

Original Article

Predictive Validity of Entrance Selection Tests for Academic Performance and Study Completion in Postgraduate Education

Sendi Ramdhani^{1*} , Amalia Sapriati¹ , Kristanti Ambar Puspitasari¹ 

Abstract: This study examined the predictive validity of postgraduate entrance-test scores for cumulative grade point average (GPA) and study duration at the Postgraduate School of Universitas Terbuka, Indonesia. Using retrospective institutional records, the study included 112 alumni admitted under comparable testing procedures implemented from 2021 onward; analyses involving English proficiency were based on 111 cases because one score was recorded on a non-comparable scale. Pearson correlations and multiple linear regression were used to assess predictive relationships. Numerical reasoning showed the strongest bivariate association with GPA ($r = .227$, $p = .016$), followed by verbal reasoning ($r = .195$, $p = .040$). No entrance-test component was significantly associated with study duration. In the multivariable model, English proficiency ($\beta = -.233$, $p = .011$), numerical reasoning ($\beta = .271$, $p = .017$), and figural reasoning ($\beta = -.286$, $p = .004$) were significant predictors of GPA, with the model explaining 17.9% of its variance (adjusted $R^2 = .139$). However, after adjustment for study program, only numerical reasoning remained statistically significant ($\beta = .230$, $p = .044$). Entrance-test scores provided little evidence of predictive utility for study duration after program adjustment. Overall, the findings indicate limited and component-specific predictive validity, with program-level heterogeneity influencing the observed relationships. These results highlight the need to evaluate entrance-test components in relation to specific academic outcomes and institutional contexts rather than treating aggregate test performance as a uniformly reliable predictor of postgraduate success.

Keywords: Academic performance; Distance education; Postgraduate admissions; Predictive validity; Study duration



Author Affiliation:

¹ Universitas Terbuka, Tangerang Selatan, Indonesia

***Corresponding author(s):**

Sendi Ramdhani, School of Graduate Studies, Universitas Terbuka, Jalan Cabe Raya, Pondok Cabe, Pamulang, Kota Tangerang Selatan 15437, Banten, Indonesia.

 sendi@ecampus.ut.ac.id

Article History:

Received 6 June 2026; Revised 21 August 2026; Accepted 22 September 2026

Available online 20 October 2026

INTRODUCTION

The Postgraduate School of Universitas Terbuka (SPs UT) plays a strategic role in fostering students' academic development, critical thinking, and professional competence through effective teaching and learning. Postgraduate education extends beyond advanced

knowledge acquisition to developing students' abilities to conduct critical inquiry, communicate ideas, and apply knowledge to complex academic and professional challenges. Its effectiveness can be reflected in graduates' ability to transfer acquired knowledge and skills into professional practice and career development (Stemler, 2012). Research also highlights its contributions to professional development, improved work practices, and expanded professional networks (Patac *et al.*, 2024). Furthermore, student success is increasingly understood as a multidimensional construct encompassing academic achievement, engagement, persistence, and timely progression, rather than GPA alone (Tinto, 1997; van der Zanden *et al.*, 2019). By fostering higher-order thinking and innovation, postgraduate education can also prepare graduates for knowledge-intensive and high-technology industries (Zhang, 2024). Therefore, evaluating postgraduate education requires consideration of both learning processes and educational outcomes that support students' academic success and professional development.

Postgraduate admission decisions commonly rely on prior academic achievement, discipline-specific assessments, English-language proficiency, and general reasoning ability to evaluate applicants' academic readiness. However, the predictive validity of these indicators varies across academic programs, student populations, and institutional contexts (Al-Qahtani & Alanzi, 2018; Ferrão & Almeida, 2019; Jaehn *et al.*, 2025; Vulperhorst *et al.*, 2018). Research on English language proficiency tests indicates that their associations with academic achievement are generally small to moderate, suggesting that language proficiency should be considered alongside other indicators of academic readiness rather than as a sole predictor of student success (Ihlenfeldt & Rios, 2023; Martirosyan *et al.*, 2015; Neumann *et al.*, 2019). These findings underscore the need to assess admission tests not only for their ability to differentiate applicants but also for their capacity to predict subsequent academic performance and progression. This is particularly important in open and distance postgraduate education, where student outcomes may also depend on learning engagement and institutional or program-specific conditions. Therefore, the predictive relevance of admission-test scores needs to be empirically examined within specific educational contexts (de Boer & van Rijnsoever, 2022; Paris & Heiser, 2022; Paris *et al.*, 2023).

As an open and distance university, the Postgraduate School of Universitas Terbuka (SPs UT) serves students from diverse academic and professional backgrounds, underscoring the need to evaluate the predictive validity of its admission tests. Although admission assessments can predict postgraduate academic performance, their validity varies across tests and outcomes (Kuncel *et al.*, 2001; Paris & Heiser, 2022; Paris *et al.*, 2023). In distance education, academic success also depends on self-regulated learning, time management, and sustained engagement (Broadbent, 2017; Broadbent & Poon, 2015; Gašević *et al.*, 2016; Kintu *et al.*, 2017; Xu *et al.*, 2023). However, the predictive validity of SPs UT's admission-test components remains insufficiently documented. This study addresses this gap by examining their associations with cumulative grade point average (GPA) and study duration. Master's admission assessments include a program-specific Subject Matter Test, the Academic Potential Test (TPDA) for applicants without an acceptable existing score, and an English-proficiency requirement for applicants without

an acceptable score. Recognized academic-potential tests include the TPDA administered by Pusat Layanan Tes Indonesia (PLTI) and the TPA administered by OTO Bappenas, with a minimum TPA-equivalent score of 450 ([Universitas Terbuka, 2025](#)). Because Subject Matter Tests vary across programs in content and disciplinary focus, pooled estimates of predictive validity should be interpreted cautiously, as they may obscure program-specific differences in the relationships between admission-test scores and subsequent academic outcomes.

Despite satisfactory academic achievement, reflected in an average cumulative grade point average (CGPA) of 3.85, study completion remains a concern at the Postgraduate School of Universitas Terbuka (SPs UT). Master's students take an average of seven semesters (3 years and 3 months) to graduate, while doctoral students require eight semesters (4 years). This discrepancy highlights the need to examine whether admission assessments predict academic achievement and study progression. Accordingly, this study examines the associations of the Subject Matter Test, English Proficiency Test, Academic Potential Test (TPDA), and its verbal, numerical, and figural components with CGPA and study duration. Bivariate analyses and multivariable regression are used to assess the associations and unique predictive contributions of individual test components. Previous meta-analytic evidence indicates that the predictive validity of admission tests varies according to the test component and the academic outcome being assessed ([Kuncel *et al.*, 2001](#)). Given the retrospective observational design, the findings are interpreted as associations rather than causal relationships. The results may inform the evaluation of admission assessments and the development of targeted academic support to improve learning progress and timely completion, while recognizing that test scores alone cannot fully explain postgraduate academic outcomes.

THEORETICAL SUPPORT

Predictive validity concerns the extent to which information obtained at selection is related to relevant outcomes observed later in time. In admissions research, however, a statistically significant association is not sufficient by itself to establish strong predictive utility; the magnitude of the relationship, the incremental contribution of a test beyond other information, and the stability of the relationship across relevant subgroups must also be considered ([Stemler, 2012](#); [Molontay & Nagy, 2023](#)). Accordingly, this study evaluates predictive validity at two complementary levels. Pearson correlations describe zero-order relationships between each entrance-test component and subsequent GPA or study duration, whereas multiple regression evaluates whether a component contributes unique predictive information when the other admission measures are considered simultaneously. Because the data are retrospective and observational, predictive validity is interpreted as within-cohort statistical evidence rather than as causal or deterministic prediction. Recent graduate-admissions evidence likewise emphasizes that validity coefficients are usually modest and that incremental validity and criterion choice matter ([Kuncel *et al.*, 2001](#); [Kuryшева *et al.*, 2022](#); [Kuryшева *et al.*, 2023](#); [Feldon *et al.*, 2024](#); [Walters *et al.*, 2022](#); [Sorgente *et al.*, 2025](#)). Classic and contemporary work likewise shows that standardized

tests can have criterion-related validity while still accounting for only a limited proportion of outcome variance (Bridgeman *et al.*, 2009; Stemler, 2012).

The three admission domains examined in this study represent conceptually different forms of academic readiness. The Subject Matter Test is intended to capture discipline-specific prior knowledge; English proficiency reflects language resources that may support access to academic literature and scholarly communication; and TPDA represents broader reasoning capacity through verbal, numerical, and figural components. Prior admissions research indicates that the usefulness of such measures is context dependent: discipline-related and prior-achievement indicators may relate more strongly to academic performance in some settings, whereas language and general reasoning measures often show smaller or variable relationships across programs and student populations (Vulperhorst *et al.*, 2018; Ferrão & Almeida, 2019; Ihlenfeldt & Rios, 2023). This framework therefore does not assume that all entrance-test components should predict the two outcomes equally. Program alignment can also influence criterion relationships, particularly when prior preparation differs in cognitive relatedness to the graduate field (de Boer & Van Rijnsoever, 2023).

A second element of the framework is that postgraduate success, particularly in open and distance education, is not determined by admission-test performance alone. Self-regulated learning, motivation, engagement, employment demands, learning conditions, and institutional support have been associated with persistence and achievement in previous research (Broadbent & Poon, 2015; Jansen *et al.*, 2017; Xu *et al.*, 2023). These factors were not measured in the present institutional dataset and are therefore treated as potential contextual influences and sources of unmeasured confounding, not as explanations demonstrated by the current results. Study program is also theoretically important because both the content of the Subject Matter Test and the academic demands of postgraduate study differ across disciplines. For this reason, the study supplements pooled analyses with program-adjusted sensitivity models and interprets any apparent predictive relationship in light of program heterogeneity. Recent evidence from distance learners and postgraduate contexts further reinforces the relevance of self-regulation and student-level contextual factors (Theobald, 2021; Guntur & Purnomo, 2024; Vos *et al.*, 2025; Botha *et al.*, 2025). Broader meta-analytic evidence also links academic performance to self-efficacy, motivational, and learning-strategy variables (Richardson *et al.*, 2012), while engagement and instructional conditions are important in online settings (Gašević *et al.*, 2016; Martin & Bolliger, 2018). Indonesian evidence also supports the relevance of learner-regulation and engagement variables in technology-mediated learning. Adam *et al.* (2022) examined self-regulated learning strategies in remote university lecturing, while Dora *et al.* (2023) identified attitude, experience, motivation, and perception as relevant to university students' decisions regarding online courses. In the Universitas Terbuka postgraduate context, Rosita *et al.* (2023) found that patterns of LMS access and online-tutorial participation were associated with tutorial outcomes. These studies reinforce the view that admission scores capture only part of the conditions related to subsequent learning behavior and achievement in distance education.

METHOD

This study employed a retrospective correlational design using secondary institutional records to evaluate the predictive validity of postgraduate admission tests for cumulative grade point average (CGPA) and study duration among master's graduates. Predictors included Subject Matter Test scores, English Proficiency Test scores, overall Academic Potential Test (TPDA) scores, and verbal, numerical, and figural TPDA subscores. Pearson's correlations assessed zero-order associations, while multiple linear regression examined the unique contributions of admission-test components. Overall TPDA scores and their subscores were analyzed in separate models to avoid redundancy, and a study-program-adjusted sensitivity analysis assessed the robustness of the estimates across academic programs. Because the study used observational data, findings were interpreted as predictive associations rather than causal effects (Cohen *et al.*, 2003; Shmueli, 2010).

The target population comprised 893 master's graduates from the Postgraduate School of Universitas Terbuka (SPs UT); doctoral graduates were excluded because only two were available. To ensure comparability of admission procedures, the analysis was restricted to three cohorts admitted in March 2021, September 2021, and March 2022, yielding 120 eligible records. Eight records (6.7%) with incomplete entrance-test data were excluded, leaving 112 complete cases for analyses not involving English proficiency. One additional English score (483) was excluded from English-related analyses because its scale was incompatible with the institutional 1–100 scale and no verified conversion rule was available, resulting in 111 cases. Accordingly, the study analyzed all available eligible institutional records rather than a sample drawn from the initial population of 893 graduates. Findings are limited to master's graduates in the eligible cohorts, and potential selection bias due to incomplete records should be considered.

Secondary quantitative data were extracted from official admission and academic records maintained by the Postgraduate School of Universitas Terbuka (SPs UT) for the March 2021, September 2021, and March 2022 admission cohorts. Variables included Subject Matter Test scores, English Proficiency Test scores, overall TPDA scores and verbal, numerical, and figural subscores, final cumulative GPA (on a 0.00–4.00 scale), and study duration in days. Records were linked using institutional identifiers and anonymized before analysis. Study duration was calculated from the official program-entry date to the final examination or graduation milestone recorded in the institutional database. Under SPs UT regulations, the standard master's program duration is four semesters, with students exceeding this period classified as Study Duration Exceeded and a maximum study period of four years (Universitas Terbuka, 2025). Because the package-based curriculum does not provide academic leave, no leave periods were deducted, and delayed registration remained part of the elapsed duration. Accordingly, shorter study duration was interpreted as faster completion relative to the institutional timeline, not as a direct indicator of educational quality. The online TPDA, administered by Pusat Layanan Tes Indonesia (PLTI), assesses verbal, numerical, and figural abilities through eight subtests comprising 160 items and uses a CEEB standard scale ranging from 200 to 800, with a normative mean of 500 (PLTI, n.d.). SPs UT recognizes PLTI's TPDA and OTO Bappenas's TPA for

admission, with a minimum TPA-equivalent score of 450 (Universitas Terbuka, 2025). As no verifiable reliability or technical validity coefficients were available for the specific TPDA records analyzed, no psychometric coefficients were assumed or reported.

To ensure clarity in the measurement and interpretation of the variables, this study provides operational definitions for each variable analyzed. The variables were categorized into predictor variables and outcome variables. The predictor variables consisted of entrance selection test components, while the outcome variables consisted of postgraduate students' Grade Point Average (GPA) and study duration. The operational definitions of the variables are presented in Table 1.

Table 1. Operational definitions of predictor and outcome variables

Variable	Type	Operational Definition	Data Source
GPA	Outcome variable	Final cumulative GPA at graduation, recorded on the institutional 0.00–4.00 scale.	Academic records
Study Duration	Outcome variable	Elapsed time from official program entry to the recorded final examination/graduation milestone, measured in days. SPs UT has a four-semester standard curriculum, a maximum study period of four years, and no academic leave; delayed registration remains counted in study duration.	Academic records
Substance Test	Predictor variable	Program-specific subject-matter entrance-test score. Scores are analyzed on the recorded admission scale; content differs by study program.	Admission records
English Proficiency Test	Predictor variable	Institutional English proficiency score on a 1–100 scale. One record reported on a non-comparable TOEFL/TOEP-type scale (483) was excluded from English-related analyses because no verified conversion rule was available.	Admission records
TPDA	Predictor variable	Overall TPDA score derived from verbal, numerical, and figural domains. PLTI reports TPDA on a CEEB standard scale of 200–800 (mean 500); SPs UT recognizes TPDA/TPA for admission with a minimum TPA-equivalent requirement of 450.	Admission records
Verbal Aspect	Predictor variable	TPDA subscore measuring verbal reasoning on the PLTI CEEB-standard reporting scale.	TPDA records
Numerical Aspect	Predictor variable	TPDA subscore measuring numerical reasoning on the PLTI CEEB-standard reporting scale.	TPDA records
Figural Aspect	Predictor variable	TPDA subscore measuring figural or abstract reasoning on the PLTI CEEB-standard reporting scale.	TPDA records

Data were analyzed in two stages. First, Pearson's product moment correlations were calculated between each admission-test component and the two outcomes, with sample size (N), correlation coefficient (r), two-sided p-value, and 95% confidence interval based on Fisher's z transformation reported. Second, multiple linear regression examined the unique predictive contributions of admission measures using two specifications for each outcome: Model 1 included Subject Matter Test, English Proficiency Test, and overall TPDA scores, whereas Model 2 replaced the overall TPDA score with its verbal, numerical, and figural subscores to avoid redundancy. Regression results included standardized and unstandardized coefficients, 95% confidence intervals, p-values, R^2 , adjusted R^2 , and the overall F-test. Effect magnitudes were interpreted separately from statistical significance, and variance inflation factors (1.04–1.59) indicated no problematic

multicollinearity. A sensitivity analysis incorporating study-program indicators assessed the robustness of pooled estimates to program-level heterogeneity. Complete-case analysis was used for multivariable models, excluding eight of the 120 eligible records with incomplete entrance-test data; the single non-comparable English score was excluded only from English-related analyses. Potential selection bias and unmeasured confounding by factors such as age, prior academic achievement, professional experience, and employment status were acknowledged. Given the observational design, results were interpreted as predictive associations rather than causal effects.

RESULT AND DISCUSSION

This section presents the findings and discusses the predictive validity of postgraduate entrance-selection tests for academic performance and study completion. The results focus on the relationships between entrance-test scores and cumulative grade point average (GPA) and study duration, the unique contributions of individual test components, and the robustness of these relationships after accounting for differences across study programs. Table 2 presents the zero-order correlations between entrance-test scores and the two outcomes, while subsequent multiple regression analyses assess the unique contributions of individual test components and a program-adjusted sensitivity analysis evaluates the robustness of the findings. The correlation results indicate that numerical and verbal reasoning had small positive associations with GPA, whereas none of the entrance-test components was significantly associated with study duration, suggesting that the predictive validity of the selection measures was limited and varied across test components and academic outcomes.

Table 2. Summary of correlations between entrance test scores and postgraduate student outcomes

Entrance Test Component	GPA: N, r, p, 95% CI	Interpretation	Study Duration: N, r, p, 95% CI	Interpretation
Substance Test	112; r = .140; p = .140; 95% CI [-.046, .318]	Small, non-significant	112; r = .041; p = .665; 95% CI [-.145, .225]	Negligible, non-significant
English Proficiency Test	111; r = -.182; p = .056; 95% CI [-.356, .005]	Small, non-significant	111; r = -.002; p = .981; 95% CI [-.189, .184]	Negligible, non-significant
TPDA	112; r = .102; p = .283; 95% CI [-.085, .283]	Small, non-significant	112; r = .030; p = .756; 95% CI [-.157, .214]	Negligible, non-significant
Verbal Aspect	112; r = .195; p = .040; 95% CI [.010, .367]	Small positive	112; r = -.022; p = .819; 95% CI [-.207, .164]	Negligible, non-significant
Numerical Aspect	112; r = .227; p = .016; 95% CI [.043, .396]	Small positive	112; r = -.043; p = .652; 95% CI [-.227, .144]	Negligible, non-significant
Figural Aspect	112; r = -.143; p = .132; 95% CI [-.320, .044]	Small, non-significant	112; r = .176; p = .063; 95% CI [-.010, .350]	Small, non-significant

Table 2 presents the correlations between postgraduate entrance-test scores and two academic outcomes: cumulative grade point average (GPA) and study duration. The results indicate that the overall Substance Matter Test, English Proficiency Test, and TPDA scores were not significantly correlated with either outcome ($p > .05$). However, the verbal and numerical aspects of the TPDA showed small but statistically significant positive

correlations with GPA ($r = .195$, $p = .040$, and $r = .227$, $p = .016$, respectively), suggesting that higher scores in these components were associated with slightly higher academic achievement. In contrast, none of the entrance-test components showed a statistically significant correlation with study duration. These findings suggest that the predictive relationships between entrance-test performance and postgraduate outcomes were limited and component-specific, with verbal and numerical reasoning demonstrating modest associations with GPA but no clear relationship with the time required to complete the degree. The regression results are presented in Table 3.

Table 3. Multiple regression results for GPA and study duration

Outcome / Model	Predictor	B	SE	Standardized beta	p	95% CI for B	VIF
GPA Model 1	Substance Test	.00137	.00089	.157	.129	[-.00040, .00314]	1.20
GPA Model 1	English	-.00170	.00074	-.219	.023	[-.00315, -.00024]	1.04
GPA Model 1	TPDA	.000086	.000127	.069	.500	[-.000165, .000337]	1.18
GPA Model 2	Substance Test	.00067	.00089	.077	.449	[-.00108, .00243]	1.31
GPA Model 2	English	-.00180	.00070	-.233	.011	[-.00318, -.00042]	1.04
GPA Model 2	Verbal	.000121	.000112	.118	.286	[-.000102, .000343]	1.55
GPA Model 2	Numerical	.000268	.000111	.271	.017	[.000049, .000488]	1.59
GPA Model 2	Figural	-.000245	.000082	-.286	.004	[-.000409, -.000082]	1.18
Duration Model 1	Substance Test	.529	1.591	.035	.740	[-2.624, 3.682]	1.20
Duration Model 1	English	-.147	1.309	-.011	.911	[-2.741, 2.448]	1.04
Duration Model 1	TPDA	.038	.225	.018	.867	[-.409, .485]	1.18
Duration Model 2	Substance Test	1.119	1.638	.074	.496	[-2.129, 4.367]	1.31
Duration Model 2	English	-.019	1.291	-.001	.989	[-2.578, 2.541]	1.04
Duration Model 2	Verbal	-.088	.208	-.050	.674	[-.499, .324]	1.55
Duration Model 2	Numerical	-.222	.204	-.130	.279	[-.628, .183]	1.59
Duration Model 2	Figural	.334	.152	.227	.030	[.032, .637]	1.18

Note. $N = 111$ for all regression models because English proficiency was included and one English score was non-comparable. GPA Model 1: $R^2 = .069$, adjusted $R^2 = .043$, $F(3, 107) = 2.65$, $p = .052$. GPA Model 2: $R^2 = .179$, adjusted $R^2 = .139$, $F(5, 105) = 4.56$, $p < .001$. Duration Model 1: $R^2 = .002$, adjusted $R^2 = -.026$, $F(3, 107) = 0.07$, $p = .976$. Duration Model 2: $R^2 = .050$, adjusted $R^2 = .004$, $F(5, 105) = 1.10$, $p = .367$. Individual coefficients in an overall non-significant model should not be interpreted as robust predictive evidence.

Table 3 presents the multiple regression results examining the independent contributions of entrance-test components to grade point average and study duration. For grade point average, the first model was not statistically significant overall, although the

English Proficiency Test score showed a significant negative association with academic achievement. The second model was statistically significant, with numerical ability showing a significant positive association and figural ability showing a significant negative association with grade point average. Neither regression model for study duration was statistically significant overall. Although figural ability had a significant positive coefficient in the second duration model, this result should be interpreted cautiously because the overall model was not statistically significant. The variance inflation factors indicated no problematic multicollinearity among the predictors. Overall, the findings suggest that the predictive contributions of entrance-test components were limited and varied across academic outcomes.

To assess the robustness of these findings, a program-adjusted sensitivity analysis was conducted by incorporating study-program indicators into the regression models. This analysis examined whether the relationships between entrance-test scores and academic outcomes remained consistent after accounting for potential differences across academic programs. The results of this analysis are presented in Table 4.

Table 4. Program adjusted sensitivity analysis

Outcome / Model	Predictor	Standardized beta	p	Interpretation
GPA Model 1 + program	Substance Test	.047	.645	Not significant
GPA Model 1 + program	English	-.106	.293	Not significant
GPA Model 1 + program	TPDA	.154	.103	Not significant
GPA Model 2 + program	Substance Test	.025	.811	Not significant
GPA Model 2 + program	English	-.136	.180	Not significant
GPA Model 2 + program	Verbal	.078	.481	Not significant
GPA Model 2 + program	Numerical	.230	.044	Small unique positive contribution
GPA Model 2 + program	Figural	-.100	.324	Not significant
Duration Model 1 + program	Substance Test	.039	.645	Not significant
Duration Model 1 + program	English	-.128	.124	Not significant
Duration Model 1 + program	TPDA	.006	.939	Not significant
Duration Model 2 + program	Substance Test	.042	.629	Not significant
Duration Model 2 + program	English	-.111	.195	Not significant
Duration Model 2 + program	Verbal	.064	.492	Not significant
Duration Model 2 + program	Numerical	-.084	.374	Not significant
Duration Model 2 + program	Figural	.012	.885	Not significant

Note. Study program was included as indicator variables. Adding entrance-test components did not significantly improve the prediction of GPA or study duration beyond study program. These adjusted models should be interpreted cautiously because some programs had small sample sizes. GPA = Grade Point Average; TPDA = Academic Potential Test. N = 112, except for English proficiency (N = 111). Confidence intervals are 95% Fisher z intervals. Statistical significance does not necessarily indicate practical importance or causality.

The pattern in Table 2 is important for two reasons. First, statistical significance and practical magnitude do not point to strong prediction: even the two statistically significant GPA correlations numerical reasoning ($r = .227$, $p = .016$, 95% CI [.043, .396]) and verbal reasoning ($r = .195$, $p = .040$, 95% CI [.010, .367]) represent small effects. Second, the remaining GPA coefficients were small and non-significant, including the Subject Matter Test ($r = .140$), overall TPDA ($r = .102$), English proficiency ($r = -.182$), and figural reasoning ($r = -.143$). All correlations with study duration were also non-significant. Thus, the bivariate analysis provides limited evidence that the entrance tests, considered individually, are informative about later postgraduate outcomes.

Relationship Between Entrance Test Scores and GPA

English Proficiency Test

English proficiency showed a small inverse bivariate association with GPA ($r = -.182$, $N = 111$, $p = .056$, 95% CI [-.356, .005]). Because the confidence interval included zero and the p -value was slightly above .05, this result is interpreted as statistically inconclusive rather than as evidence of a reliable negative relationship.

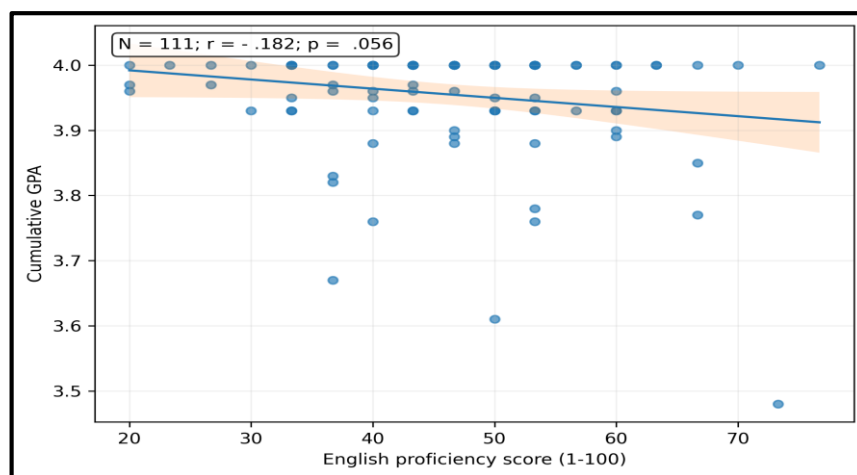


Figure 1. English proficiency score and GPA ($N = 111$, $r = -.182$, $p = .056$). The solid line represents the fitted regression, with the shaded band indicating the 95% confidence interval.

Figure 1 illustrates substantial dispersion in GPA across the English-proficiency range and only a slight downward fitted trend. The visual pattern is consistent with the small correlation and reinforces the conclusion that English scores provided limited zero-order information about GPA in this cohort. This finding is compatible with literature showing that English proficiency can be relevant to academic participation while still displaying only weak or variable relationships with GPA across higher-education contexts (Ihlenfeldt & Rios, 2023; Neumann *et al.*, 2019). In the present study, however, the English score should not be interpreted as either beneficial or detrimental to academic performance on the basis of the bivariate coefficient alone. Recent studies similarly report context-sensitive predictive relationships for English-language measures and academic outcomes (Bo *et al.*, 2023; Isaacs *et al.*, 2023; Chakiso *et al.*, 2025; Althewini, 2026). Earlier work similarly cautions that language proficiency is only one of several contributors to academic adaptation and achievement (Andrade, 2006; Martirosyan *et al.*, 2015).

The multivariable results further demonstrate why a zero-order coefficient should not be labeled a predictor without qualification. English proficiency had a significant inverse coefficient in the unadjusted regression models, but that coefficient was no longer significant after study program was controlled. This instability suggests that the apparent unique relationship was sensitive to program composition and should not be interpreted as robust incremental predictive validity.

Measurement comparability is an additional consideration. The institutional dataset contained one English score reported on a TOEFL/TOEP-type scale rather than the institutional 1–100 scale. Because no verified conversion rule was available, that score was excluded only from English-related analyses. This decision avoided an arbitrary transformation but also highlights the need for SPs UT to maintain a clearly documented common metric when different English assessments are accepted for admission. Variables such as motivation, learning strategies, research skills, academic support, employment demands, and time management were not available in the dataset. They therefore cannot be used to explain the English GPA pattern observed here. Their relevance is limited to identifying plausible domains for future prospective studies of postgraduate success. Taken together, the present evidence does not support using English proficiency as a stand-alone predictor of GPA. Its more defensible role in the current admissions system is as one component of a broader readiness profile and, potentially, as a diagnostic indicator for academic-language support, subject to separate validation of that use.

Academic Potential Test (TPDA)

The analysis of the Academic Potential Test (TPDA) examined whether its individual components were associated with postgraduate academic performance. Among the three components, numerical reasoning showed a small but statistically significant positive correlation with GPA, suggesting that higher numerical reasoning scores were associated with slightly higher academic achievement. This relationship is illustrated in Figure 2, which shows the positive association between TPDA numerical reasoning scores and cumulative grade point average (GPA), along with the fitted regression line and its 95% confidence interval.

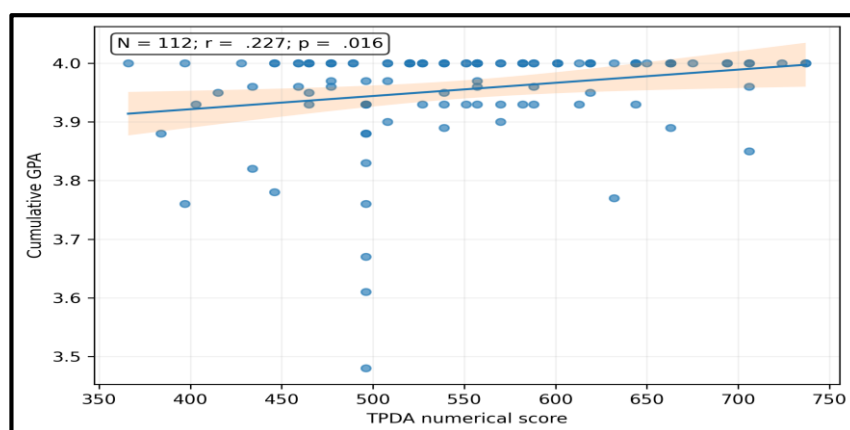


Figure 2. TPDA numerical reasoning score and GPA ($N = 112$, $r = .227$, $p = .016$), with the fitted regression line and 95% confidence interval.

Among the TPDA subcomponents, numerical reasoning showed the largest positive bivariate association with GPA ($r = .227$, $p = .016$), followed by verbal reasoning ($r = .195$, $p = .040$). Figure 2 shows the numerical relationship. Although statistically detectable, the effect was small, indicating that numerical reasoning explains only a limited portion of individual differences in cumulative GPA. A plausible theoretical interpretation is that numerical reasoning captures general analytical processes that are repeatedly used in postgraduate coursework and research-related tasks. Prior research has likewise found that quantitative or achievement-related indicators can contribute to academic-performance prediction when they align with the cognitive demands of the program (Ferrão & Almeida, 2019; Vulperhorst *et al.*, 2018). The present data do not establish the mechanism, however; they only show that the numerical component had the most consistent positive relationship with GPA among the TPDA domains examined.

The TPDA subcomponents did not behave uniformly. Verbal reasoning had a small positive bivariate relationship with GPA, whereas figural reasoning had a small negative, non-significant zero-order relationship. In the unadjusted multivariable model, numerical and figural coefficients were significant in opposite directions, but after program adjustment only numerical reasoning remained significant. This pattern suggests that interpreting TPDA solely as a composite score may conceal meaningful differences among its components and that program composition affects the apparent predictive contribution of those components. The differing behavior of verbal, numerical, and figural scores is also consistent with the broader psychometric point that reasoning domains need not have equivalent criterion relevance across academic fields (Lohman, 2005).

Accordingly, numerical reasoning can be described as the most consistent TPDA component in this dataset, but not as a universally 'best predictor.' Its standardized coefficient after program adjustment remained small ($\beta = .230$, $p = .044$), and the entrance-test block as a whole did not add a statistically significant amount of explained GPA variance beyond study program. Any change in admission weighting would therefore require prospective replication and evaluation of incremental validity in new cohorts.

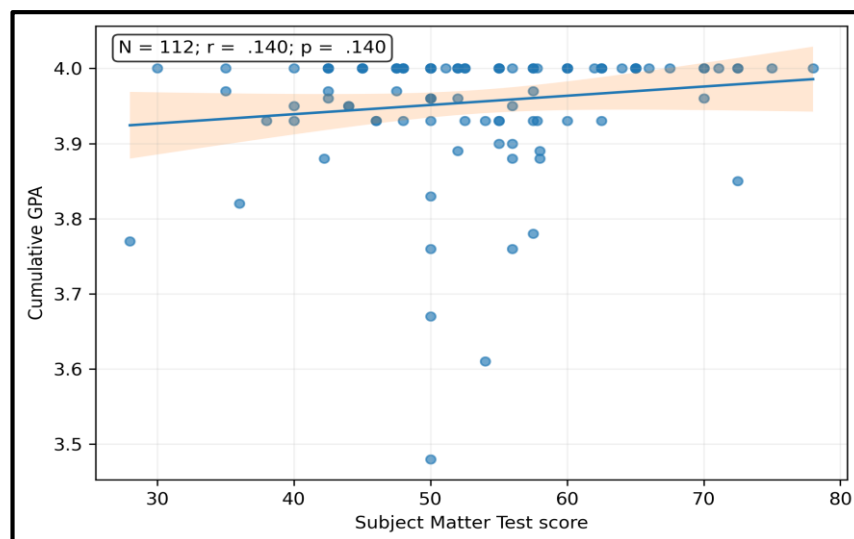


Figure 3. Subject matter test score and cumulative GPA ($N = 112$, $r = .140$, $p = .140$). The solid line is the fitted linear regression line and the shaded band is the 95% confidence interval for the mean fitted value.

Figure 3 shows only a slight positive fitted relationship between Subject Matter Test score and GPA ($r = .140$, $p = .140$). The confidence interval included zero, indicating that the pooled association was uncertain and small. Discipline-specific tests are theoretically attractive because they assess knowledge directly related to the field of study. Nevertheless, theoretical relevance should not be equated with demonstrated predictive strength. In the present pooled sample, the Subject Matter Test was not significantly associated with GPA, and it did not show a significant unique contribution in either the unadjusted or program-adjusted regression models.

The pooled coefficient must also be interpreted in light of measurement heterogeneity. Unlike the English test or TPDA, the Subject Matter Test is program specific; therefore, students from different master's programs were not necessarily answering equivalent content domains. Combining these scores assumes sufficient comparability of the recorded scale across programs, an assumption that cannot be fully evaluated from the available secondary data. Program-specific validation is therefore more appropriate than drawing a single institution-wide conclusion from the pooled coefficient. For SPs UT, the immediate implication is not to increase or decrease the Subject Matter Test weight on the basis of this study alone. A stronger evidence base would require examining score reliability and content validity within each program, estimating program-specific criterion relationships where sample sizes permit, and then testing whether the score adds predictive information beyond other admission components.

Relationship Between Entrance Tests and Study Duration

English Proficiency Test

The study-duration results differed clearly from the GPA results. None of the entrance-test components showed a statistically significant zero-order relationship with time to degree, and all absolute correlations were .176 or smaller. Thus, the present data provide little evidence that individual admission-test scores can identify which graduates will complete their programs more quickly.

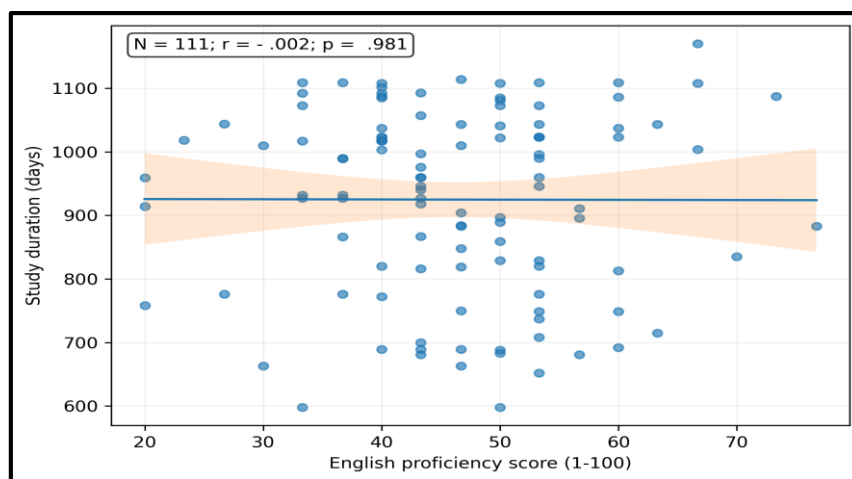


Figure 4. English proficiency score and study duration in days ($N = 111$, $r = -.002$, $p = .981$). The solid line is the fitted linear regression line and the shaded band is the 95% confidence interval for the mean fitted value.

Figure 4 shows an essentially flat fitted relationship between English proficiency and study duration, consistent with $r = -.002$ ($p = .981$). English proficiency was essentially unrelated to study duration ($r = -.002$, $N = 111$, $p = .981$, 95% CI $[-.189, .184]$). The verified analysis therefore does not support the earlier interpretation that higher English proficiency is associated with faster completion. The near-zero coefficient is best interpreted as absence of meaningful evidence in this cohort, not as evidence that language proficiency can never matter for particular stages of postgraduate study.

The weak entrance-test relationships become more understandable when program differences are considered. Study program alone explained 54.7% of the observed variance in study duration in the sensitivity analysis, whereas adding the five entrance-test components increased R^2 by only .016 and did not significantly improve the model. This indicates that the pooled variation in time to degree was far more strongly structured by program membership than by the available admission-test scores. Because the institutional dataset did not include employment status, leave patterns, supervision processes, family responsibilities, or other potential determinants of time to degree, the study cannot specify why programs differed in completion duration. These variables should be measured directly in future longitudinal research rather than invoked as post hoc explanations of the present coefficients.

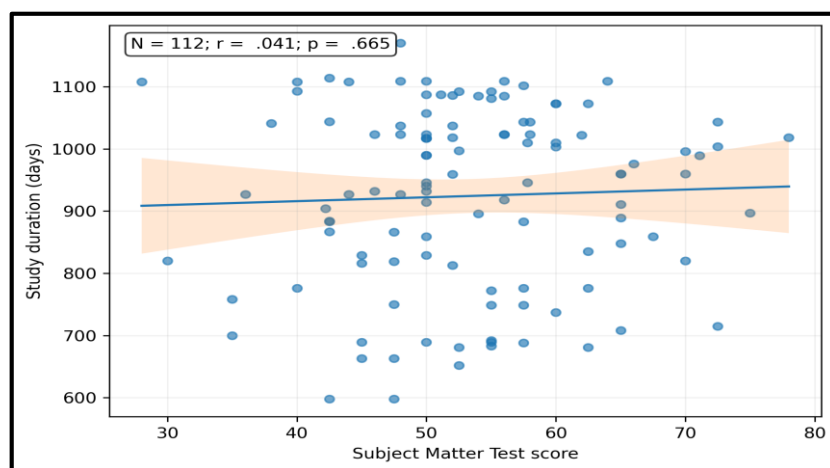


Figure 5. Subject Matter Test scores and study duration ($N = 112$, $r = .041$, $p = .665$). The solid line represents the fitted regression line, and the shaded band indicates the 95% confidence interval.

Figure 5 shows a negligible relationship between Subject Matter Test scores and study duration ($r = .041$, $p = .665$), indicating no significant association between discipline-specific knowledge and time to degree. Given the negligible magnitude of the coefficient, elaborate substantive explanations would be unwarranted. The more informative result is that admission knowledge scores contributed little to prediction of study duration in the pooled cohort, especially after program differences were taken into account.

Correlation Between Entrance Test Aspects and Study Duration

The correlations between TPDA components and study duration were examined to determine whether specific reasoning abilities were associated with time to degree. Figure 6 illustrates the relationship between figural reasoning scores and study duration.

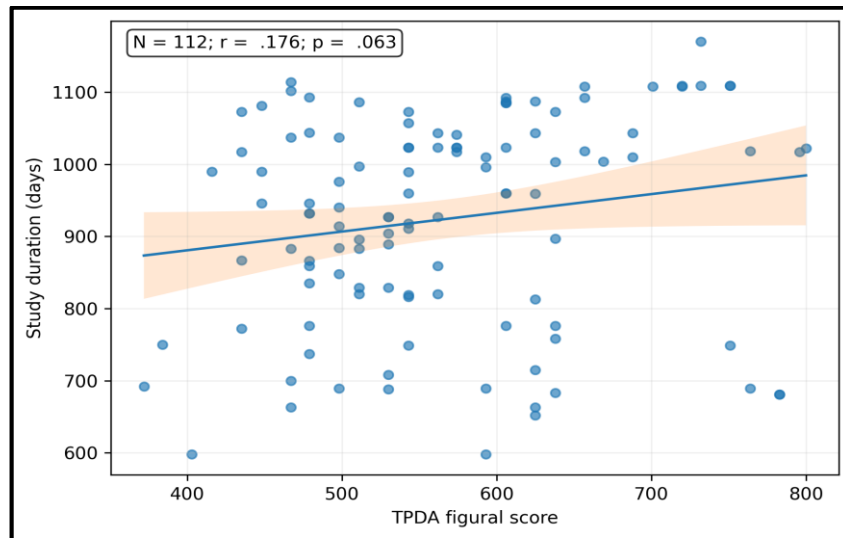


Figure 6. TPDA figural reasoning score and study duration in days ($N = 112$, $r = .176$, $p = .063$). The solid line is the fitted linear regression line and the shaded band is the 95% confidence interval for the mean fitted value.

The TPDA subcomponents were similarly weak in relation to study duration: verbal $r = -.022$ ($p = .819$), numerical $r = -.043$ ($p = .652$), and figural $r = .176$ ($p = .063$). Figure 6 displays the largest of these coefficients. Even this figural association remained small and statistically non-significant at the bivariate level. Numerical reasoning, which showed the clearest positive relationship with GPA, was negligibly related to study duration ($r = -.043$, $p = .652$). This contrast indicates that a component associated with academic performance need not also predict time to degree, supporting the treatment of GPA and study duration as distinct criterion outcomes rather than interchangeable indicators of 'academic success.' Although figural reasoning produced the largest TPDA coefficient for study duration, the bivariate result did not reach statistical significance. Its coefficient became significant in the unadjusted multivariable model, but the overall model was non-significant and the figural effect disappeared after adjustment for study program. The finding is therefore not regarded as robust predictive evidence.

Taken together, the study-duration analyses show little criterion-related predictive utility for the entrance-test components. Future validation should examine time to degree using designs that explicitly model program structure and interruptions in enrollment, and where suitable data are available consider longitudinal or time-to-event approaches rather than relying only on linear models.

Multivariable Predictive Analysis

The multivariable analyses sharpen the distinction between observed association and predictive contribution. For GPA, Model 1 (Subject Matter Test, English proficiency, and overall TPDA) explained 6.9% of the variance and narrowly missed the conventional threshold for the overall test, $F(3, 107) = 2.65$, $p = .052$, adjusted $R^2 = .043$. Model 2, which replaced overall TPDA with its verbal, numerical, and figural components, was significant, $F(5, 105) = 4.56$, $p < .001$, $R^2 = .179$, adjusted $R^2 = .139$. In this unadjusted model, English proficiency ($\beta = -.233$, $p = .011$), numerical reasoning ($\beta = .271$, $p = .017$), and figural reasoning ($\beta = -.286$, $p = .004$) had significant unique coefficients. Because the English

and figural coefficients differed from their zero-order patterns, these estimates were treated as potentially sensitive to suppression or program composition rather than as immediately actionable predictive effects.

For study duration, the evidence was substantially weaker. Model 1 explained virtually none of the variance, $R^2 = .002$, $F(3, 107) = 0.07$, $p = .976$, while Model 2 explained 5.0% and was also non-significant, $F(5, 105) = 1.10$, $p = .367$. Although figural reasoning had an individually significant coefficient in Model 2 ($\beta = .227$, $p = .030$), interpreting that coefficient as predictive validity would be inappropriate because the overall model was non-significant and the effect was not stable in the program-adjusted sensitivity analysis. Program adjustment materially changed the interpretation of both outcomes. For GPA, the program-only model explained 32.4% of the variance. Adding the five entrance-test components increased R^2 by .058, but this increment was not statistically significant, $F(5, 95) = 1.77$, $p = .126$. Within the adjusted subcomponent model, only numerical reasoning retained a statistically significant unique coefficient ($\beta = .230$, $p = .044$). For study duration, program alone explained 54.7% of the variance, and adding the entrance-test block increased R^2 by only .016, $F(5, 95) = 0.68$, $p = .640$. None of the entrance-test components remained significant. These results show that predictive validity is limited, component specific, and strongly contingent on program context rather than uniform across SPs UT.

The practical implications should therefore be incremental rather than prescriptive. The findings justify continued validation of the numerical TPDA component, separate reporting of TPDA subcomponents for evaluation purposes, program-specific review of the Subject Matter Test, and standardized documentation of English-score equivalence. They do not justify immediate changes to admission cut scores or test weights. A future admissions-policy study should prospectively test whether alternative weighting schemes improve prediction beyond program membership and other available applicant characteristics, preferably using a new cohort for cross-validation. This cautious position is consistent with contemporary work on holistic and context-sensitive admissions, which recommends evaluating multiple criteria without allowing a single standardized metric to dominate decision making ([Wilson et al., 2019](#); [Almalki et al., 2024](#); [Punongbayan, 2025](#); [Seyfried & Hollenberg, 2024](#)).

CONCLUSION

This study provides evidence of limited and component-specific predictive validity of postgraduate entrance-selection tests at the Postgraduate School of Universitas Terbuka (SPs UT), Indonesia. Numerical reasoning showed the most consistent relationship with academic performance, exhibiting a small positive correlation with cumulative GPA ($r = .227$) and remaining statistically significant after adjustment for study program ($\beta = .230$, $p = .044$). By contrast, verbal reasoning showed only a small bivariate association, while the Subject Matter Test, overall TPDA, and English proficiency demonstrated no stable independent predictive contributions after program adjustment. Entrance-test scores also provided little evidence of predictive utility for study duration, and the test components did not significantly improve either program-adjusted model as a block. These findings

suggest that entrance-test scores provide some information about subsequent academic performance but should not be interpreted as uniformly reliable predictors or causal determinants of postgraduate success.

These conclusions should be interpreted in light of the retrospective design, the 112 complete cases drawn from comparable post-2021 admission cohorts, limited representation within some study programs, and the absence of relevant applicant characteristics and contextual factors. Moreover, the models were not validated in independent cohorts; therefore, the findings establish within-cohort associations rather than out-of-sample predictive accuracy. For SPs UT, the results support further validation rather than immediate changes to admission criteria. Future research should incorporate prior academic achievement and relevant applicant characteristics, evaluate subject-specific tests within adequately represented programs, standardize English-proficiency measurement, and assess predictive performance prospectively using independent cohorts. Such evidence would provide a stronger basis for determining whether adjustments to test weighting, cut-off scores, or broader admission procedures are warranted.

Acknowledgments

This research was funded by the Institute for Research and Community Service (LPPM), Universitas Terbuka. The authors thank the Postgraduate School of Universitas Terbuka for authorized access to the institutional records used in this study. Conflict of interest: The authors declare no conflict of interest.

Author Contributions

Sendi Ramdhani contributed to conceptualization, methodology, data analysis, and manuscript drafting. Amalia Sapriati contributed to research supervision, methodological review, and critical revision of the manuscript. Kristanti Ambar Puspitasari contributed to data interpretation, literature review, and manuscript editing. All authors reviewed and approved the final manuscript and accepted accountability for its content.

Funding

This research was funded by the Institute for Research and Community Service (LPPM), Universitas Terbuka.

Conflict of Interest

The authors declare that they have no conflicts of interest that could have influenced the conduct, analysis, interpretation, or reporting of this study.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to institutional policies, ethical considerations, and applicable data confidentiality requirements.

Declaration of Generative AI and AI-Assisted Technologies

The authors used ChatGPT to assist with language refinement and readability. The authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Ethical Approval and Informed Consent

The authors confirm that this study used anonymized secondary institutional data and involved no direct participant contact or intervention. The authors are responsible for ensuring that data use was appropriately authorized and complied with applicable institutional policies, ethical requirements, and data-protection regulations. Individual informed consent was not sought because the study used secondary institutional data. The authors assume full responsibility for the accuracy of the ethical declarations and the lawful use and confidentiality of the data.

REFERENC

- Adam, N. S., Kartowagiran, B., Yansa, H., Marlina, & Amiruddin, M. (2022). Self-regulated learning strategy in conducting remote lecturing during COVID-19 pandemic. *Jurnal Pendidikan Indonesia*, 11(2), 274–282. <https://doi.org/10.23887/jpiundiksha.v11i2.34801>
- Al-Qahtani, M. F., & Alanzi, T. M. (2019). Comparisons of the predictive values of admission criteria for academic achievement among undergraduate students of health and non-health science professions: A longitudinal cohort study. *Psychology Research and Behavior Management*, 12, 1–6. <https://doi.org/10.2147/PRBM.S183651>
- Almalki, S. A., AlJameel, A. H., Alghomlas, Z., Alothman, T., & Alhajri, F. (2024). Assessing the predictive validity of pre-admission criteria on dental students' academic performance: A cross-sectional study. *BMC Oral Health*, 24, Article 90. <https://doi.org/10.1186/s12903-023-03839-6>
- Andrade, M. S. (2006). International students in English-speaking universities: Adjustment factors. *Journal of Research in International Education*, 5(2), 131–154. <https://doi.org/10.1177/1475240906065589>
- Althewini, A. (2026). How well do admission criteria predict student performance in one medical course? *BMC Medical Education*, 26, Article 312. <https://doi.org/10.1186/s12909-025-08378-y>
- Bo, W. V., Fu, M., & Lim, W. Y. (2023). Revisiting English language proficiency and its impact on the academic performance of domestic university students in Singapore. *Language Testing*, 40(1), 133–152. <https://doi.org/10.1177/02655322211064629>
- Botha, L., Viviers, H. A., Fouché, J. P., & Kruger, P. (2025). Success factors for postgraduate accountancy and related business students in higher education: A systematic review. *Journal of Further and Higher Education*, 49(6), 749–765. <https://doi.org/10.1080/0309877X.2025.2495957>
- Bridgeman, B., Burton, N., & Cline, F. (2009). A note on presenting what predictive validity numbers mean. *Applied Measurement in Education*, 22(2), 109–119. <https://doi.org/10.1080/08957340902754577>
- Broadbent, J. (2017). Comparing online and blended learners' self-regulated learning strategies and academic performance. *The Internet and Higher Education*, 33, 24–32. <https://doi.org/10.1016/j.iheduc.2017.01.004>
- Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies and academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, 27, 1–13. <https://doi.org/10.1016/j.iheduc.2015.04.007>
- Chakiso, Z. T., Bushisso, E. W., & Wanna, W. (2025). Unveiling predictive validity of English language exam on student achievement: Mediated by self-rated English proficiency. *Language Testing in Asia*, 15, Article 26. <https://doi.org/10.1186/s40468-025-00356-x>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates.
- de Boer, T., & van Rijnsoever, F. (2022). In search of valid non-cognitive student selection criteria. *Assessment & Evaluation in Higher Education*, 47(5), 783–800. <https://doi.org/10.1080/02602938.2021.1958142>
- de Boer, T., & van Rijnsoever, F. J. (2023). One field too far? Higher cognitive relatedness between bachelor and master leads to better predictive validity of bachelor grades during admission. *Assessment & Evaluation in Higher Education*, 48(7), 966–979. <https://doi.org/10.1080/02602938.2022.2158453>
- Dora, Y. M., Widokarti, J. R., Ismail, A., & Buana, D. M. A. (2023). Online course: Factors influencing students' decision. *Jurnal Organisasi dan Manajemen*, 19(2), 316–331. <https://doi.org/10.33830/jom.v19i2.5103.2023>
- Feldon, D. F., Litson, K., Cahoon, B., Feng, Z., Walker, A., & Tofel-Grehl, C. (2024). The predictive validity of the GRE across graduate outcomes: A meta-analysis of trends over time. *The Journal of Higher Education*, 95(1), 120–148. <https://doi.org/10.1080/00221546.2023.2187177>

- Ferrão, M., & Almeida, L. (2019). Differential effect of university entrance score on first-year students' academic performance in Portugal. *Assessment & Evaluation in Higher Education*, 44(4), 610–622. <https://doi.org/10.1080/02602938.2018.1525602>
- Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28, 68–84. <https://doi.org/10.1016/j.iheduc.2015.10.002>
- Guntur, M., & Purnomo, Y. W. (2024). A meta-analysis of self-regulated learning interventions studies on learning outcomes in online and blended environments. *Online Learning*, 28(3). <https://doi.org/10.24059/olj.v28i3.4025>
- Ihlenfeldt, S. D., & Rios, J. A. (2023). A meta-analysis on the predictive validity of English language proficiency assessments for college admissions. *Language Testing*, 40(2), 276–299. <https://doi.org/10.1177/02655322221112364>
- Isaacs, T., Hu, R., Trenkic, D., & Varga, J. (2023). Examining the predictive validity of the Duolingo English Test: Evidence from a major UK university. *Language Testing*, 40(3), 748–770. <https://doi.org/10.1177/02655322231158550>
- Jaehn, M., Hissbach, J., Frickhoeffer, M., Weppert, D., Zimmerhofer, A., Hampe, W., Kadmon, M., & Becker, N. (2025). Predictive validity of admission tests and educational attainment on preclinical academic performance: A multisite study. *BMC Medical Education*, 25, Article 1255. <https://doi.org/10.1186/s12909-025-07974-2>
- Jansen, R. S., van Leeuwen, A., Janssen, J., Kester, L., & Kalz, M. (2017). Validation of the self-regulated online learning questionnaire. *Journal of Computing in Higher Education*, 29(1), 6–27. <https://doi.org/10.1007/s12528-016-9125-x>
- Kintu, M. J., Zhu, C., & Kagambe, E. (2017). Blended learning effectiveness: The relationship between student characteristics, design features and outcomes. *International Journal of Educational Technology in Higher Education*, 14, Article 7. <https://doi.org/10.1186/s41239-017-0043-4>
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, 127(1), 162–181. <https://doi.org/10.1037/0033-2909.127.1.162>
- Kuryшева, A., Koning, N., Fox, C. M., van Rijen, H. V. M., & Dilaver, G. (2022). Once the best student always the best student? Predicting graduate study success using undergraduate academic indicators: Evidence from research masters' programs in the Netherlands. *International Journal of Selection and Assessment*, 30(4), 579–595. <https://doi.org/10.1111/ijsa.12397>
- Kuryшева, A., van Rijen, H. V. M., Stolte, C., & Dilaver, G. (2023). Validity, acceptability, and procedural issues of selection methods for graduate study admissions in the fields of science, technology, engineering, and mathematics: A mapping review. *International Journal of STEM Education*, 10, Article 55. <https://doi.org/10.1186/s40594-023-00445-4>
- Locke, L. F., Silverman, S. J., & Spirduso, W. W. (2009). *Reading and understanding research*. SAGE.
- Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly*, 49(2), 111–138. <https://doi.org/10.1177/001698620504900203>
- Martin, F., & Bolliger, D. U. (2018). Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. *Online Learning*, 22(1), 205–222. <https://doi.org/10.24059/olj.v22i1.1092>
- Martirosyan, N. M., Hwang, E., & Wanjohi, R. (2015). Impact of English proficiency on academic performance of international students. *Journal of International Students*, 5(1), 60–71. <https://doi.org/10.32674/jis.v5i1.443>
- Molontay, R., & Nagy, M. (2023). How to improve the predictive validity of a composite admission score? A case study from Hungary. *Assessment & Evaluation in Higher Education*, 48(4), 419–437. <https://doi.org/10.1080/02602938.2022.2093835>
- Neumann, H., Padden, N., & McDonough, K. (2019). Beyond English language proficiency scores: Understanding the academic performance of international undergraduate students during the first year of study. *Higher Education Research & Development*, 38(2), 324–338. <https://doi.org/10.1080/07294360.2018.1522621>
- Paris, J., & Heiser, R. (2022). An examination of the predictive validity of a measure of college admissions applicants' attributions of success and failure. *Journal of Postsecondary Student Success*, 1(4), 112–140. https://doi.org/10.33009/fsop_jpss130205
- Paris, J. H., Beckowski, C. P., & Fiorot, S. (2023). Predicting success: An examination of the predictive validity of a measure of motivational-developmental dimensions in college admissions. *Research in Higher Education*, 64(8), 1191–1216. <https://doi.org/10.1007/s11162-023-09743-w>

- Patac, A., Jr., Patac, L. P., & Caadan, J. G. (2024). Graduate school's impact on careers: Insights and implications from master of arts in education graduates. *Contemporary Educational Researches Journal*, 14(3), 154–167. <https://doi.org/10.18844/cerj.v14i3.9526>
- Pusat Layanan Tes Indonesia. (n.d.). *Tes Potensi Dasar Akademik (TPDA)*. <https://plti.co.id/tpda>
- Punongbayan, J. C. B. (2025). Balancing excellence and equity in predictive college admissions: Insights from the University of the Philippines. *Education Economics*, 33(5), 794–817. <https://doi.org/10.1080/09645292.2024.2421165>
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353–387. <https://doi.org/10.1037/a0026838>
- Rosita, T., Ramdhani, S., & Sapriati, A. (2023). Mapping student engagement: Analyzing the nexus of course engagement, resource utilization, and online tutorial participation. *Al-Ishlah: Jurnal Pendidikan*, 15(4), 5639–5649. <https://doi.org/10.35445/alishlah.v15i4.4475>
- Seyfried, M., & Hollenberg, S. (2024). Student selection in higher education: The organisational performance dilemma. *Journal of Higher Education Policy and Management*, 46(6), 671–686. <https://doi.org/10.1080/1360080X.2024.2365514>
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310. <https://doi.org/10.1214/10-STS330>
- Sorgente, A., Pietrabissa, G., Antonietti, A., Bonanomi, A., Castelnuovo, G., Lanz, M., Tagliabue, S., & Traficante, D. (2025). How to select candidates for an undergraduate degree in psychology? Combining high-school GPA and admission test score. *Asia Pacific Education Review*, 26, 459–471. <https://doi.org/10.1007/s12564-024-09978-x>
- Stemler, S. E. (2012). What should university admissions tests predict? *Educational Psychologist*, 47(1), 5–17. <https://doi.org/10.1080/00461520.2011.611444>
- Theobald, M. (2021). Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. *Contemporary Educational Psychology*, 66, Article 101976. <https://doi.org/10.1016/j.cedpsych.2021.101976>
- Tinto, V. (1997). Classrooms as communities: Exploring the educational character of student persistence. *The Journal of Higher Education*, 68(6), 599–623. <https://doi.org/10.1080/00221546.1997.11779003>
- Universitas Terbuka. (2025). *Katalog program magister Sekolah Pascasarjana Universitas Terbuka 2025/2026*. <https://www.ut.ac.id/wp-content/uploads/2025/06/04-Katalog-Program-Magister-Sekolah-Pascasarjana-Universitas-Terbuka-2025-2026-05062025.pdf>
- van der Zanden, P. J. A. C., Denessen, E., Cillessen, A. H. N., & Meijer, P. C. (2019). Patterns of success: First-year student success in multiple domains. *Studies in Higher Education*, 44(11), 2081–2095. <https://doi.org/10.1080/03075079.2018.1493097>
- Vos, S. G., Wijnia, L., Martens, R. L., & de Groot, R. H. M. (2025). How do distance learners self-regulate learning? A latent profile analysis of self-regulated learning and its relationship with self-regulatory efficacy, motivation, need satisfaction, need frustration, academic achievement, and well-being. *Learning and Individual Differences*, 120, Article 102678. <https://doi.org/10.1016/j.lindif.2025.102678>
- Vulperhorst, J., Lutz, C., de Kleijn, R., & van Tartwijk, J. (2018). Disentangling the predictive validity of high school grades for academic success in university. *Assessment & Evaluation in Higher Education*, 43(3), 399–414. <https://doi.org/10.1080/02602938.2017.1353586>
- Walters, T., Abeyta, A., Bean, A. J., & Wilson, M. A. (2022). The impact of holistic review on correlations between doctoral student outcomes, and GPA and GRE scores in the biomedical sciences. *PLOS ONE*, 17(12), Article e0279258. <https://doi.org/10.1371/journal.pone.0279258>
- Wilson, M. A., Odem, M. A., Walters, T., DePass, A. L., & Bean, A. J. (2019). A model for holistic review in graduate admissions that decouples the GRE from race, ethnicity, and gender. *CBE—Life Sciences Education*, 18(1), Article ar7. <https://doi.org/10.1187/cbe.18-06-0103>
- Xu, Z., Zhao, Y., Zhang, B., Liew, J., & Kogut, A. (2023). A meta-analysis of the efficacy of self-regulated learning interventions on academic achievement in online and blended environments in K–12 and higher education. *Behaviour & Information Technology*, 42(16), 2911–2931. <https://doi.org/10.1080/0144929X.2022.2151935>
- Zhang, Z. (2024). The impact of graduate education scale on the innovation capability of the tertiary industry. *International Journal of Education and Humanities*, 13(2), 7–10. <https://doi.org/10.54097/g5yb5d05>