- RQ3: Does socioeconomic status moderate the relationship between assessment condition and student outcomes?

Based on the extant literature, it was hypothesized that students in formative assessment classrooms would demonstrate significantly higher academic achievement, self-efficacy, engagement, and intrinsic motivation, and significantly lower test anxiety, compared to students in summative assessment classrooms (H1). It was further hypothesized that the benefits of formative assessment would be more pronounced for students from low-income backgrounds (H2), consistent with findings suggesting that formative feedback disproportionately benefits students with fewer external academic support resources (Wiliam, 2011).

II. LITERATURE REVIEW

2.1. Formative Assessment: Theoretical Foundations and Evidence

Formative assessment, also described as assessment for learning, encompasses a range of practices through which evidence of student learning is elicited, interpreted, and used by teachers and students to make instructional and learning decisions (Black & Wiliam, 1998; Wiliam, 2011). The seminal meta-analysis by Black and Wiliam (1998), reviewing over 250 studies, reported effect sizes of between 0.4 and 0.7 standard deviations for formative assessment interventions on student achievement, among the largest effects documented for any educational intervention at the time. These findings catalyzed a global policy movement toward embedding formative assessment practices within classroom instruction.

Hattie and Timperley's (2007) influential model of feedback provides a theoretical framework for understanding how formative assessment produces its effects. The model distinguishes four levels at which feedback operates: task, process, self-regulation, and self. Feedback directed at the task and process levels, which characterizes effective formative assessment, is associated with enhanced learning, while feedback directed at the self level (praise or criticism of the learner as a person) is typically ineffective or counterproductive. Consistent with Self-Determination Theory (Deci & Ryan, 2000), feedback that supports students' sense of competence and autonomy, both hallmarks of well-implemented formative assessment, is predicted to enhance intrinsic motivation and engagement.

2.2. Summative Assessment and Its Consequences

Summative assessment, or assessment of learning, is designed to evaluate student achievement at defined endpoints and to certify, grade, or rank performance. In secondary school contexts, summative assessments typically include end-of-unit examinations, standardized tests, and high-stakes national or regional assessments that determine progression and postsecondary access (Harlen, 2005). While summative assessments serve essential accountability and certification functions, research indicates that environments dominated by high-stakes summative assessment are associated with elevated test anxiety (Putwain, 2007), performance-avoidance goal orientations (Ryan & Brown, 2005), surface rather than deep learning approaches (Birenbaum, 2003), and reduced intrinsic motivation (Harlen & Deakin Crick, 2003).

Putwain (2007) conducted a systematic review of test anxiety in secondary school students and found that prevalence rates ranged from 10% to 40% across studies, with substantial negative effects on examination performance, well-being, and long-term academic self-concept. The mechanisms through which high-stakes summative assessment elevates test anxiety include the attribution of significant consequences to performance, the infrequency of assessment opportunities (limiting recovery from poor performance), and the competitive ranking of students, all of which are antithetical to the psychologically safe environments that support intrinsic motivation and deep learning (Ryan & Brown, 2005).

2.3. Assessment and Equity

A consistent finding in the assessment literature is that the consequences of assessment practices are not uniformly distributed across student populations. Students from economically disadvantaged backgrounds are disproportionately harmed by high-stakes summative assessment cultures, both because they are less likely to have access to private tutoring and test preparation resources and because the consequences of underperformance, including exclusion from academic tracks and postsecondary pathways, are more severe for students with limited alternative routes to advancement (Sadler, 2010; Stobart, 2008). Wiliam (2011) argued that formative assessment is inherently more equitable than summative assessment because it provides all students with actionable information about how to improve, rather than simply certifying current performance levels.

III. METHOD

3.1. Research Design

The study employed a quasi-experimental, nonequivalent comparison group design (Campbell & Stanley, 1963), in which intact classrooms were assigned to formative or summative assessment conditions. Random assignment of individual students was not feasible given the intact nature of school class groupings; however, schools were matched on key demographic and academic performance indicators prior to condition assignment to maximize comparability between groups. Institutional ethics approval was obtained from the university's Human Research Ethics Committee, and informed consent was secured from school principals, teachers, parents, and students prior to data collection.

3.2. Participants

Participants were 364 secondary school students (Grade 9-12) enrolled in eight schools across two school districts in a mid-sized metropolitan region. Schools were selected through purposive sampling based on willingness to participate and comparability on district-reported demographic and achievement indicators. Students in four schools were assigned to the formative assessment condition (n = 186; 52.7% female, 44.1% male, 3.2% non-binary or other), and students in the remaining four schools formed the summative assessment comparison group (n = 178; 52.8% female, 43.8% male, 3.4% non-binary or other). Table 1 presents full demographic characteristics of the sample.

Table 1. Participant Demographic Characteristics by Assessment Condition

Variable	Formative Group (n = 186)	Summative Group (n = 178)	Total (N = 364)
Grade Level			
Grade 9	47 (25.3%)	44 (24.7%)	91 (25.0%)
Grade 10	48 (25.8%)	46 (25.8%)	94 (25.8%)
Grade 11	46 (24.7%)	45 (25.3%)	91 (25.0%)
Grade 12	45 (24.2%)	43 (24.2%)	88 (24.2%)
Gender			
Female	98 (52.7%)	94 (52.8%)	192 (52.7%)
Male	82 (44.1%)	78 (43.8%)	160 (44.0%)
Non-binary/Other	6 (3.2%)	6 (3.4%)	12 (3.3%)
Socioeconomic Status			
Low income	58 (31.2%)	55 (30.9%)	113 (31.0%)
Middle income	89 (47.8%)	86 (48.3%)	175 (48.1%)
High income	39 (21.0%)	37 (20.8%)	76 (20.9%)

Note. SES categories are based on eligibility for free or reduced-price school lunch programs (low income) and parental self-report (middle and high income).

3.3. Intervention

The intervention period spanned one full academic semester (18 weeks). Teachers in the formative assessment condition received four full days of professional development training in formative assessment strategies prior to the intervention, followed by fortnightly coaching sessions throughout the semester. Formative assessment practices implemented included: daily exit tickets and learning checks; peer and self-assessment activities using co-constructed success criteria; frequent low-stakes quizzes with individualized written feedback; and student-led conferences on learning progress. Teachers in the summative assessment condition continued their existing assessment practices, which consisted primarily of end-of-unit examinations, mid-semester tests, and a final semester examination, with minimal formative feedback between assessments. To ensure fidelity, all formative assessment classrooms were observed on three occasions by trained observers using a validated observation checklist (Leahy et al., 2005).

3.4. Measures

Academic achievement was measured using a standardized end-of-semester achievement test developed collaboratively by district curriculum officers and independently validated against national curriculum standards. Scores ranged from 0 to 100. Academic self-efficacy was assessed using the Academic Self-Efficacy Scale (Bandura, 1997), a 10-item Likert-format instrument (1 = not at all confident, 5 = completely confident) with well-established reliability (alpha = .87 to .92 across previous studies). Test anxiety was measured using the Revised Test Anxiety Scale (Benson & El-Zahhar, 1994), a 20-item scale assessing cognitive, affective, and behavioral dimensions of test anxiety (alpha = .90 to .95). Student engagement was assessed using the Student Engagement Instrument (Appleton et al., 2006), capturing cognitive and affective engagement dimensions (alpha = .88). Intrinsic motivation was measured using the Intrinsic Motivation Inventory academic subscale (Ryan & Deci, 2000), assessing interest, enjoyment, and perceived choice in academic activities (alpha = .86).

3.5. Data Analysis

Data were analyzed using IBM SPSS Statistics version 28. Preliminary analyses included examination of distributional assumptions, missing data patterns (less than 2% of data were missing, addressed through multiple imputation), and equivalence of groups on pre-intervention covariates using independent samples t-tests and chi-square tests. The primary analysis was a one-way Multivariate Analysis of Variance (MANOVA) with assessment condition as the independent variable and the five outcome measures as dependent variables. Wilks' lambda was used as the multivariate test statistic. Statistically significant multivariate effects were followed by univariate ANOVAs with Bonferroni correction for multiple comparisons.

Effect sizes were estimated using partial eta-squared and Cohen's d. Moderation analyses examining the role of socioeconomic status were conducted using hierarchical multiple regression with assessment condition by SES interaction terms.

IV. RESULTS

4.1. Preliminary Analyses

Pre-intervention equivalence checks confirmed that the formative and summative assessment groups did not differ significantly on baseline academic performance records ($t(362) = 1.14$, $p = .255$), SES distribution (chi-squared(2) = 0.04, $p = .980$), or gender composition (chi-squared(2) = 0.02, $p = .990$). Box's M test indicated that the assumption of equality of covariance matrices was satisfied ($M = 24.17$, $p = .163$). Mardia's tests confirmed multivariate normality was not severely violated. No significant outliers were identified using Mahalanobis distance.

4.2. Multivariate and Univariate Results

The one-way MANOVA revealed a statistically significant multivariate effect of assessment condition on the combined set of outcome variables, Wilks' lambda = .74, $F(5, 358) = 25.43$, $p < .001$, partial eta-squared = .26, indicating that assessment condition accounted for approximately 26% of the variance in the multivariate outcome space. Table 2 presents descriptive statistics and between-group effect sizes for each outcome variable, and Table 3 presents univariate ANOVA results.

Table 2. Descriptive Statistics and Effect Sizes by Assessment Condition

Variable	Formative Group M (SD)	Summative Group M (SD)	Cohen's d
Academic Achievement	74.82 (9.14)	68.37 (10.63)	0.65
Academic Self-Efficacy	3.91 (0.72)	3.44 (0.81)	0.62
Test Anxiety	2.31 (0.84)	3.12 (0.91)	0.93
Student Engagement	4.02 (0.68)	3.58 (0.74)	0.62
Intrinsic Motivation	3.87 (0.77)	3.29 (0.88)	0.71

Note. M = mean; SD = standard deviation. Academic achievement scores are percentage-based (0-100). All other variables are on a 1-5 Likert scale. Cohen's d values represent the standardized mean difference (formative minus summative, with direction reversed for test anxiety).

Table 3. Univariate ANOVA Results for All Outcome Variables

Dependent Variable	F (1, 362)	p	eta-p2	95% CI	Effect Size
Academic Achievement	41.73	< .001	.103	[0.06, 0.15]	Medium
Academic Self-Efficacy	37.84	< .001	.095	[0.05, 0.14]	Medium
Test Anxiety	68.29	< .001	.159	[0.10, 0.22]	Large
Student Engagement	35.11	< .001	.088	[0.05, 0.13]	Medium
Intrinsic Motivation	47.56	< .001	.116	[0.07, 0.16]	Medium

Note. eta-p2 = partial eta-squared. Effect size benchmarks follow Cohen (1988): small = .01, medium = .06, large = .14. All p values are Bonferroni-corrected. CI = confidence interval for partial eta-squared.

Follow-up univariate ANOVAs confirmed statistically significant differences between conditions on all five outcome variables after Bonferroni correction. The largest effect was observed for test anxiety (partial eta-squared = .159, large effect), indicating that students in formative assessment classrooms reported substantially lower test anxiety than their summative group peers. Medium effects were observed for academic achievement, self-efficacy, engagement, and intrinsic motivation, all favoring the formative assessment condition.

4.3. Moderation by Socioeconomic Status

Hierarchical regression analyses examined whether SES moderated the relationship between assessment condition and each outcome variable. For academic achievement, the Assessment Condition x SES interaction was statistically significant (beta = .18, $t(360) = 3.42$, $p < .001$, R-squared change = .03), indicating that the achievement advantage associated with formative assessment was significantly more pronounced for low-income students ($d = 0.81$) than for middle-income ($d = 0.63$) or high-income students ($d = 0.51$). A similar pattern was observed for self-efficacy and intrinsic motivation, with significant Assessment Condition x SES interactions in each case (all $p < .01$). The interaction was not statistically significant for test anxiety or student engagement, suggesting that the anxiety-reducing and engagement-promoting effects of formative assessment were relatively uniform across SES groups.

V. DISCUSSION

The findings of this study provide substantial quantitative support for the proposition that formative assessment practices produce superior educational outcomes across multiple dimensions compared to predominantly summative assessment approaches in secondary school contexts. Students in formative assessment classrooms demonstrated higher academic achievement, greater self-efficacy, lower test anxiety, higher engagement, and stronger intrinsic motivation than their peers in summative assessment classrooms, with effect sizes ranging from medium to large. These findings are broadly consistent with the foundational meta-analytic evidence of Black and Wiliam (1998) and with the theoretical predictions of Self-Determination Theory (Deci & Ryan, 2000) and feedback research (Hattie & Timperley, 2007).

The particularly large effect size observed for test anxiety ($d = 0.93$) deserves special attention. This finding suggests that the shift from a predominantly summative to a predominantly formative assessment environment produces a substantial

reduction in the psychological burden of academic assessment for secondary school students, consistent with Putwain's (2007) analysis of the mechanisms through which high-stakes summative environments elevate anxiety. Given the well-documented negative consequences of test anxiety for both academic performance and student well-being, this finding has considerable practical significance for secondary school assessment policy.

The moderation analyses provide compelling evidence that the benefits of formative assessment are not uniformly distributed but are particularly pronounced for students from low-income backgrounds, at least with respect to academic achievement, self-efficacy, and intrinsic motivation. This finding aligns with Wiliam's (2011) equity argument and extends it with quantitative evidence: formative assessment appears to function as an equalizing force in secondary education, providing students who lack access to private tutoring and supplementary academic support with the ongoing, actionable feedback that more advantaged students may obtain through private means.

Several limitations of the present study warrant acknowledgment. First, the quasi-experimental design, while appropriate given the constraints of school-based research, does not permit causal inference with the confidence of a fully randomized controlled trial. Although groups were well-matched on observed pre-intervention characteristics, unobserved confounds cannot be ruled out. Second, the intervention was delivered over a single semester, and it is possible that longer implementation periods would yield larger or more differentiated effects. Third, the study was conducted in a single metropolitan region, and replication across diverse geographic, cultural, and institutional contexts will be necessary to establish the generalizability of findings.

VI. CONCLUSION

This quantitative study contributes to the evidence base on assessment practices in secondary education by demonstrating, through a large-sample quasi-experimental design, that formative assessment classrooms produce significantly better outcomes than predominantly summative classrooms across achievement, motivational, and affective dimensions. The findings support continued policy investment in formative assessment professional development for secondary teachers and underscore the equity implications of assessment design, with particular urgency for schools serving economically disadvantaged student populations.

For school leaders and policymakers, the evidence suggests that structural reforms reducing the dominance of high-stakes summative assessment in secondary education may yield significant dividends not only in achievement but in student motivation and psychological well-being. For researchers, future studies employing randomized controlled designs, longer intervention periods, and more diverse samples will be essential to further refine understanding of the conditions under which formative assessment produces its most powerful effects and for which student populations its benefits are greatest.

REFERENCES

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology, 44*(5), 427–445.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Benson, J., & El-Zahhar, N. (1994). Further refinement and validation of the Revised Test Anxiety Scale. *Structural Equation Modeling, 1*(3), 203–221. https://doi.org/10.1080/10705519409539975

Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment* (pp. 13–36). Springer.

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

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Harlen, W. (2005). Teachers' summative practices and assessment for learning: Tensions and synergies. *Curriculum Journal, 16*(2), 207–223. https://doi.org/10.1080/09585170500136093

Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice, 10*(2), 169–207. https://doi.org/10.1080/0969594032000121270

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28–37. https://doi.org/10.1111/j.1745-3992.2011.00220.x

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership, 63*(3), 18–24.

Putwain, D. W. (2007). Researching academic stress and anxiety in students: Some methodological considerations. *British Educational Research Journal, 33*(2), 207–219. https://doi.org/10.1080/01411920701208258

Ryan, A. M., & Brown, J. L. (2005). Assessment for learning, for keeps: Understanding formative assessment in secondary school contexts. *Theory Into Practice, 44*(1), 5–14.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535–550. https://doi.org/10.1080/02602930903541015

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Routledge.

Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.