JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

CONSISTENCY OF POSTGRADUATE ACADEMIC PERFORMANCE AT MUET, JAMSHORO: A WELCH EQUIVALENCE-TESTING AND EXPLORATORY REGRESSION STUDY

Fabiha Shaikh[1], Shauban Ali Solangi[*2]

[1]*Department of Mathematics, Hyderabad Institute for Technology & Management Sciences (HITMS), Hyderabad, Sindh, Pakistan*
[*2]*Department of Computer Science & Related Studies, Hyderabad Institute for Technology & Management Sciences (HITMS), Hyderabad, Sindh, Pakistan*

[*2]dr.shauban@faculty.hitms.edu.pk

DOI: https://doi.org/10.5281/zenodo.23530061

Keywords
student performance; CGPA; Welch's t-test; equivalence testing; multiple linear regression; Applied Mathematics; MS/MBA

Article History
Received: 14 July 2026
Accepted: 05 October 2026
Published: 21 October 2026

Copyright @Author
Corresponding Author: * Shauban Ali Solangi

Abstract

Universities need evidence on whether the academic performance of successive postgraduate batches is stable and whether programmes differ, yet non-significant test results are often misread as proof of consistency. This study examines the CGPA records of 206 students at Mehran University of Engineering & Technology (MUET), Jamshoro: 53 in M.Phil Applied Mathematics (five batches) and 153 in MS/MBA at the Mehran University Institute of Science, Technology and Development (MUISTD; four batches), for the batches of 2014–2017. Batch means were compared with Welch's t-tests (Holm-adjusted), Welch's one-way ANOVA, Hedges' g with 95% confidence intervals, and two one-sided equivalence tests (TOST) with a ±0.30 CGPA margin; a sensitivity analysis excluded zero-CGPA records. Mean CGPA ranged from 3.45 to 3.72 (Applied Mathematics) and from 3.37 to 3.59 (MUISTD). No comparison between batches (Holm-adjusted $p \geq 0.95$ in Applied Mathematics and 1.00 in MUISTD) or between programmes (≥ 0.42) was significant, and the omnibus tests agreed ($p = 0.42$ and 0.61). However, equivalence within ±0.30 CGPA was demonstrated for only 4 of 20 comparisons (16 of 20 within ±0.50), so "no significant difference" cannot be equated with "consistency" at these sample sizes. Excluding zero-CGPA records raised the MUISTD batch means to 3.50–3.69 and brought the omnibus test close to significance ($p = 0.064$). A batch-level regression of on-time graduates on enrolment and CGPA fitted well in sample ($R^2 = 0.95$) but was dominated by two high-leverage batches (Cook's $D = 12.9$ and 18.3) and failed leave-one-out validation, so it is reported as exploratory. Student-level data are needed for confirmatory inference.

1. INTRODUCTION

Completing a postgraduate degree within the prescribed time, and with a sound cumulative grade point average (CGPA), is a key indicator of academic performance. Universities therefore need to know whether the performance of successive batches is stable and whether it differs between programmes, because the answer informs admission policy, supervision capacity and quality assurance. Regression analysis, a

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal
ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

statistical tool for quantifying the relationship between a response variable and one or more explanatory variables, has been widely used to predict students' academic performance [1, 2] and learning outcomes [3–6]. A regression with more than one explanatory variable is called multiple regression.

Recent systematic reviews show that this literature is mostly predictive and student-level: prior academic achievement, including CGPA, is the factor used most often, and most studies target performance at graduation or in a single course [7, 8]. Two questions that matter to programme managers are much less well served. The first is descriptive and inferential: is performance in a programme consistent from one batch to the next, and do programmes differ? The second concerns practice: "consistency" is routinely claimed from non-significant t-tests, although a non-significant result may simply reflect small batches and large variances. Both difficulties are acute for postgraduate programmes, where batches are small and records are often kept in aggregate form.

This paper studies two programmes of Mehran University of Engineering & Technology (MUET), Jamshoro—M.Phil Applied Mathematics and MS/MBA at the Mehran University Institute of Science, Technology and Development (MUISTD)—for the batches 2014–2017. Its objectives are (i) to describe CGPA in each batch; (ii) to test whether mean CGPA differs between batches within each programme and between the two programmes, using Welch's tests with multiplicity control; (iii) to test whether batches can be declared practically equivalent, using two one-sided tests (TOST); (iv) to examine the sensitivity of the conclusions to zero-CGPA records; and (v) to fit and critically diagnose a multiple linear regression model relating the number of students who obtained the degree on time to the number enrolled and their CGPA. The contribution is therefore as much inferential as empirical: we show how far batch-level data can, and cannot, support the claim that performance is consistent.

The paper is organised as follows. Section 2 reviews related work, Section 3 describes the data and methods, Section 4 presents the results, and Sections 5–7 discuss implications, limitations and conclusions.

2. Literature Review

2.1 Prediction and determinants of academic performance

Student success has been defined both narrowly, as academic achievement, and more broadly, as persistence to the degree [7]. Reviewing the predictive literature, Alyahyan and Düştegör [7] found that prior academic achievement—pre-university results and university grades such as the CGPA—was the factor category used most often (44% of the studies reviewed), ahead of student demographics (25%). Alwarthan et al. [8] likewise report that most earlier studies predicted performance at graduation time or at the level of a specific course. Educational data mining and learning analytics have since matured into established fields that apply classification, clustering and regression to institutional records [9]; an applied example is the work of Mueen et al. [10], who used Naïve Bayes, neural-network and decision-tree classifiers on students' academic records and forum participation.

Regression remains the workhorse where interpretability matters. Yang et al. [1] combined multiple linear regression with principal component analysis; Zacharis [3], Tempelaar et al. [4] and Huang and Fang [5] predicted outcomes in blended, data-rich and engineering courses; Yoo and Kim [6] predicted project performance from online discussion features; and Abro et al. [2] investigated student performance in mathematics. These studies rest on student-level observations, in contrast to the handful of batch-level aggregates available here.

2.2 Persistence, completion and time to degree

Persistence and completion are the second face of student success. Tinto's influential synthesis frames departure from higher education as the outcome of a student's academic and social integration into the institution [11]. At doctoral level, van de Schoot et al. [12] studied Dutch candidates and asked whether those who will experience delay can be predicted, noting that delay can end in attrition and institutional cost. These studies justify using the number of

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal
ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

students graduating within the stipulated duration (ND) as an outcome, but they also model individual candidates with covariates, which a batch-level data set cannot do.

2.3 Statistical considerations for comparing batches

Four statistical issues shape the design of this study. First, group comparison: Student's pooled-variance t-test assumes equal variances, an assumption that rarely holds. Welch's test [13] controls the Type I error rate better when variances differ and loses little when they do not, and has been recommended as the default [14]; the same logic favours Welch's one-way ANOVA for omnibus comparisons [15]. Second, interpretation: a p-value does not measure the size or importance of an effect [16], and absence of evidence of a difference is not evidence of absence [17]. Showing that two batches are similar requires an equivalence test, in which effects as large as a pre-specified smallest effect of interest are statistically rejected [18–20]; reporting effect sizes with confidence intervals conveys the same information more directly [21, 22], with standardised effects read against the conventions of [23]. Third, multiplicity: comparing every pair of batches inflates the family-wise error rate unless corrected, for instance by Holm's sequentially rejective procedure [24]. Fourth, small-sample regression: guidance on sample size for multiple regression is of the order of 50 + 8k observations for the overall test with k predictors [25], i.e. 66 for the two-predictor model used here, against four or five available batches.

Data quality matters as well. A zero in a grade-point field may be a genuine grade, a code for withdrawal, or a missing-data marker; treating it as an observation can distort means and variances, and the sensitivity of conclusions to such choices should be examined explicitly [26]. CGPA is also bounded and tends to cluster near the top of the scale, a pattern of grade compression documented in other systems [27], which further argues for effect-size and robustness reporting rather than reliance on p-values alone.

2.4 Research gap

To the authors' knowledge, little published work examines the batch-to-batch consistency of postgraduate CGPA within Pakistani programmes using procedures that distinguish "no detectable difference" from "practically equivalent", and fewer still report the diagnostics needed to judge small batch-level regressions. This paper addresses that gap.

3. Data and Methodology

3.1 Study design and data

The study is retrospective, quantitative, descriptive and inferential. CGPA records were collected for 53 students of M.Phil Applied Mathematics (batches 14, 15S, 15F, 16 and 17) and 153 students of MS/MBA at MUISTD (batches 14–17), for the batches of 2014–2017. Applied Mathematics batch 15 consists of two cohorts, labelled 15S and 15F. CGPA is reported on a 4.00 scale. For each Applied Mathematics batch, the number of students enrolled (NS) and the number who obtained the degree within the stipulated programme duration (ND) were also recorded.

All inferential statistics in Sections 4.3–4.5 were recomputed from the batch-level sample sizes, means and standard deviations in Table 1. For the classical pooled-variance t statistics this reproduces published values to within ±0.01, which confirms the inputs. Standard deviations are sample (n−1) standard deviations.

3.2 Descriptive statistics and roll-order trend lines

The number of students, mean, standard deviation (SD) and variance (VAR) of CGPA were computed for every batch. To check for systematic drift, CGPA was also plotted against the student serial number within each batch and a linear trend line $y = b_1x + b_0$ was fitted by least squares. The trend lines are descriptive: they indicate drift only to the extent that the roll order is not arbitrary, and no inferential test is attached to them.

3.3 Welch tests, effect sizes and multiplicity control

Mean CGPA of two batches (or of the same batch in the two programmes) was compared

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

with Welch's two-sided t-test at the 5% level. With sample sizes n_1, n_2, means and standard deviations s_1, s_2, the statistic and the Welch–Satterthwaite degrees of freedom are

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (1)$$

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1-1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2-1)}} \quad (2)$$

where $d = \bar{x_1} - \bar{x_2}$ is the difference in means; its 95% confidence interval is $d \pm t_{0.975,\nu} \times SE$. Standardised differences are reported as Hedges' g [22],

$$g = \left(1 - \frac{3}{4(n_1+n_2)-9}\right)\frac{d}{s_p} \quad (3)$$

where s_p is the pooled standard deviation. Because every pair of batches within a programme is compared, Holm's step-down adjustment [24] was applied within each family of comparisons (10 pairs for Applied Mathematics, 6 for MUISTD and 4 batch-wise comparisons between programmes). For the omnibus hypothesis of equal batch means within a programme, Welch's one-way ANOVA [15] was used, with the classical F test and η^2 reported for reference.

3.4 Equivalence testing

Equivalence was assessed with the two one-sided tests (TOST) procedure [18–20]. For a margin Δ, the null hypothesis $|\mu_1 - \mu_2| \geq \Delta$ is rejected when both one-sided Welch tests

$$t_L = \frac{d+\Delta}{SE}, \quad t_U = \frac{d-\Delta}{SE} \quad (4)$$

are significant, i.e. when $p_{TOST} = \max(p_L, p_U) < 0.05$, which is equivalent to the 90% confidence interval of d lying wholly inside $(-\Delta, +\Delta)$. The margin was set at $\Delta = 0.30$ CGPA points, roughly the step between adjacent plus/minus grades on a 4.00 scale and about one-tenth of the scale, with ±0.50 as a lenient alternative. Any margin is a judgement and should be fixed before data are seen [19, 20]; here both are reported so that readers can apply their own. The minimum difference detectable with 80% power (two-sided $\alpha = 0.05$) is also reported to show what the batch sizes can resolve.

3.5 Sensitivity to zero-CGPA records

The MUISTD records contain zero CGPA entries (one each in batches 14 and 16, two in batch 15, none in batch 17). The records do not state whether these are genuine grades, withdrawals or missing-data codes. The primary analyses retain them; every analysis was repeated after removing them [26]. Because a zero adds nothing to the sum or to the sum of squares, the reduced-sample means and standard deviations follow exactly from Table 1.

3.6 Regression model and diagnostics

For Applied Mathematics (five batches) and MUISTD (four batches) a multiple linear regression of the batch-level number of on-time graduates on enrolment and mean CGPA was fitted,

$$ND = b_0 + b_1 NS + b_2 CGPA + \varepsilon \quad (5)$$

where ND is the number of students who obtained the degree on time, NS the number of students enrolled, and CGPA the batch mean. Goodness of fit was summarised by

$$R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-k-1} \quad (6)$$

where $k = 2$ is the number of independent variables and n the number of batches. Because n is far below the sample size recommended for regression [25], the models are treated as exploratory and were diagnosed with leverage h_{ii} (the diagonal of the hat matrix), Cook's distance

$$D_i = \frac{e_i^2 h_{ii}}{p s^2 (1-h_{ii})^2} \quad (7)$$

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

(e_i the residual, p = 3 the number of parameters, s^2 the residual variance) and leave-one-out (LOO) prediction error. Since ND cannot exceed NS, an on-time rate model ND/NS = a + b CGPA was also fitted to the Applied Mathematics batches, with Pearson and Spearman correlations. Welch, TOST, ANOVA and diagnostic computations were carried out in Python (NumPy, SciPy).

4. Results and Discussion

4.1 Descriptive statistics

Table 1. Number of students, mean, standard deviation (SD) and variance (VAR) of CGPA, batches 2014–2017.

Programme / batch	n	Mean	SD	VAR
M.Phil Applied Mathematics				
Batch 14	8	3.6738	0.2986	0.0892
Batch 15S	8	3.7212	0.3861	0.1491
Batch 15F	10	3.6780	0.2133	0.0455
Batch 16	20	3.5835	0.2707	0.0733
Batch 17	7	3.4486	0.2798	0.0783
MS/MBA (MUISTD)				
Batch 14	26	3.3658	0.7734	0.5982
Batch 15	35	3.4620	0.9039	0.8169
Batch 16	39	3.5926	0.6652	0.4425
Batch 17	53	3.5504	0.3552	0.1262

Zero CGPA entries in the source records (MUISTD: one each in batches 14 and 16, two in batch 15) are retained in Table 1 and in the primary analyses; Section 4.5 reports the results without them.

In Applied Mathematics (Table 1, Fig. 1a) the maximum mean (3.7212), standard deviation (0.3861) and variance (0.1491) all occurred in batch 15S, and the lowest mean in batch 17 (3.4486). In MUISTD (Fig. 1b) the maximum mean (3.5926) occurred in batch 16, whereas the maximum standard deviation (0.9039) and variance (0.8169) occurred in batch 15. The large spread of MUISTD batches 14–16 is mainly due to the zero entries (Section 4.5); batch 17, which has none, has a much smaller spread (SD = 0.3552).

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

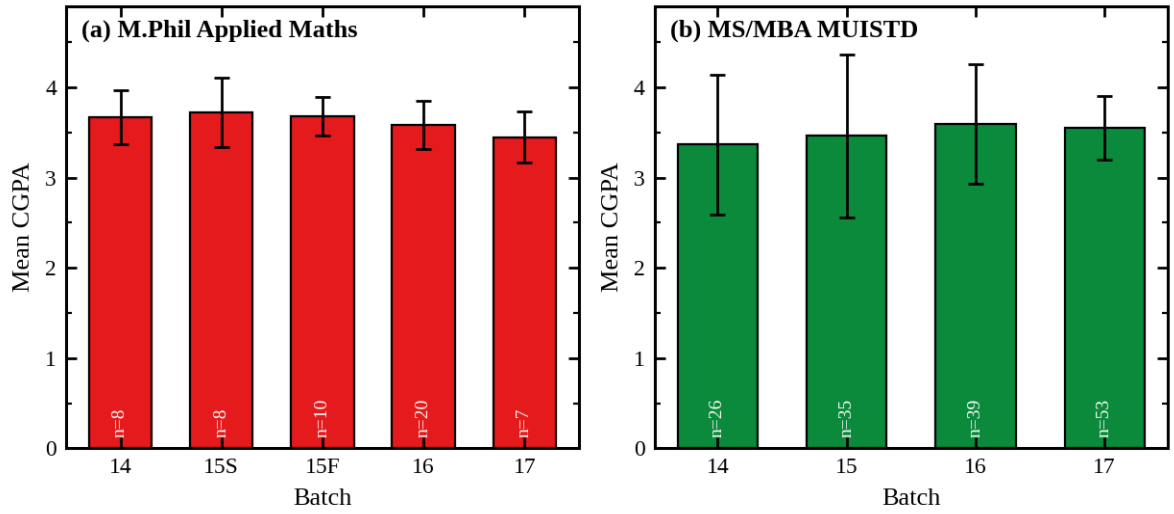


Fig. 1. Mean CGPA ± one standard deviation for each batch: (a) M.Phil Applied Mathematics and (b) MS/MBA (MUISTD). The number of students is given inside the bars.

4.2 Student-wise CGPA graphs and trend lines

The student-wise CGPA and the fitted trend lines are shown in Figs. 2 and 3 and the trend equations in Table 2. Every slope is small ($|b_1| \leq 0.0263$), so CGPA shows no appreciable drift along the roll sequence. In Applied Mathematics the slope is positive for batches 14 (0.0132) and 15S (0.0263, the largest) and negative for batches 15F, 16 and 17; in MUISTD all four slopes are negative. As noted in Section 3.2, these lines are descriptive only.

Table 2. Trend-line equations of CGPA (y) against student serial number (x) for each batch.

Programme / batch	Trend line
M.Phil Applied Mathematics	
Batch 14	y = 0.0132x + 3.6143
Batch 15S	y = 0.0263x + 3.6029
Batch 15F	y = −0.0166x + 3.7693
Batch 16	y = −0.0141x + 3.7318
Batch 17	y = −0.0093x + 3.4857
MS/MBA (MUISTD)	
Batch 14	y = −0.0175x + 3.6019
Batch 15	y = −0.0070x + 3.5878
Batch 16	y = −0.0015x + 3.6225
Batch 17	y = −0.0074x + 3.7503

Trend lines are descriptive; see Section 3.2.

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

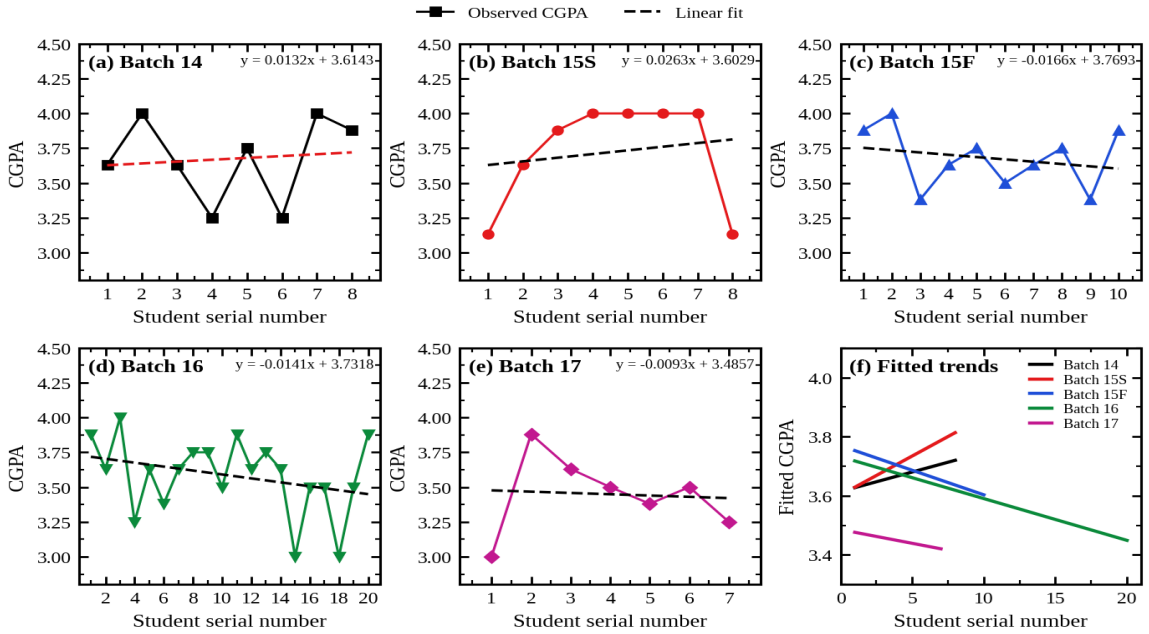


Fig. 2. M.Phil Applied Mathematics: student CGPA with linear trend lines for (a) batch 14, (b) batch 15S, (c) batch 15F, (d) batch 16 and (e) batch 17; (f) comparison of the fitted trends.

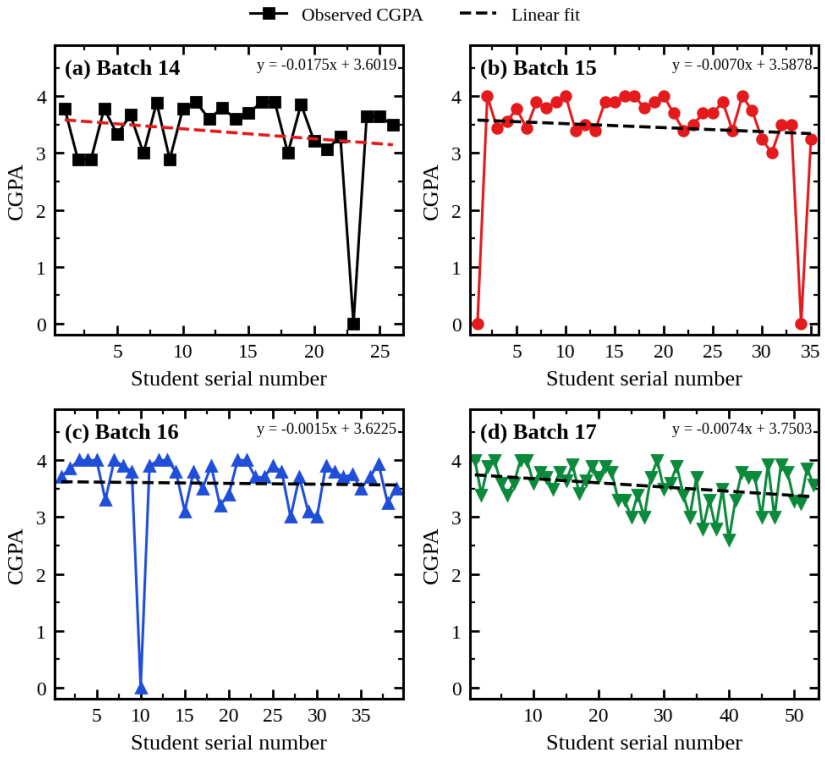


Fig. 3. MS/MBA (MUISTD): student CGPA with linear trend lines for (a) batch 14, (b) batch 15, (c) batch 16 and (d) batch 17.

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

4.3 Consistency across batches within programmes

Table 3. Welch tests, effect sizes and equivalence tests (±0.30 CGPA) for the difference between batch means within programmes.

Comparison	Δ mean	95% CI	Welch t (df)	p	p_{Holm}	g	p_{TOST}
M.Phil Applied Mathematics							
14–15S	−0.05	[−0.42, +0.32]	−0.27 (13.2)	0.788	1.000	−0.13	0.083
14–15F	0.00	[−0.28, +0.27]	−0.03 (12.3)	0.974	1.000	−0.02	**0.018**
14–16	+0.09	[−0.18, +0.36]	0.74 (11.9)	0.472	1.000	+0.31	0.055
14–17	+0.23	[−0.10, +0.55]	1.51 (12.9)	0.156	1.000	+0.73	0.313
15S–15F	+0.04	[−0.29, +0.38]	0.28 (10.4)	0.782	1.000	+0.14	0.061
15S–16	+0.14	[−0.20, +0.47]	0.92 (9.9)	0.378	1.000	+0.44	0.151
15S–17	+0.27	[−0.10, +0.65]	1.58 (12.6)	0.139	1.000	+0.75	0.438
15F–16	+0.09	[−0.09, +0.28]	1.04 (22.4)	0.308	1.000	+0.36	**0.017**
15F–17	+0.23	[−0.05, +0.51]	1.83 (10.7)	0.095	0.954	+0.90	0.293
16–17	+0.13	[−0.14, +0.41]	1.11 (10.2)	0.294	1.000	+0.48	0.102
MS/MBA (MUISTD), all records							
14–15	−0.10	[−0.53, +0.33]	−0.45 (57.8)	0.657	1.000	−0.11	0.174
14–16	−0.23	[−0.60, +0.15]	−1.22 (48.1)	0.227	1.000	−0.32	0.347
14–17	−0.18	[−0.51, +0.14]	−1.16 (30.3)	0.256	1.000	−0.35	0.237
15–16	−0.13	[−0.50, +0.24]	−0.70 (62.0)	0.486	1.000	−0.16	0.183
15–17	−0.09	[−0.41, +0.24]	−0.55 (41.0)	0.585	1.000	−0.14	0.097
16–17	+0.04	[−0.19, +0.28]	0.36 (53.9)	0.720	1.000	+0.08	**0.016**

Δ mean = first minus second batch; p_{Holm} = Holm-adjusted p within each family (10 Applied Mathematics comparisons, 6 MUISTD comparisons); g = Hedges' g; p_{TOST} = larger of the two one-sided Welch p values for the margin ±0.30 CGPA (bold: equivalence demonstrated at $\alpha = 0.05$). MUISTD results retain the zero-CGPA records. Statistics recomputed from Table 1.

Table 3 and Fig. 4 give the Welch tests for every pair of batches. None of the 16 comparisons is significant: unadjusted p ranges from 0.095 to 0.974 in Applied Mathematics and from 0.227 to 0.720 in MUISTD, and after Holm adjustment the smallest p is 0.954. The Welch decisions coincide with those of classical pooled-variance tests, although the statistics differ where variances are unequal (for example MUISTD batches 14 and 17: $|t| = 1.16$ with Welch against 1.46 pooled). The omnibus tests agree (Table 4): Welch's $W = 1.02$ ($p = 0.42$) for Applied Mathematics and 0.61 ($p = 0.61$) for MUISTD.

These null results must be read with their precision in mind. The batches are small (7 to 20 students in Applied Mathematics), so the minimum difference detectable with 80% power is 0.27–0.52 CGPA points in Applied Mathematics and 0.33–0.61 in MUISTD. Batch 17 of Applied Mathematics ($n = 7$) has the lowest mean (3.4486), 0.23–0.27 points below batches 14, 15S and 15F, a standardised difference of $g = 0.73$–0.90; such differences are not statistically significant here, but they are large in standardised terms and imprecisely estimated, not evidence of sameness. Formally, equivalence within ±0.30 CGPA is demonstrated for only 3 of the 16 comparisons (14–15F and 15F–16 in Applied Mathematics; 16–17 in MUISTD), while equivalence within ±0.50 is demonstrated for 14 of 16 (9 of 10 and 5 of 6). Batch means are therefore compatible with consistency and any batch-to-batch difference is probably under half a CGPA point, but the data do not establish tighter consistency.

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

Table 4. Omnibus tests of equal batch means within each programme.

Sample	Welch W (df_1, df_2)	p	Classical F (df_1, df_2)	p	η^2
M.Phil Applied Mathematics	1.02 (4, 18.4)	0.422	1.13 (4, 48)	0.354	0.086
MS/MBA (MUISTD), all records	0.61 (3, 62.2)	0.612	0.73 (3, 149)	0.535	0.015
MS/MBA (MUISTD), zero-CGPA records excluded	2.52 (3, 70.9)	0.064	2.58 (3, 145)	0.056	0.051

Computed from the batch sample sizes, means and standard deviations. η^2 is the classical proportion of variance between batches.

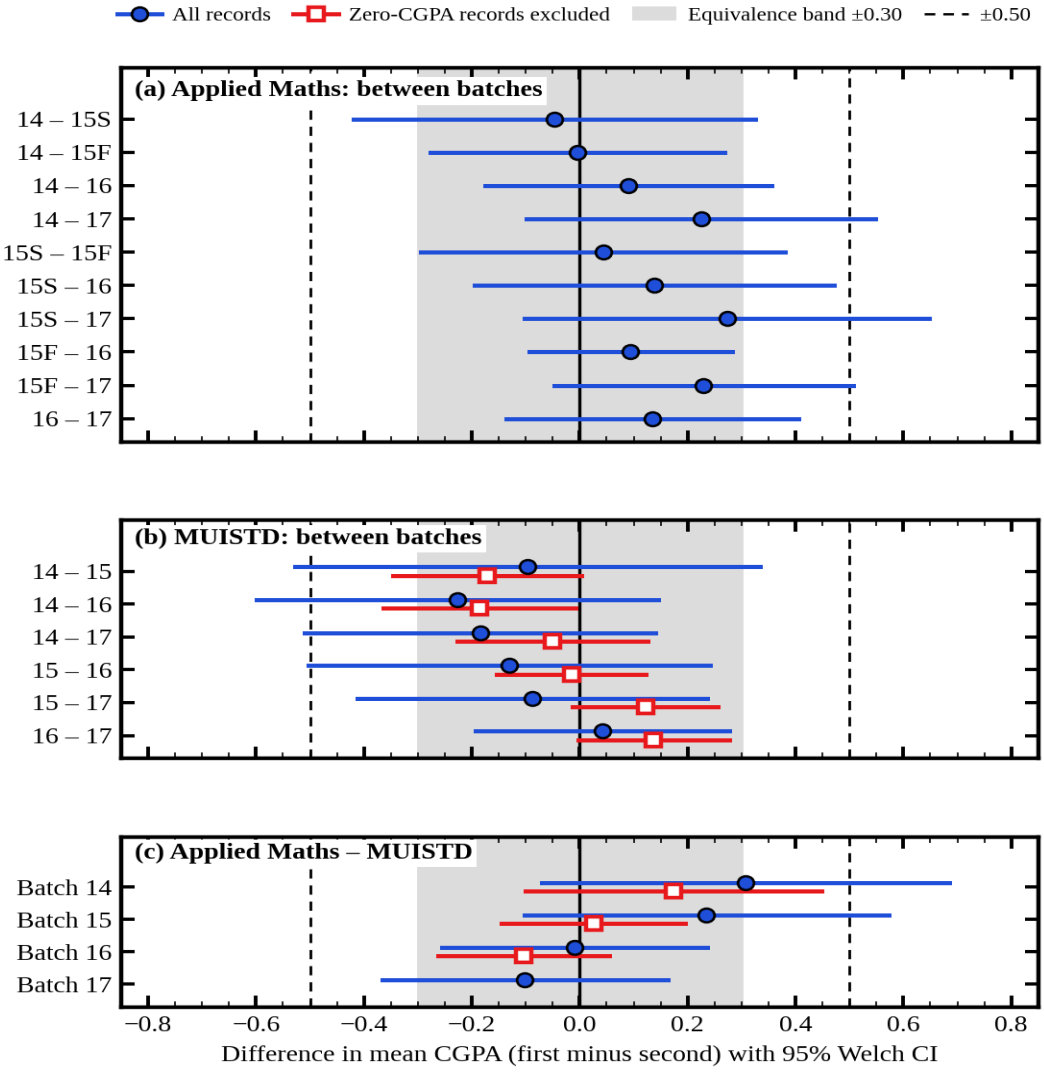


Fig. 4. Differences in mean CGPA with 95% Welch confidence intervals: (a) between batches of Applied Mathematics, (b) between batches of MUISTD and (c) between the two programmes within each batch. Blue circles: all records; red squares: MUISTD zero-CGPA records excluded (omitted where unchanged). The grey band marks the equivalence margin ±0.30 CGPA and dashed lines ±0.50. An interval inside a band is sufficient for equivalence at that margin; TOST itself uses the 90% interval.

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

4.4 Consistency between the two programmes

Table 5. Welch tests, effect sizes and equivalence tests (±0.30 CGPA) for the difference in mean CGPA between Applied Mathematics and MUISTD, batch by batch.

Batch	n (AM / MU)	Mean (AM / MU)	Δ mean [95% CI]	Welch t (df)	p (p_{Holm})	g	p_{TOST}
All records							
Batch 14	8 / 26	3.6738 / 3.3658	+0.31 [−0.07, +0.69]	1.67 (30.0)	0.106 (0.424)	+0.43	0.517
Batch 15	18 / 35	3.6972 / 3.4620	+0.24 [−0.10, +0.57]	1.40 (45.5)	0.168 (0.503)	+0.31	0.350
Batch 16	20 / 39	3.5835 / 3.5926	−0.01 [−0.25, +0.24]	−0.07 (55.0)	0.941 (0.941)	−0.02	**0.011**
Batch 17	7 / 53	3.4486 / 3.5504	−0.10 [−0.37, +0.16]	−0.87 (8.8)	0.405 (0.811)	−0.29	0.062
MUISTD zero-CGPA records excluded							
Batch 14	8 / 25	3.6738 / 3.5004	+0.17 [−0.10, +0.45]	1.35 (14.3)	0.197 (0.780)	+0.48	0.170
Batch 15	18 / 33	3.6972 / 3.6718	+0.03 [−0.14, +0.20]	0.30 (32.8)	0.764 (0.811)	+0.09	**0.001**
Batch 16	20 / 38	3.5835 / 3.6871	−0.10 [−0.26, +0.06]	−1.32 (43.7)	0.195 (0.780)	−0.34	**0.008**
Batch 17	7 / 53	3.4486 / 3.5504	−0.10 [−0.37, +0.16]	−0.87 (8.8)	0.405 (0.811)	−0.29	0.062

AM = Applied Mathematics; MU = MUISTD. Batch 15 of Applied Mathematics combines cohorts 15S and 15F (n = 18, mean 3.6972, SD 0.2932). Δ mean = AM minus MU. Holm adjustment is applied within each set of four batch-wise comparisons. Bold p_{TOST}: equivalence within ±0.30 demonstrated.

Applied Mathematics students had a higher mean CGPA than MUISTD students in batches 14 and 15 (by 0.31 and 0.24 points) and a very similar mean in batches 16 and 17 (differences of −0.01 and −0.10; Table 5, Fig. 4c). No difference is significant (Holm-adjusted $p \geq 0.42$), and the effect sizes are small to moderate ($|g| \leq 0.43$). Welch's test matters here: for batch 14 the Welch statistic (1.67) differs markedly from the pooled value (1.09) because the MUISTD variance is about seven times larger. Equivalence within ±0.30 is shown only for batch 16 (p_{TOST} = 0.011) and within ±0.50 for batches 16 and 17. When the zero-CGPA records are excluded, the mean differences become +0.17, +0.03, −0.10 and −0.10 for batches 14–17 and equivalence within ±0.30 is shown for batches 15 and 16. The programmes therefore show no detectable difference in mean CGPA, but equivalence cannot be claimed for every batch.

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal
ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

4.5 Sensitivity to zero-CGPA records

Table 6. MUISTD batch statistics with all records and with zero-CGPA records excluded.

Batch	n	Mean	SD	n	Mean	SD
MS/MBA (MUISTD)						
Batch 14	26	3.3658	0.7734	25	3.5004	0.3635
Batch 15	35	3.4620	0.9039	33	3.6718	0.2713
Batch 16	39	3.5926	0.6652	38	3.6871	0.3105
Batch 17	53	3.5504	0.3552	53	3.5504	0.3552

Left block: all records; right block: zero-CGPA records excluded. Reduced-sample statistics are exact functions of Table 1.

Removing the zero entries has a large effect on MUISTD (Table 6). The batch standard deviations fall from 0.67–0.90 to 0.27–0.36, and the means of batches 14, 15 and 16 rise to 3.50, 3.67 and 3.69. The conclusion of consistency across MUISTD batches then weakens: batch 14 becomes the lowest-scoring batch, 0.17 and 0.19 points below batches 15 and 16 ($g \approx 0.55$; unadjusted $p = 0.055$ and 0.040), and the omnibus Welch test gives $W = 2.52$ ($p = 0.064$; Table 4). None of the pairwise differences survives Holm adjustment (smallest adjusted $p = 0.24$), so the evidence is suggestive rather than conclusive, but the reduction in variance also improves precision: the minimum detectable difference falls to 0.19–0.25 CGPA points, and equivalence within ±0.30 is demonstrated for four of the six pairs. Whether the zero entries represent genuine failures, withdrawals or missing codes cannot be determined from the records, and the substantive conclusion for MUISTD depends on that answer.

4.6 Multiple linear regression

The fitted equations are ND = −59.12 + 0.73 NS + 16.29 CGPA for Applied Mathematics and ND = 242.50 − 0.18 NS − 63.72 CGPA for MUISTD. The input data of the Applied Mathematics model are given in Table 7 and the regression statistics, ANOVA and coefficients in Tables 8 and 9.

Table 7. Input data of the Applied Mathematics regression: on-time graduates (ND), students enrolled (NS), batch-mean CGPA, and on-time rate ND/NS.

Batch	ND	NS	CGPA	ND/NS
Batch 14	8	8	3.67	1.00
Batch 15S	7	8	3.72	0.88
Batch 15F	7	10	3.68	0.70
Batch 16	14	20	3.58	0.70
Batch 17	2	7	3.45	0.29

The MUISTD row is incomplete in the source data; the MUISTD model below is reported as fitted but cannot be diagnosed until the inputs are supplied.

Table 8. Regression statistics and analysis of variance.

Statistic	Applied Mathematics	MUISTD
Multiple R	0.97	0.84
R^2	0.95	0.71
Adjusted R^2	0.90	0.14
Standard error	1.36	8.69

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal
ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

Statistic	Applied Mathematics	MUISTD
Observations	5	4
Regression SS (d.f.)	69.51 (2)	188.42 (2)
Residual SS (d.f.)	3.69 (2)	75.58 (1)
F	18.86	1.25
Significance F	0.050	0.54

Applied Mathematics: Significance F is 0.0503 before rounding.

Table 9. Estimated regression coefficients with 95% confidence intervals.

Term	Coefficient	Standard error	t	p-value	Lower 95%	Upper 95%
M.Phil Applied Mathematics						
Intercept	−59.12	23.00	−2.57	0.12	−158.07	39.83
NS	0.73	0.13	5.77	0.03	0.19	1.28
CGPA	16.29	6.31	2.58	0.12	−10.86	43.44
MS/MBA (MUISTD)						
Intercept	242.50	257.42	0.94	0.52	−3028.38	3513.37
NS	−0.18	0.69	−0.26	0.84	−8.98	8.62
CGPA	−63.72	79.33	−0.80	0.57	−1071.73	944.30

For Applied Mathematics, R = 0.97 and R^2 = 0.95: 95% of the variation in ND is explained by NS and CGPA (adjusted R^2 = 0.90; standard error 1.36). The overall F-test (F = 18.86, 2 and 2 d.f.) has a significance level of 0.05 (0.0503 before rounding), so the model is only marginally significant. NS is individually significant (b_1 = 0.73, t = 5.77, p = 0.03, 95% CI 0.19–1.28), whereas CGPA is not (p = 0.12) and its confidence interval includes zero (Fig. 5a).

For MUISTD, R^2 = 0.71 but the adjusted R^2 drops to 0.14 because only four observations are available for three parameters; the model is not significant (F = 1.25, significance F = 0.54) and neither NS (p = 0.84) nor CGPA (p = 0.57) is a significant predictor (Fig. 5c). The poor adjusted fit of the MUISTD model, together with its very wide confidence intervals, means that its coefficients cannot be interpreted.

Table 10. Influence diagnostics of the Applied Mathematics regression.

Batch	ND	NS	CGPA	Fitted ND	Leverage h_{ii}	Cook's D	LOO error
Batch 14	8	8	3.67	6.6	0.31	0.26	2.1
Batch 15S	7	8	3.72	7.3	0.46	0.04	−0.6
Batch 15F	7	10	3.68	8.1	0.27	0.12	−1.5
Batch 16	14	20	3.58	13.8	0.98	12.90	8.2
Batch 17	2	7	3.45	2.2	0.98	18.32	−9.8

Fitted values, leverage and Cook's distance are computed from the unrounded batch means of Table 1, so fitted values differ slightly from those implied by the rounded coefficients of Table 9. LOO error = observed minus leave-one-out prediction (students). The average leverage is 3/5 = 0.60.

These in-sample figures should not be taken at face value. With five observations and three parameters the Applied Mathematics model has only two residual degrees of freedom, and NS is

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

mechanically related to the response (ND cannot exceed NS; r(ND, NS) = 0.88). Influence diagnostics (Table 10, Fig. 5b) show that batches 16 and 17 carry almost all the leverage (h_{ii} = 0.98 each, against a mean of 0.60) and have Cook's distances of 12.9 and 18.3, far above the usual threshold of 1. Batch 16 has by far the largest enrolment (NS = 20) and batch 17 the lowest ND and CGPA; when batch 16 is omitted the NS coefficient falls from 0.73 to 0.06, so the "significant" enrolment effect is carried by a single batch. Leave-one-out validation confirms the instability: the LOO root-mean-square error is 5.83 students, more than four times the in-sample standard error (1.36) and larger than that of a mean-only model (4.78).

Because ND is bounded by NS, the on-time rate ND/NS is the more natural response. Across the five Applied Mathematics batches it fell from 1.00 (batch 14) to 0.29 (batch 17) (Table 7) and was positively related to batch-mean CGPA (rate = −7.33 + 2.22 CGPA; r = 0.89, R^2 = 0.80, p = 0.041), although the rank correlation is weaker and not significant (Spearman ρ = 0.62, p = 0.27; n = 5). The association between CGPA and on-time completion is therefore a hypothesis for student-level testing, not a finding. Influence diagnostics for the MUISTD model can be added once its batch-level inputs are inserted in Table 7.

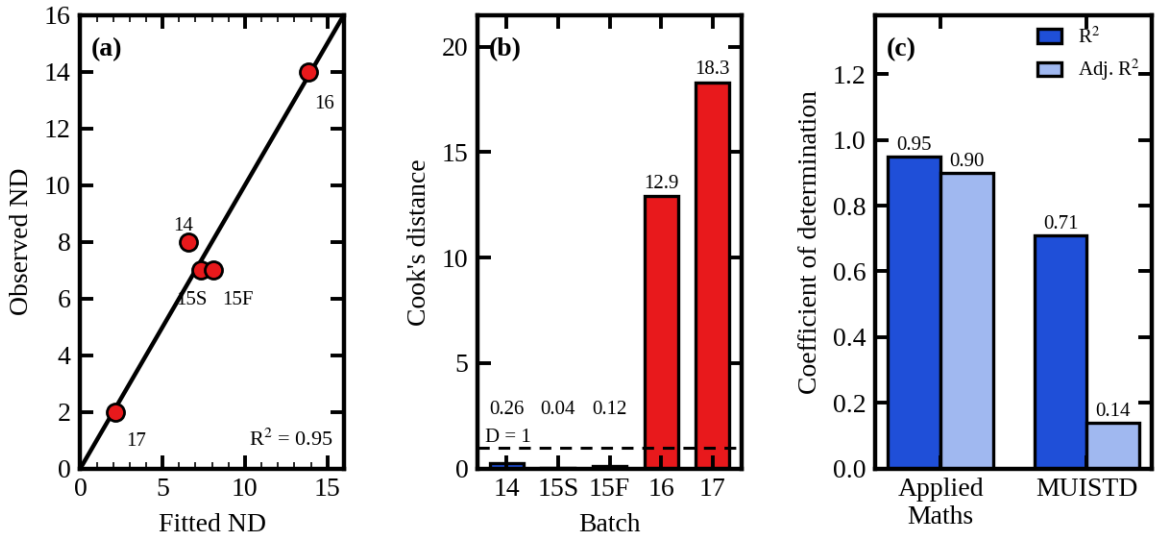


Fig. 5. (a) Observed versus fitted number of on-time graduates (ND) for Applied Mathematics; the solid line is the 1:1 line and labels denote the batch. (b) Cook's distance by batch (dashed line: D = 1). (c) R^2 and adjusted R^2 of the Applied Mathematics and MUISTD models.

5. Discussion

The central empirical result is a negative one with a positive methodological message. In both programmes, no comparison of batch means, within or between programmes, reached significance, and omnibus Welch tests agreed. Read naively, that supports the statement that performance is consistent. The equivalence tests, effect sizes and detectable-difference calculations show why that reading is premature: with batches of 7 to 53 students and standard deviations of 0.2–0.9, the data largely exclude differences of about half a CGPA point but cannot exclude differences of 0.3. This distinction between failing to reject and demonstrating equivalence [17, 19, 20] is the main lesson for programme reviews that rely on non-significant t-tests.

The sensitivity analysis shows that data handling can matter as much as the choice of test. Zero entries inflated the MUISTD standard deviations roughly two- to three-fold, concealed a modest dip in batch 14 and, through the larger variances, reduced the power of every MUISTD comparison. Recording a status code for withdrawn, failed or deferred students, rather

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

than a zero, would remove this ambiguity [26]. The regression results illustrate a related point. A high R^2 from five batches is easily obtained and easily over-read; influence and leave-one-out diagnostics show that the Applied Mathematics model is driven by two batches and does not predict held-out batches. This agrees with the predictive literature, where models are built and validated on student-level records [7, 8, 10], and with studies of completion and delay, where individual covariates are essential [12].

Several practical implications follow. First, programme-quality reports should present effect sizes with confidence intervals and, where consistency is the claim, equivalence tests against a pre-specified margin [19, 21]. Second, completion status should be tracked for each student, so that on-time rates can be modelled with logistic or survival methods rather than from batch totals. Third, the lower CGPA and low on-time rate of Applied Mathematics batch 17 (2 of 7 on time) merit follow-up, bearing in mind that a recent batch may simply not yet have had time to complete.

6. Limitations and Future Work

The study covers two programmes at one university and nine batches, so its findings are descriptive of these cohorts and should not be generalised. The regression models rest on four or five batch-level observations, far below recommended sample sizes [25]. Inference was carried out from batch summary statistics; normality of CGPA within batches, which is doubtful for a bounded and compressed variable [27], could not be assessed, and rank-based checks (Mann–Whitney and Kruskal–Wallis tests) on the raw records are advisable. The meaning of the zero-CGPA entries is unknown, the roll-order trend lines are interpretable only if the order is informative, and the completion status of the most recent batches may be right-censored by the date of data extraction. The M.Phil and MS/MBA programmes differ in curriculum and structure, so between-programme comparisons are of CGPA outcomes, not of teaching quality. Entry scores, gender, funding and other student-level covariates were unavailable. Future work should use student-level records from more programmes and batches, model on-time completion with logistic, survival or mixed-effects models, and fix equivalence margins in advance.

7. Conclusion

This study presented the CGPA records of 206 students from M.Phil Applied Mathematics and MS/MBA (MUISTD), batches 2014–2017, which were analysed with Welch and equivalence methods and a diagnosed regression. The main findings are that the Mean CGPA ranged from 3.45 to 3.72 in Applied Mathematics and from 3.37 to 3.59 in MUISTD; batch-wise trend slopes were very small ($|b_1| \leq 0.0263$), indicating no systematic drift along the roll sequence (a descriptive finding only). No difference between batches or programmes was statistically significant (Welch, Holm-adjusted $p \geq 0.42$; omnibus $p = 0.42$ and 0.61), but equivalence within ±0.30 CGPA was demonstrated for only 4 of 20 comparisons (16 of 20 within ±0.50). The data are compatible with consistency but cannot establish it, especially for small batches such as Applied Mathematics batch 17. Moreover, the MUISTD conclusions depend on zero-CGPA records: excluding them lowered batch standard deviations from 0.67–0.90 to 0.27–0.36 and brought the omnibus test close to significance ($p = 0.064$). The nature of these records should be verified. Also, the regression models are exploratory as the Applied Mathematics model ($R^2 = 0.95$) is driven by two high-leverage batches (Cook's D = 12.9 and 18.3) and fails leave-one-out validation (RMSE 5.83), and the MUISTD model is not significant (adjusted $R^2 = 0.14$). The positive association between batch CGPA and on-time rate ($r = 0.89$, $n = 5$) is a hypothesis for student-level testing. Therefore, the programme reviews should report effect sizes, confidence intervals and equivalence tests, and universities should keep student-level completion and status records so that consistency of postgraduate performance can be tested with adequate power.

Declarations

Author contributions: Fabiha Shaikh: Idea, Methodology, Validation, Editing. **Shauban Ali Solangi:** Conceptualization, Supervision,

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

Methodology, Original draft preparation, Data curation, Software, Writing - Investigation & editing.

Funding: No funding is received for this research.

Competing interests: There is no conflict of interest among the authors.

Data availability: The data will be available upon reasonable request from the corresponding author.

References

S. J. H. Yang, O. H. T. Lu, A. Y. Q. Huang, J. C. H. Huang, H. Ogata and A. J. Q. Lin, "Predicting students' academic performance using multiple linear regression and principal component analysis," Journal of Information Processing, vol. 26, pp. 170–176, 2018.

S. Abro, M. Solangi and A. Shaikh, "An investigation on the performance of students in mathematics: A case study," Sindh University Research Journal (Science Series), vol. 48, no. 2, 2016.

N. Z. Zacharis, "A multivariate approach to predicting student outcomes in web-enabled blended learning courses," The Internet and Higher Education, vol. 27, pp. 44–53, 2015.

D. T. Tempelaar, B. Rienties and B. Giesbers, "In search for the most informative data for feedback generation: Learning analytics in a data-rich context," Computers in Human Behavior, vol. 47, pp. 157–167, 2015.

S. Huang and N. Fang, "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models," Computers & Education, vol. 61, pp. 133–145, 2013.

J. Yoo and J. Kim, "Predicting learner's project performance with dialogue features in online Q&A discussions," in Proc. Int. Conf. on Intelligent Tutoring Systems, Springer, Berlin, Heidelberg, Jun. 2012, pp. 570–575.

E. Alyahyan and D. Düştegör, "Predicting academic success in higher education: Literature review and best practices," International Journal of Educational Technology in Higher Education, vol. 17, Art. no. 3, 2020, doi: 10.1186/s41239-020-0177-7.

S. A. Alwarthan, N. Aslam and I. U. Khan, "Predicting student academic performance at higher education using data mining: A systematic review," Applied Computational Intelligence and Soft Computing, vol. 2022, Art. no. 8924028, 2022, doi: 10.1155/2022/8924028.

C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," WIREs Data Mining and Knowledge Discovery, vol. 10, no. 3, Art. no. e1355, 2020.

A. Mueen, B. Zafar and U. Manzoor, "Modeling and predicting students' academic performance using data mining techniques," International Journal of Modern Education and Computer Science, vol. 8, no. 11, pp. 36–42, 2016.

V. Tinto, "Dropout from higher education: A theoretical synthesis of recent research," Review of Educational Research, vol. 45, no. 1, pp. 89–125, 1975.

R. van de Schoot, M. A. Yerkes, J. M. Mouw and H. Sonneveld, "What took them so long? Explaining PhD delays among doctoral candidates," PLoS ONE, vol. 8, no. 7, Art. no. e68839, 2013.

B. L. Welch, "The generalization of 'Student's' problem when several different population variances are involved," Biometrika, vol. 34, no. 1–2, pp. 28–35, 1947.

M. Delacre, D. Lakens and C. Leys, "Why psychologists should by default use Welch's t-test instead of Student's t-test," International Review of Social Psychology, vol. 30, no. 1, pp. 92–101, 2017.

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 6 • October 2026

M. Delacre, C. Leys, Y. L. Mora and D. Lakens, "Taking parametric assumptions seriously: Arguments for the use of Welch's F-test instead of the classical F-test in one-way ANOVA," International Review of Social Psychology, vol. 32, no. 1, Art. no. 13, 2019.

R. L. Wasserstein and N. A. Lazar, "The ASA statement on p-values: Context, process, and purpose," The American Statistician, vol. 70, no. 2, pp. 129–133, 2016.

D. G. Altman and J. M. Bland, "Absence of evidence is not evidence of absence," BMJ, vol. 311, no. 7003, p. 485, 1995.

D. J. Schuirmann, "A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability," Journal of Pharmacokinetics and Biopharmaceutics, vol. 15, no. 6, pp. 657–680, 1987.

D. Lakens, "Equivalence tests: A practical primer for t tests, correlations, and meta-analyses," Social Psychological and Personality Science, vol. 8, no. 4, pp. 355–362, 2017.

D. Lakens, A. M. Scheel and P. M. Isager, "Equivalence testing for psychological research: A tutorial," Advances in Methods and Practices in Psychological Science, vol. 1, no. 2, pp. 259–269, 2018.

G. Cumming, "The new statistics: Why and how," Psychological Science, vol. 25, no. 1, pp. 7–29, 2014.

L. V. Hedges, "Distribution theory for Glass's estimator of effect size and related estimators," Journal of Educational Statistics, vol. 6, no. 2, pp. 107–128, 1981.

J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1988.

S. Holm, "A simple sequentially rejective multiple test procedure," Scandinavian Journal of Statistics, vol. 6, no. 2, pp. 65–70, 1979.

S. B. Green, "How many subjects does it take to do a regression analysis?" Multivariate Behavioral Research, vol. 26, no. 3, pp. 499–510, 1991.

J. L. Schafer and J. W. Graham, "Missing data: Our view of the state of the art," Psychological Methods, vol. 7, no. 2, pp. 147–177, 2002.

S. Rojstaczer and C. Healy, "Where A is ordinary: The evolution of American college and university grading, 1940–2009," Teachers College Record, vol. 114, no. 7, pp. 1–23, 2012.

JOURNAL FOR SOCIAL SCIENCE STUDIES
ADVANCING KNOWLEDGE, INSPIRING SOCIETY