Journal of Advanced Sciences and Mathematics Education
Volume 6, Issue 4, 115-129
e_ISSN: 2798-1606
DOI: 10.58524/jasme.v6i4.1668

Rising scores, persistent gaps: Mathematics readiness after gasing matriculation at Institut Teknologi Del in the 2026/2027 academic year

Oktofianus Sedubun*
Univertitas Pendidikan Indonesia, INDONESIA

Mubiar Agustin
Univertitas Pendidikan Indonesia, INDONESIA

Asep Deni Gustiana
Univertitas Pendidikan Indonesia, INDONESIA

Yohanes Surya
IT del, INDONESIA

Mai Rosalli Matondang
Univertitas Pendidikan Indonesia, INDONESIA

Article Info

Article history:
Received: Aug 25, 2026
Revised: Oct 10, 2026
Accepted: Oct 26, 2026

Keywords:
GASING Method; Mathematics Readiness; Engineering Education; Normalised Gain; Matriculation; Secondary–Tertiary Transition.

Abstract

Background: Incoming engineering students often have unequal mathematical preparation, while evaluations of intensive matriculation programmes tend to emphasise average improvements rather than persistent differences in readiness.
Aim: This study examined changes in mathematics readiness following GASING-based matriculation at Institut Teknologi Del, focusing on differences across instructional groups and students' initial mathematical abilities.
Method: A quantitative one-group pre-test–post-test design was employed involving 282 students from eleven instructional groups and ten study programmes. The programme comprised 48 instructional hours over six days. Data were analysed using paired-samples t-tests, normalised gain, analysis of covariance, correlation analysis, baseline quartiles, and descriptive dispersion measures.
Result: Mean scores increased from 16.72 to 57.05, representing a difference of 40.34 points (95% CI [38.45, 42.22]). The within-subject effect size was large (Cohen's dz = 2.51), and 99.6% of participants achieved higher final scores. However, only 45.0% reached the 60-point threshold. Mean normalised gain was 0.507, with a positive association between initial ability and normalised gain (r = .600). Absolute score dispersion increased, although relative dispersion decreased.
Conclusion: The programme was followed by substantially higher mathematics scores, but differences associated with initial ability persisted. Given the non-equivalent assessments and absence of a comparison group, causal effectiveness cannot be established. The findings suggest the importance of diagnostic assessment and differentiated mathematical support for students with lower initial readiness.

To cite this article: Sedubun, O., Agustin, M., Gustiana, A, D., Surya, Y., & Matondang, M, R. (2026). Rising scores, persistent gaps: mathematics readiness after gasing matriculation at Institut Teknologi Del in the 2026/2027 academic year. *Journal of Advanced Sciences and Mathematics Education, 6*(4), 115-129.

INTRODUCTION

The transition from secondary to university mathematics can expose a mismatch between knowledge assumed by lecturers and knowledge available to incoming students. Students beginning engineering degrees are expected to use algebra, functions, and trigonometry while learning increasingly abstract material, often with little time for remediation. International work describes this transition not only as a cognitive discontinuity but also as a change in academic expectations and students' relationships with mathematics (Di Martino et al., 2023a; Di Martino et al., 2023b). Difficulties at this point are consequential because early mathematics courses structure progression through many engineering curricula (Charalambides et al., 2023; Curran et al., 2026; Morán-Soto et al., 2026). The educational problem is therefore not simply that some entrants score poorly, but that a single first-semester curriculum must accommodate substantially different starting points.

Unequal mathematics preparation is particularly pertinent in Indonesia, where performance on international assessment has raised questions about foundational proficiency. In PISA 2022, 18%

* **Corresponding author:**
Oktofianus Sedubun, Univertitas Pendidikan Indonesia, Indonesia
oktofianus@upi.edu

of Indonesian students reached at least Level 2 in mathematics, compared with an OECD average of 69% (OECD, 2023). These figures describe school-age populations and do not directly measure the entrants at Institut Teknologi Del, but they establish why readiness warrants institutional attention. Evidence from Germany and Sweden likewise indicates that entering mathematical competencies vary over time and depend on educational trajectories (Büchele & Feudel, 2023; Hed et al., 2025). Incoming preparation may also intersect with differential academic opportunities (Salehi et al., 2020). A short matriculation programme can provide a common learning opportunity, yet shared exposure does not ensure that all students attain a common level.

Mathematical prerequisites are also linked to later performance, not merely to initial placement. Research with engineering students has documented the importance of prior experience, confidence, and specific conceptual weaknesses at the start of university study (Charalambides et al., 2023; Mahadewsing et al., 2024). Mosia et al. (2025) found that secondary mathematics preparation and university mathematics credits were associated with subsequent performance, while Guerra et al. (2026) proposed readiness indicators for earlier academic support decisions. Studies of interest and self-concept show that cognitive preparation exists alongside affective dimensions of transition (Geisler, Rolka, & Rach, 2023; Kirsten et al., 2026). These strands suggest that readiness should be assessed with attention to both average attainment and the students who remain furthest from prerequisite competence.

Bridge and preparatory programmes address this transition by making prerequisite mathematics an explicit instructional target before first-semester coursework begins. Bradford et al. (2021) found positive average outcomes across university STEM summer bridge programmes, whereas Lynch et al. (2023) synthesised more modest effects on mathematics achievement across summer programmes. Studies of preparatory-course participation, delivery format, and learner motivation have shown that success depends partly on what is taught, how it is assessed, and the population served (Gerdes et al., 2022; Kirsten & Greefrath, 2024; Mayerhofer et al., 2025). Research in related STEM settings likewise illustrates that short preparatory courses cannot be treated as interchangeable interventions (Eitemüller & Habig, 2020). More recent engineering evidence relates mathematical preparedness and tailored support to subsequent academic performance (Brumfield et al., 2025; Morán-Soto et al., 2026; Msomi & Rzyankina, 2025). Consequently, average programme outcomes should be interpreted together with initial differences in attainment and the demands represented by the assessment.

The programme examined here used Matematika GASING (Gampang, Asyik, dan Menyenangkan), developed by Yohanes Surya, which sequences learning from concrete representations towards more abstract mathematical procedures. The higher-education application described by Kusuma and Sulistiawati (2014) concerned multiplication, whereas the present matriculation extended from arithmetic and algebra to topics associated with introductory calculus. A systematic review of 35 GASING-related articles found encouraging evidence for basic arithmetic learning but also recurring concerns about small samples, selection bias, missing controls, and limited cross-context evaluation (Gesiradja et al., 2025). In a more recent community-based evaluation, Waimbo et al. (2026) described short-term arithmetic score gains in two non-formal groups, again at a considerably narrower content level. A critical synthesis has also cautioned that a uniform emphasis on speed or procedural success may create new distinctions among learners, although that possibility has not been directly tested in a university matriculation setting (Samaloisa et al., 2026). These studies support investigating how GASING is experienced at a different educational level without assuming that evidence from elementary arithmetic transfers unchanged to engineering readiness.

An additional gap concerns the distribution of measured progress rather than its average magnitude. Cumulative-advantage accounts suggest that students who already possess relevant knowledge may find it easier to incorporate new procedures (Merton, 1968; Stanovich, 1986), and contemporary higher-education evidence links growth trajectories to initial competencies (Büchele & Marten, 2026). Yet a meta-analysis of domain-specific prior knowledge indicates that the form of this relationship is sensitive to learning context and outcome measurement (Simonsmeier et al., 2022). Normalised gain deserves particular caution because its denominator is determined by the initial score, and empirical work identifies biases that can affect comparisons across groups (Coletta & Phillips, 2005; Nissen et al., 2018; Navarrete-Ulloa, 2025; Navarrete-Ulloa et al., 2026). A large

cohort mean therefore cannot, by itself, establish that weaker entrants have caught up or that stronger entrants have learned more. Assessing both raw and normalised gains alongside quartile-level attainment and dispersion offers a more discriminating, although still measurement-dependent, description of the programme's reach.

This study evaluated the 48-hour GASING-based mathematics matriculation programme delivered over six instructional days to 282 incoming students from ten study programmes at Institut Teknologi Del. It asked how performance differed between the initial and final assessments, how those differences varied across eleven intact groups after accounting for entry scores, and whether final attainment remained differentiated by starting ability. Its contribution is not an experimental estimate of the teaching method's effectiveness: there was no comparison group and the two tests assessed partly different mathematical content. Instead, the study joins individual gains, adjusted group comparisons, entry-quartile patterns, and absolute and relative dispersion in a single institutional evaluation. Conceptually, it tests whether a favourable cohort average can coexist with persistent low-end readiness and whether interpretations based on normalised gain align with those based on raw attainment. Practically, the findings can inform diagnostic assessment and the design of additional support for students who remain least prepared for first-semester mathematics.

METHOD

Research design and setting

The study adopted a quantitative, one-group pre-test–post-test programme-evaluation design with repeated score measurement in intact instructional groups (Campbell & Stanley, 1963). The institution required incoming students to participate in matriculation, so no untreated comparison group was formed. Accordingly, the analysis assessed change associated with the programme period rather than a causal effect of the GASING approach. Data were recorded at Institut Teknologi Del in Laguboti, Toba Regency, North Sumatra, between 27 August and 2 September 2026. Teaching occurred on six instructional days, namely 27, 28, 29, and 31 August and 1 and 2 September; 30 August was excluded because it was a Sunday.

Participants and analytical records

The register initially contained 339 names across ten study programmes. The analytical inclusion criterion was the availability of both pre-test and post-test scores; 36 records lacked an initial score, 33 lacked a final score, and one row contained aggregate means rather than a participant record. Because missingness could overlap across the two tests, the numbers of missing scores should not be treated as mutually exclusive categories. Filtering produced 282 complete pairs, equivalent to 83.2% of the registered cohort, and the study therefore used a complete-case census of eligible participants rather than probability sampling. Eleven intact groups contained between 23 and 28 analysed students each, with group composition based on study programme rather than random allocation. Six groups represented single study programmes and five combined two or three related programmes.

For students with a pre-test score who did not complete the post-test (n = 21), mean entry score was 15.67 compared with 16.72 among complete cases. The reported Welch comparison gave $t(23.1) = -0.35$, $p = .73$, providing no evidence of an entry-score difference in this observable subset. This check does not rule out selection associated with unmeasured characteristics or with the participants who missed the initial assessment. Complete-case comparisons and the limitations of missing data are therefore retained in the interpretation.

Matriculation programme and data collection

The GASING-based programme provided 48 contact hours, eight hours per instructional day. Four sessions were scheduled daily from 08.00 to 10.00, 10.00 to 12.00, 13.00 to 16.00, and 16.00 to 17.00, with the final period reserved for questions and answers. Lecture and question-and-answer sessions took place in the auditorium, while video-based consolidation was provided in classrooms and during dormitory study periods. Content progressed from arithmetic, fractions, and decimals to lines and inequalities; logarithms and exponents; series and polynomials; matrices and trigonometry; and limits and introductory calculus. Derivatives and integrals appearing on the final assessment were introduced through video-based consolidation. Yohanes Surya served as principal instructor and five assistants supported participants who experienced difficulty.

The evaluation sequence began with administrative registration and an initial constructed-response mathematics test, followed by the six instructional days and a final constructed-response test. Assistants awarded partial credit for written solutions and recorded scores on a 0–100 scale. The final analytical file was formed by matching scores and removing records without complete pairs. Scores were then summarised for the full cohort, intact groups, academic study programmes, and quartiles of entry ability. Figure 1 presents this workflow as a programme evaluation rather than a controlled experiment.

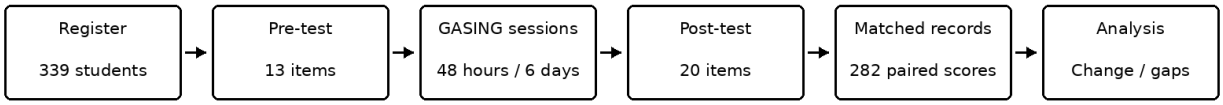


Figure 1. Research procedure for the GASING-based mathematics matriculation programme

Figure 1 shows the sequence from registration to statistical evaluation. The initial and final assessments were administered around the six-day matriculation programme. Complete-case matching reduced the analytical dataset to 282 paired records. Subsequent analyses addressed score change, intact-group differences, and distributions by initial ability. No control or random-assignment step appears in the diagram because the evaluation did not include either procedure.

Instruments and assessment comparability

The pre-test comprised 13 constructed-response items weighted at six points for each of the first two items and eight points for each of the remaining eleven items, for a total of 100 points. The post-test comprised 20 constructed-response items with weights of 3, 3, 3, 4, 3, 4, 3, 3, 4, 5, 6, 5, 6, 5, 5, 6, 8, 8, 8, and 8 points, also totalling 100. The pre-test covered radicals, linear equations, exponents, quadratic equations, and quadratic functions. The post-test included arithmetic, radicals, exponents, linear equations, ratio and proportion, trigonometry, derivatives, and integrals. Only 44 pre-test points and 25 post-test points concerned overlapping topics, and 38 final-test points were assigned to derivatives and integrals. Therefore, the two total scores shared a numeric scale but did not constitute equivalent measures of identical content. Differences between them should be understood as changes in test performance during programme implementation, not as a calibrated estimate of growth on a common test form.

The assistants scored constructed responses using partial credit under direct guidance from the principal instructor. Item-level response records were not retained, so empirical internal-consistency coefficients and inter-rater agreement statistics were unavailable. Alignment between the test topics and the programme syllabus provides a descriptive content map but does not establish that the two instruments have equivalent difficulty or measurement properties. The absence of item-level information also limits decomposition of gains by common content domain. These restrictions are central to the interpretation of the large observed score difference.

Data analysis

Analyses were conducted in Python 3 using pandas (McKinney, 2010), SciPy 1.17.1 (Virtanen et al., 2020), and statsmodels 0.15.0 (Seabold & Perktold, 2010). Descriptive summaries included the mean, standard deviation, median, quartiles, range, and proportions above a reported score threshold of 60. Difference-score normality was assessed with Shapiro–Wilk (Shapiro & Wilk, 1965), $W = 0.993$, $p = .230$, before calculating a paired-samples t-test, Cohen's dz (Cohen, 1988), and a 95% confidence interval for the average difference. Individual normalised gain was calculated as (post-test minus pre-test)/(100 minus pre-test) and classified using conventional thresholds of 0.30 and 0.70 (Hake, 1998). Because normalised gain depends on baseline scores and may be influenced by test reliability, raw gain, correlations, quartile summaries, and score dispersion were examined alongside it (Coletta & Phillips, 2005; Nissen et al., 2018; Navarrete-Ulloa, 2025; Navarrete-Ulloa et al., 2026). All reported tests were two-tailed with $\alpha = .05$.

Levene's tests (Levene, 1960) indicated unequal variances for group comparisons, $p < .001$, so group-level normalised gain was evaluated by Welch analysis of variance (Welch, 1951) and corroborated with Kruskal–Wallis (Kruskal & Wallis, 1952); Games–Howell (Games & Howell, 1976) was used for post-hoc comparisons. The association of group membership with final scores conditional on entry score was evaluated using analysis of covariance, following a test of the group-by-entry-score interaction. Because groups were intact and differed at baseline, adjusted means were

interpreted descriptively, not as randomised group treatment effects (Miller & Chapman, 2001). Pearson and Spearman coefficients described associations between entry score and normalised gain; Pearson correlations were also calculated using study-programme means. Baseline quartiles and descriptive comparisons of standard deviation, interquartile range, and coefficient of variation were used to characterise the distribution of observed scores. The reported Levene statistic comparing pre-test and post-test distributions is retained descriptively, but its usual independent-group interpretation is not appropriate for paired measurements.

Permissions and researcher positionality

Institut Teknologi Del permitted the use of programme activity records, and Yohanes Surya permitted use of the data and description of the instructional approach. Participant names were replaced by codes before analysis and no identifying records are reproduced here. The records originated in an institutional educational activity rather than a separately administered research intervention. The first author was a teaching assistant in the programme, a role that could be associated with delivery and scoring decisions and is disclosed accordingly. The evaluation is limited to immediate test outcomes; first-semester calculus performance was not tracked.

RESULTS AND DISCUSSION

Observed results

Among 339 registered entrants, 282 (83.2%) had paired scores suitable for analysis. Their initial mathematics scores differed markedly across the eleven intact groups: group pre-test means ranged from 6.86 to 22.93, Welch $F(10, 104.8) = 11.45$, $p < .001$. As the instructional groups were organised by study programme, the differences confirm that the students did not begin from a common achievement baseline. They also justify interpreting end-of-programme contrasts against the initial-score distribution rather than presenting unadjusted group rankings as evidence of different instruction quality.

Across the full cohort, mean score increased from 16.72 (SD = 13.29) on the pre-test to 57.05 (SD = 22.76) on the post-test. The mean paired difference was 40.34 points (SD = 16.08, 95% CI [38.45, 42.22]), and median score moved from 13.0 to 56.0. The observed ranges were 0–68 initially and 13–100 finally. At the fixed descriptive score threshold of 60, 3 students (1.1%) reached the threshold initially, compared with 127 (45.0%) at the end. Table 1 details the descriptive distributions and the raw differences underlying these summaries.

Table 1. Descriptive statistics for pre-test, post-test, and gain scores (n = 282)

Statistic	Pre-test	Post-test	Gain
Mean	16.72	57.05	40.34
Standard deviation	13.29	22.76	16.08
Median	13.00	56.00	41.00
First quartile	7.00	39.00	30.00
Third quartile	23.00	75.00	51.00
Minimum	0.00	13.00	-3.00
Maximum	68.00	100.00	81.00
Coefficient of variation	0.795	0.399	-
Students scoring 60 or above	3 (1.1%)	127 (45.0%)	-

Table 1 indicates a 40.34-point average difference between the two assessments. The median difference of 41 points is close to the mean, although individual differences ranged from -3 to 81 points. Scores were more dispersed at the final measurement, as reflected in the standard deviations. The increase in the number of students at or above 60 shows a substantial change in the reported score distribution. Because the test blueprints were not equivalent, the difference should not be interpreted as a direct measure of improvement on identical mathematics content.

The difference scores were approximately normal under the reported Shapiro–Wilk test, $W = 0.993$, $p = .230$. A paired-samples t-test showed a statistically reliable difference, $t(281) = 42.12$, $p < .001$, with Cohen's $d_z = 2.51$. A total of 281 students (99.6%) scored higher on the final assessment, while one scored lower. Mean normalised gain was 0.507 (SD = 0.235); 65 students fell in the high

category, 156 in the medium category, and 61 in the low category. The pre-test and post-test correlation was r = .721, p < .001, consistent with appreciable continuity in relative standing. Table 2 collects the paired-test and normalised-gain statistics.

Table 2. Pre-post comparison and distribution of normalised gain (n = 282)

Statistic	Value	Note
Shapiro-Wilk W (difference scores)	0.993	p = .230
Paired-samples t	42.12	df = 281, p < .001
Mean difference [95% CI]	40.34 [38.45, 42.22]	points
Cohen's dz	2.51	within-subject
Pre-post correlation (r)	0.721	p < .001
Students with higher post-test score	281 (99.6%)	1 student scored lower
Mean normalised gain	0.507 (SD = 0.235)	medium category
High normalised gain (g >= 0.70)	65 (23.0%)	-
Medium normalised gain (0.30 <= g < 0.70)	156 (55.3%)	-
Low normalised gain (g < 0.30)	61 (21.6%)	-

Table 2 shows that the difference in scores was statistically reliable for the complete-case cohort. The effect-size statistic represents a within-participant contrast, rather than an effect estimated against a separate untreated group. The normalised-gain distribution included students in all three conventional gain categories. The substantial correlation between the two test totals suggests that students' relative ordering was not completely reshuffled. The findings support a description of widespread higher final-test performance, subject to the instrument-comparability limitation.

Normalised gain differed between intact groups, with group means ranging from 0.35 (Group 8) to 0.62 (Group 10). Welch's test yielded F(10, 104.3) = 8.03, p < .001, and Kruskal–Wallis yielded H(10) = 42.69, p < .001. Eight of the 55 Games–Howell comparisons reached the .05 threshold, with the largest difference observed between Groups 8 and 10 (difference = 0.278, p = .012). In the adjusted post-test model, the homogeneous-slopes interaction was not significant, F(10, 260) = 0.93, p = .508. Entry score was associated with final score, F(1, 270) = 259.67, p < .001, partial η^2 = .490, while the conditional group term was F(10, 270) = 3.77, p < .001, partial η^2 = .123; model R^2 was .578. Table 3 shows the observed and adjusted means needed to distinguish the two descriptions.

Table 3. Observed and covariate-adjusted outcomes by intact study group

Group	n	Pre-test M	Post-test M	Gain M	Normalised gain M	Adjusted post-test M
1	27	21.48	66.30	44.81	0.60	60.66
2	23	19.22	64.91	45.70	0.59	61.96
3	26	17.00	61.35	44.35	0.56	61.01
4	25	18.20	52.44	34.24	0.44	50.69
5	28	6.86	42.89	36.04	0.39	54.54
6	25	21.52	61.68	40.16	0.54	56.00
7	24	11.08	58.79	47.71	0.54	65.45
8	25	12.72	41.08	28.36	0.35	45.80
9	25	11.96	46.12	34.16	0.40	51.74
10	28	22.93	69.00	46.07	0.62	61.66
11	26	20.77	62.85	42.08	0.54	58.06

Table 3 shows meaningful variation in both the initial and final group means. Groups with similar final scores sometimes had different baseline scores and, conversely, similar starting means could accompany different final outcomes. Adjustment changed the ordering of groups, with Group 7 reaching the highest adjusted post-test mean. The observed range of group final-score means was larger than the range of adjusted means. These values do not isolate an instructional group effect because baseline differences and unmeasured allocation factors remain possible explanations.

The study-programme summaries in Table 4 offer a second description of heterogeneity at a different unit of aggregation. Across ten programmes, the reported correlation between mean initial score and mean normalised gain was r = .819. However, several programmes contained relatively

small numbers of complete cases and their mean values may be unstable. These programme-level associations should not be treated as direct evidence about individual students or programme-specific instruction.

Table 4. Outcomes by study programme

Study programme	n	Pre-test M	Post-test M	Gain M	Norm. gain M
Metallurgical Engineering	37	23.22	69.27	46.05	0.63
Information Systems	30	22.23	68.83	46.60	0.62
Engineering Management	39	15.51	59.36	43.85	0.54
Informatics	23	22.13	63.09	40.96	0.54
Electrical Engineering	4	19.75	62.25	42.50	0.54
Applied Software Engineering	68	13.47	54.31	40.84	0.49
Bioprocess Engineering	11	17.09	52.55	35.45	0.45
Biotechnology	14	19.07	52.36	33.29	0.43
Information Technology	36	11.86	43.97	32.11	0.38
Computer Technology	20	9.85	42.95	33.10	0.37

Table 4 displays differences in the size and starting position of the ten academic programmes. Metallurgical Engineering and Information Systems had comparatively high observed average final scores in this dataset. Computer Technology and Information Technology recorded lower final means and lower mean normalised gains. Electrical Engineering contained only four analysed participants, making its mean particularly sensitive to individual scores. Programme-level patterns are therefore informative for support planning but do not justify causal comparisons between degree programmes.

When baseline scores were partitioned into four quartiles, mean normalised gain increased monotonically from 0.366 to 0.429, 0.539, and 0.716. The associated final-score means were 39.35, 48.73, 61.86, and 81.24. Pearson correlation between initial score and individual normalised gain was $r = .600$, $p < .001$, and Spearman $\rho = .589$, $p < .001$; by comparison, the correlation with raw gain was $r = .194$, $p = .001$. The divergence between the two correlation magnitudes matters because normalised gain is a transformation that incorporates the pre-test score. Table 5 provides the distribution of baseline quartiles and their recorded outcomes.

Table 5. Outcomes by baseline ability quartile

Quartile	n	Pre-test range	Pre-test M	Post-test M	Gain M	Norm. gain M (SD)
Q1 (lowest)	75	0-7	4.40	39.35	34.95	0.366 (0.156)
Q2	78	8-13	10.32	48.73	38.41	0.429 (0.183)
Q3	59	14-23	17.61	61.86	44.25	0.539 (0.215)
Q4 (highest)	70	24-68	36.29	81.24	44.96	0.716 (0.221)

Table 5 shows that post-test means were ordered in the same direction as baseline means. Students in the lowest quartile finished with a mean of 39.35 points, compared with 81.24 in the highest quartile. Raw mean gains were positive in all four groups, from 34.95 to 44.96 points. Normalised mean gains rose more steeply across quartiles than did raw gains. This descriptive pattern is consistent with a persistent entry gap, but its magnitude is sensitive to differences between the pre-test and post-test content.

The standard deviation rose from 13.29 to 22.76, and the interquartile range increased from 16 to 36 points. In contrast, the coefficient of variation declined from 0.795 to 0.399 because the mean score increased substantially. The reported Levene statistic was $F = 99.94$, $p < .001$, but a classical independent-samples comparison of variances does not account for the repeated-measures structure of these scores. Table 6 therefore places emphasis on the descriptive dispersion indices rather than using this test to establish a formal paired-variance difference.

Table 6. Absolute and relative dispersion of mathematics scores

Dispersion indicator	Pre-test	Post-test
Standard deviation	13.29	22.76
Interquartile range	16.00	36.00
Coefficient of variation	0.795	0.399
Levene test, pre versus post	$F = 99.94$	$p < .001$

Table 6 indicates that the absolute spread of observed scores was greater on the final test. The larger interquartile range likewise reflects a greater distance between the middle fifty percent of

recorded scores. The coefficient of variation moved in the opposite direction because it expresses spread relative to the corresponding mean. Absolute and relative dispersion therefore answer different descriptive questions and should not be collapsed into a single assertion about convergence. Differences in test content further limit interpretation of dispersion as a change in a single latent ability distribution.

The distributions of pre-test and post-test totals and the individual gain categories are visualised in Figure 2. The figure complements the descriptive statistics by showing where the observed scores concentrate, rather than reporting means alone.

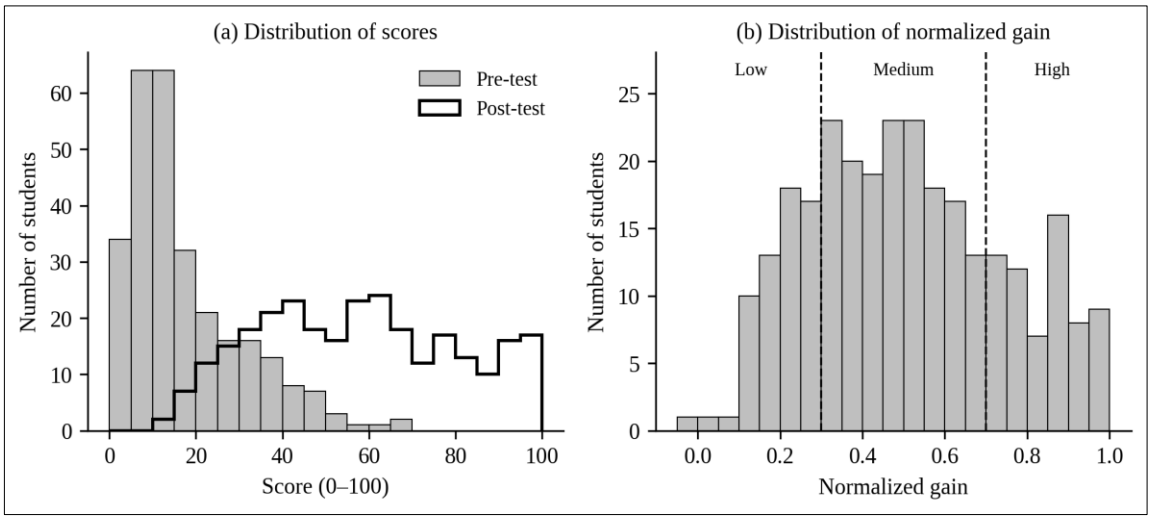


Figure 2. Distribution of (a) pre-test and post-test scores and (b) normalised gain (n = 282)

Figure 2a shows that initial scores were concentrated at the lower end of the scale. The final assessment has a much broader spread, extending to the scale maximum. Figure 2b presents students across low, medium, and high normalised-gain intervals. The middle category contains the largest number of students according to Table 2. The figure depicts differences in score distributions without establishing equivalent measurement across the two assessments.

Figure 3 visualises the monotonic quartile pattern and the positive association among programme means. These displays are useful for distinguishing distributional differences from the magnitude of the cohort mean change.

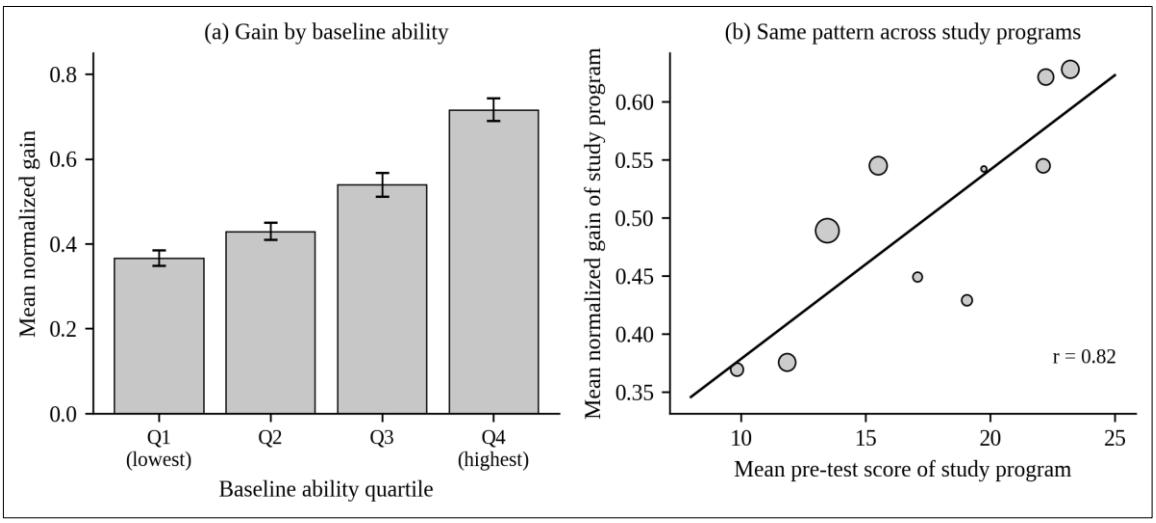


Figure 3. Normalised gain by (a) baseline ability quartile and (b) study programme (n = 282; ten programmes).

Figure 3a shows increasingly high average normalised gains from the lowest to the highest entry quartile. Its error bars indicate variability around the quartile means. Figure 3b plots each programme as a unit and shows the reported positive association with baseline score. The

programme-level plot does not describe within-programme variability or causal pathways. Both panels should be interpreted alongside raw gains and the non-equivalence of test content.

Interpretation and comparison with previous research

The most visible finding was a near-universal increase in mathematics scores across the six instructional days. This pattern is consistent in direction with positive short-term outcomes reported by STEM bridge programmes (Bradford et al., 2021; Mayerhofer et al., 2025; Brumfield et al., 2025) and with arithmetic-focused GASING evaluations (Gesiradja et al., 2025; Waimbo et al., 2026). The size of the within-student contrast, however, cannot be compared directly with controlled between-group effect estimates, because they describe different quantities. Nor does an increase in the mean show that students have reached a common prerequisite level: fewer than half attained 60 on the final assessment. The distinctive contribution of this evaluation is therefore the simultaneous documentation of widespread higher final scores and a substantial remaining lower-readiness group. For institutional practice, a cohort-wide increase is one indicator of programme performance, but cannot serve as its only criterion.

The comparison of initial and final performance is constrained by assessment construction. The pre-test used thirteen items centred largely on algebraic foundations, whereas the post-test contained twenty items and sampled additional topics, including calculus. Because these assessments lack a common content blueprint or equated scale, the 40.34-point mean difference cannot be assigned solely to learning in a stable measurement domain. Scholarship on mathematics assessment emphasises the importance of aligning tasks with the construct being interpreted (Evans & Jeong, 2023), and recent methodological work shows that change estimates can depend strongly on score transformations (Navarrete-Ulloa, 2025; Navarrete-Ulloa et al., 2026). A broader final-test syllabus is not proof that the final instrument was harder on a common scale. Accordingly, the responsible claim is that observed totals were higher after matriculation, while how much comparable mathematical competence changed remains unresolved by the existing records. Future evaluations would benefit from common anchor items and explicit scoring rubrics, rather than from further reanalysis of these non-equivalent totals.

A second result concerns systematic differences between intact instructional groups. Statistically detectable variation was driven mainly by a limited set of pairwise contrasts, which cautions against interpreting the omnibus test as evidence that every group followed a different learning trajectory. The adjusted model also placed a much larger partial η^2 on initial score than on group membership, consistent with the importance of prior preparation reported in engineering and university mathematics research (Charalambides et al., 2023; Mosia et al., 2025; Guerra et al., 2026). At the same time, the adjusted group term remained statistically detectable, leaving plausible influences such as peer composition, assistance, or engagement. Since groups were formed by academic programme and these variables were not measured, the model cannot establish which contextual feature produced the remaining variation. This interpretation is aligned with the methodological distinction between statistical adjustment and causal identification (Miller & Chapman, 2001).

The increasing sequence of mean normalised gains across entry quartiles resembles the cumulative-advantage pattern predicted when stronger prerequisite knowledge supports further learning (Merton, 1968; Stanovich, 1986; Büchele & Marten, 2026). Prior-knowledge studies in higher mathematics also show why algebraic foundations matter for later calculus tasks (Alreshidi, 2023; Mahadewsing et al., 2024). Nevertheless, the observed gradient is not a direct test of that learning mechanism. The positive association of entry score with raw gain was considerably weaker than its association with normalised gain, whose formula incorporates the entry score in the denominator. Furthermore, the final test included content requiring prerequisites that were not equally represented in the initial instrument. The quartile gradient therefore demonstrates a distributional pattern in these particular scores, not proof that GASING itself disproportionately benefited stronger students. This distinction extends the programme-evaluation literature by making the choice of gain indicator part of the substantive equity interpretation rather than treating it as a neutral reporting decision.

Whether a cohort is becoming more equitable also depends on the metric used to describe its spread. Standard deviation and interquartile range were greater for the final-test totals, while the

coefficient of variation declined as the mean increased. These movements are mathematically compatible and cannot be reduced to a single claim that the readiness gap either widened or narrowed in every sense. The lowest entry quartile had a final mean of 39.35, and 55.0% of the cohort remained below the descriptively reported score of 60. This pattern is especially relevant to an institution seeking to prepare all entrants for common first-semester requirements; however, 60 was not established here as a validated calculus-readiness cut score. Work on active learning and metacognitive problem-solving identifies possible directions for targeted support (Theobald et al., 2020; Follmer et al., 2026), but those approaches were not tested in this programme. The practical implication is to combine common matriculation with follow-up assessment and additional teaching where prerequisite weaknesses are concentrated, while evaluating the effects of that support separately.

International transition research increasingly treats first-year mathematical adjustment as cognitive, affective, and institutional rather than solely procedural (Di Martino et al., 2023a, 2023b; Nieminen et al., 2024). For example, Geisler, Rolka, and Rach (2023) and Kirsten et al. (2026) linked affective characteristics and students' mathematical self-concept with university study experiences. Other work has compared instructional formats, active problem-solving, and opportunities to apply mathematics in unfamiliar contexts (Kirsten & Greefrath, 2024; Baumgartner et al., 2025; Weber et al., 2025; Zakariya et al., 2025). Recent GASING synthesis likewise suggests that instructional pacing, classroom interaction, and procedural expectations deserve attention when assessing inclusion (Samaloisa et al., 2026). The present data do not include measures of motivation, anxiety, participation, assistant contact, or individual learning strategies. These mechanisms therefore remain plausible explanations drawn from the literature, not measured mediators of the observed score patterns. They nevertheless provide a rationale for designing future evaluations around more than the test average.

At the disciplinary level, the programme bridges arithmetic and basic algebra with demands that eventually reach limits, derivatives, and integrals. Studies have described how this transition requires learners to reorganise earlier mathematical procedures for new forms of reasoning (Schüler-Meyer, 2022; Kontorovich et al., 2024). Research on calculus-course mathematical modelling illustrates additional forms of mathematical engagement (Bogen et al., 2025). The importance of interpreting mathematics within other STEM disciplines has also been documented in work on the mathematics–physics interface (Hellio et al., 2025). The higher final-test totals therefore reflect performance on a different mix of mathematical tasks, not simply repeated success on the same initial questions. The instructional sequence may have helped students encounter a wider set of topics, but the present records cannot identify which components of the GASING approach were most influential. This boundary preserves the distinction between a useful institutional programme evaluation and an instructional-method experiment.

Together, these results suggest a more discriminating framework for judging short intensive matriculation programmes. Syntheses of bridging interventions often foreground average academic outcomes (Bradford et al., 2021; Lynch et al., 2023), while this case illustrates how a favourable mean change may coexist with strongly stratified final assessment scores. The conceptual contribution is the separation of three questions that are often conflated: whether final totals are higher, whether the weakest entrants attain a useful prerequisite standard, and whether apparent gain differences are stable across scoring metrics. This distinction is especially necessary when assessment forms are not content-equivalent. In practice, an institution could use diagnostic results to identify students needing additional algebra or calculus preparation, provide structured follow-up without imposing unnecessary progression barriers, and monitor subsequent course outcomes (Engelhardt et al., 2026). These are recommendations arising from a programme evaluation, not demonstrated effects of a differentiated intervention. A stronger subsequent study would combine comparable test forms with first-semester calculus outcomes and, where feasible, an appropriate comparison condition.

Limitations and implications for future research

The evaluation was conducted in one institution, without a comparison condition, and the two tests differed in content coverage and number of items. Consequently, neither the large within-student score difference nor the gradient in normalised gain identifies a causal learning effect or a verified change on an equated mathematics-readiness scale. Item-level responses were not retained,

precluding internal-consistency estimates, inter-rater reliability analysis, and calibration of overlapping content. Although the complete-case sample contained most registered entrants, attrition could still reflect unmeasured differences. Group allocation was not random, and one author served as an instructional assistant, which limits independent attribution of group differences. No downstream calculus grades or persistence outcomes were observed. Subsequent research could administer anchored pre- and post-assessments, retain scoring rubrics and item-level data, document student experiences, and follow students into their first mathematics courses. These steps would permit a clearer test of whether targeted follow-up closes prerequisite gaps rather than only changing short-term test distributions.

CONCLUSION

The 48-hour GASING-based matriculation programme at Institut Teknologi Del was followed by higher final-test totals for 281 of 282 analysed students. Mean scores differed by 40.34 points, but only 45.0% of students reached 60 on the final assessment, and the lowest entry quartile continued to have substantially lower final scores than the highest quartile. Entry score was more strongly associated with final attainment than intact-group membership, while normalised gains varied more sharply by starting ability than raw gains. Absolute score dispersion increased and relative dispersion declined, revealing different aspects of the same observed distributions. The study's principal contribution is to show why average score change alone is insufficient to describe the reach of a uniform university matriculation programme. The absence of a comparison group and the differences between the two assessments prevent claims about a causal GASING effect or a common-scale improvement in readiness. Universities may use these descriptive findings to motivate diagnostic follow-up and differentiated prerequisite support, whose effectiveness should be evaluated with comparable assessments and later-course outcomes.

ACKNOWLEDGMENT

The authors thank Prof. Yohanes Surya for permission to study the matriculation programme and describe the GASING approach, the five assistants involved in programme delivery and assessment, and Institut Teknologi Del for access to the administrative records. The authors also acknowledge the incoming 2026/2027 students whose anonymised data were analysed.

AUTHOR CONTRIBUTIONS STATEMENT

Oktofianus Sedubun contributed to conceptualisation, data collection, statistical analysis, interpretation, and original drafting. Mubiar Agustin and Asep Deni Gustiana contributed to methodology, critical review, and academic supervision. Yohanes Surya contributed to developing and delivering the GASING-based matriculation programme and to subject-matter interpretation. Mai Rosalli Matondang contributed to manuscript review and editing.

REFERENCES

Alreshidi, N. A. K. (2023). Enhancing topic-specific prior knowledge of students impacts their outcomes in mathematics. *Frontiers in Education, 8*, 1050468. https://doi.org/10.3389/feduc.2023.1050468

Baumgartner, V., Daguati, S., Trninic, D., Akveld, M., Caspar, A., Hungerbühler, N., & Kapur, M. (2025). Problem-solving before instruction for learning linear algebra in university mathematics. *Instructional Science, 53*, 1573–1602. https://doi.org/10.1007/s11251-025-09709-8

Bogen, S. C., Neequaye, B. N. D., Thomas, B., & Kohler, B. R. (2025). Mathematical modeling projects as a vehicle toward doing mathematics in a traditional calculus course. *PRIMUS, 35*(1), 65–85. https://doi.org/10.1080/10511970.2024.2445507

Bradford, B. C., Beier, M. E., & Oswald, F. L. (2021). A meta-analysis of university STEM summer bridge program effectiveness. *CBE—Life Sciences Education, 20*(2), ar21. https://doi.org/10.1187/cbe.20-03-0046

Brumfield, M., Mohammadi-Aragh, M. J., & Strawderman, L. (2025). A quantitative analysis on the impact of an engineering focused bridge program for first-time Black students on first semester

mathematics courses. *Journal of STEM Education: Innovations and Research, 26*(3), 40–47. https://doi.org/10.63504/jstem.v26i3.2707

Büchele, S., & Feudel, F. (2023). Changes in students' mathematical competencies at the beginning of higher education within the last decade at a German university. *International Journal of Science and Mathematics Education, 21*(8), 2325–2347. https://doi.org/10.1007/s10763-022-10350-x

Büchele, S., & Marten, C. (2026). Unpacking skill growth in higher education: How initial competencies and engagement shape mathematics achievement. *Higher Education, 92*(3), 1713–1740. https://doi.org/10.1007/s10734-025-01594-x

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

Charalambides, M., Panaoura, R., Tsolaki, E., & Pericleous, S. (2023). First year engineering students' difficulties with math courses: What is the starting point for academic teachers? *Education Sciences, 13*(8), 835. https://doi.org/10.3390/educsci13080835

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

Coletta, V. P., & Phillips, J. A. (2005). Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. *American Journal of Physics, 73*(12), 1172–1182.

Curran, J., Walsh, R., & Guerin, A. (2026). Identifying transient and persistent issues relating to the secondary–tertiary transition and the coping mechanisms of engineering mathematics students. *Teaching Mathematics and its Applications*. https://doi.org/10.1093/teamat/hrag007

Di Martino, P., Gregorio, F., & Iannone, P. (2023a). The transition from school to university in mathematics education research: New trends and ideas from a systematic literature review. *Educational Studies in Mathematics, 113*(1), 7–34. https://doi.org/10.1007/s10649-022-10194-w

Di Martino, P., Gregorio, F., & Iannone, P. (2023b). The transition from school to university mathematics in different contexts: Affective and sociocultural issues in students' crisis. *Educational Studies in Mathematics, 113*, 79–106. https://doi.org/10.1007/s10649-022-10179-9

Eitemüller, C., & Habig, S. (2020). Enhancing the transition? Effects of a tertiary bridging course in chemistry. *Chemistry Education Research and Practice, 21*(2), 561–569. https://doi.org/10.1039/C9RP00207C

Engelhardt, B., Hoem, B., & Johnson, M. (2026). Numeracy, major choice, and the likelihood of graduation. *International Review of Economics Education*, Advance online publication. https://doi.org/10.1016/j.ijme.2026.101471

Evans, T., & Jeong, I. (2023). Concept maps as assessment for learning in university mathematics. *Educational Studies in Mathematics, 113*, 475–498. https://doi.org/10.1007/s10649-023-10209-0

Follmer, D. J., Hut, M., & Santiago, L. Y. (2026). Promoting metacognitive problem-solving among non-calculus-ready first-year engineering students. *Learning and Individual Differences, 130*, Article 102956. https://doi.org/10.1016/j.lindif.2026.102956

Games, P. A., & Howell, J. F. (1976). Pairwise multiple comparison procedures with unequal n's and/or variances: A Monte Carlo study. *Journal of Educational Statistics, 1*(2), 113–125.

Geisler, S., Rach, S., & Rolka, K. (2023). The relation between attitudes towards mathematics and dropout from university mathematics: The mediating role of satisfaction and achievement. *Educational Studies in Mathematics, 112*, 359–381. https://doi.org/10.1007/s10649-022-10198-6

Geisler, S., Rolka, K., & Rach, S. (2023). Development of affect at the transition to university mathematics and its relation to dropout: Identifying related learning situations and deriving possible support measures. *Educational Studies in Mathematics, 113*, 35–56. https://doi.org/10.1007/s10649-022-10200-1

Gerdes, A., Halverscheid, S., & Schneider, S. (2022). Teilnahme an mathematischen Vorkursen und langfristiger Studienerfolg: Eine empirische Untersuchung [Pre-courses in mathematics and long-term academic success: An empirical study]. *Journal für Mathematik-Didaktik, 43*, 377–403. https://doi.org/10.1007/s13138-021-00194-3

Gesiradja, M. C., Sudiapermana, E., Lutfiansyach, D. Y., & Shantini, Y. (2025). Evaluating the effectiveness, challenges, and contextual applications of the GASING mathematics method in Indonesia: A systematic literature review. *Pedagogia: Jurnal Ilmu Pendidikan, 23*(3), 318–335. https://doi.org/10.17509/pdgia.v23i3.89961

Guerra, J., Imbaquingo-Esparza, D. E., Jácome-León, J. G., Flores, S., Umaquinga, A., & Ramírez, F. (2026). Predicting success in first year quantitative courses: An open engineering math readiness score for early levelling decisions. *Frontiers in Education, 11*, 1760230. https://doi.org/10.3389/feduc.2026.1760230

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74.

Hed, L., Åhag, P., & Norqvist, M. (2025). Changes in mathematical skills among freshman engineering students. *International Journal of Mathematical Education in Science and Technology, 56*(11), 2236–2250. https://doi.org/10.1080/0020739X.2024.2404427

Hellio, P., Gueudet, G., & Caussarieu, A. (2025). Supporting the transition between mathematics and physics in the first year of university. *Canadian Journal of Science, Mathematics and Technology Education, 25*, 12–30. https://doi.org/10.1007/s42330-025-00346-4

Kirsten, K., & Greefrath, G. (2024). On-campus vs distance tutorials in preparatory courses for mathematics student teachers: Performance gains and influencing factors. *International Journal of Research in Undergraduate Mathematics Education, 10*, 547–576. https://doi.org/10.1007/s40753-023-00221-3

Kirsten, K., Greefrath, G., & Geisler, S. (2026). Development of interest in mathematics and mathematical self-concept at the transition from school to university and its relation to study success. Educational Studies in Mathematics, 122, 111–133. https://doi.org/10.1007/s10649-025-10469-y

Kontorovich, I., Liu, N. Q., & Kang, S.-W. (2024). Transitioning to proof via writing scripts on the rules of a new discourse. *Educational Studies in Mathematics, 117*, 143–162. https://doi.org/10.1007/s10649-024-10324-6

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47*(260), 583–621.

Kusuma, J., & Sulistiawati. (2014). Teaching multiplication of numbers from 1 to 10 to STKIP Surya students using Matematika GASING. *Journal on Mathematics Education, 5*(1), 66–84. https://doi.org/10.22342/jme.5.1.1450.66-84

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), Contributions to probability and statistics (pp. 278–292). Stanford University Press.

Lynch, K., An, L., & Mancenido, Z. (2023). The impact of summer programs on student mathematics achievement: A meta-analysis. *Review of Educational Research, 93*(2), 275–315. https://doi.org/10.3102/00346543221105543

Mahadewsing, R., Getrouw, D., & Calor, S. M. (2024). Prior knowledge of a calculus course: The impact of prior knowledge on students' errors. *International Electronic Journal of Mathematics Education, 19*(3), em0786. https://doi.org/10.29333/iejme/14765

Mayerhofer, M., Lüftenegger, M., & Eichmair, M. (2024). The development of mathematics expectancy-value profiles during the secondary–tertiary transition into STEM fields. *International Journal of STEM Education, 11*(1), Article 31. https://doi.org/10.1186/s40594-024-00491-6

Mayerhofer, M., Lüftenegger, M., & Eichmair, M. (2025). Impact of a mathematics bridging course on the motivation and learning skills of university students. *International Journal of Research in Undergraduate Mathematics Education, 11*(1), 55–90. https://doi.org/10.1007/s40753-023-00224-0

McKinney, W. (2010). Data structures for statistical computing in Python. Proceedings of the 9th Python in Science Conference, 56–61.

Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 56–63.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*(1), 40–48.

Morán-Soto, G., Arellano Ulloa, L. H., & González Peña, O. I. (2026). Mathematics abilities as the key to successfully completing an engineering major: An analysis of a remedial mathematics course. *European Journal of Engineering Education*, Advance online publication. https://doi.org/10.1080/03043797.2025.2607635

Mosia, M., Egara, F. O., Nannim, F. A., & Basitere, M. (2025). Factors influencing students' performance in university mathematics courses: A structural equation modelling approach. *Education Sciences, 15*(2), 188. https://doi.org/10.3390/educsci15020188

Msomi, A. M., & Rzyankina, E. (2025). Bridging gaps: Enhancing holistic support in mathematics during the transition from secondary school to university. *Journal of Student Affairs in Africa, 12*(2), 51–70. https://doi.org/10.24085/jsaa.v12i2.5450

Navarrete-Ulloa, J. A. (2025). Learning rates: A correction of gain scores to assess math learning interventions. *The Journal of Experimental Education, 93*(3), 649–665. https://doi.org/10.1080/00220973.2024.2352768

Navarrete-Ulloa, J. A., Giaconi, V., & Contador, G. (2026). Normal gains: Estimators of learning rates in pretest-posttest settings. *Frontiers in Psychology, 17*, 1901493. https://doi.org/10.3389/fpsyg.2026.1901493

Nieminen, J. H., Reinholz, D. L., & Valero, P. (2024). Mathematics is a battle, but I've learned to survive: Becoming a disabled student in university mathematics. *Educational Studies in Mathematics, 116*, 5–25. https://doi.org/10.1007/s10649-024-10311-x

Nissen, J. M., Talbot, R. M., Nasim Thompson, A., & Van Dusen, B. (2018). Comparison of normalized gain and Cohen's d for analyzing gains on concept inventories. *Physical Review Physics Education Research, 14*(1), 010115.

OECD. (2023). PISA 2022 results (Volume I and II) — Country note: Indonesia. OECD Publishing.

Salehi, S., Cotner, S., & Ballen, C. J. (2020). Variation in incoming academic preparation: Consequences for minority and first-generation students. *Frontiers in Education, 5*, Article 552364. https://doi.org/10.3389/feduc.2020.552364

Samaloisa, M. Z., Irianto, D. M., & Sukardi, R. R. (2026). Pedagogical transformation through the GASING method in elementary mathematics: A systematic literature review from symbolic violence and critical pedagogy perspectives. *Journal of Educational Sciences, 10*(10), 966–989. https://doi.org/10.31258/jes.10.10.p.966-989

Schüler-Meyer, A. (2022). How transition students relearn school mathematics to construct multiply quantified statements. *Educational Studies in Mathematics, 110*, 291–311. https://doi.org/10.1007/s10649-021-10127-z

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. Proceedings of the 9th Python in Science Conference, 92–96.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3–4), 591–611.

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2022). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist, 57*(1), 31–54. https://doi.org/10.1080/00461520.2021.1939700

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360–407.

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., … Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences, 117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods, 17*, 261–272.

Waimbo, K. N., Daullu, M. A., & Ansanay, Y. O. (2026). The Matematika GASING method as an intervention for improving arithmetic fluency among school students in an under-resourced community context. *Jurnal Ilmu Pendidikan Indonesia, 14*(1), 70–84. https://doi.org/10.31957/jipi.v14i1.5368

Weber, T., Rathgeb-Schnierer, E., & Eichler, A. (2025). Development of prospective elementary teachers' beliefs about mathematics during their first semester at university, and the impact of inquiry-based tasks. *International Journal of Research in Undergraduate Mathematics Education, 11*, 146–161. https://doi.org/10.1007/s40753-023-00232-0

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika, 38*(3–4), 330–336.

Zakariya, Y. F., Berg, M. T., Gjesteland, T., Umahaba, R. E., & Abanikannda, M. O. (2025). From single to multiple assessments in a foundational mathematics course for engineering students: What do we gain? *Frontiers in Education, 10*, 1544647. https://doi.org/10.3389/feduc.2025.1544647