INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT

OPEN ACCESS JOURNAL

Vol 15, Issue 4, (2026) E-ISSN: 2226-6348

# Academic Performance in Architectural Education: Learner Characteristics and the Measurement of Grades

Chih Siong Wong[1]*, Chu Sheng Ding[1], Lim Jin Wong[2], Shi Yee Wong[1], Prescilla Anak Palis[1]

[1]Centre for Built Environment and Resourceful Societies (BERi), and School of Built Environment, University of Technology Sarawak, Malaysia, [2]Centre on Technological Readiness and Innovation in Business Technopreneurship (CONTRIBUTE), and School of Business and Management, University of Technology Sarawak, Malaysia

*Corresponding Author Email: andywong@uts.edu.my

**DOI Link:** http://dx.doi.org/10.6007/IJARPED/v15-i4/29365

***Published Online:*** 18 November 2026

**Abstract**

Architecture programmes select and teach students on assumptions about which characteristics predict success. This integrative review examines which learner-level variables have been tested as predictors of architecture students' academic performance, what associations were reported, and how performance was measured. Scopus, citation and hand searching yielded 45 reports (39 independent samples, 15 countries, 1942–2026), appraised with Joanna Briggs Institute checklists. Earlier attainment was the most consistent predictor, strongest for aggregate grades and weakest for studio and final-project marks, though nine of the ten reports behind it were rated Low. Spatial ability had a modest association with early studio grades. Programme-specific admission tests, learning styles, intelligence, personality and emotional intelligence gave weak, null, cohort-dependent or statistically unsupported results, and the largest disposition coefficients arose where predictor and outcome were not independent. Thirty-nine reports drew on one institution, and none reached a high quality rating. No study reported grade reliability, and the relative weight of attainment, spatial ability and dispositions remains unestimated; fewer than half named who assigned grades, and accreditation context was rarely stated. Early marks forecast later marks, but the evidence cannot yet identify what underlies studio performance. Multi-institution studies measuring several predictors against studio grades of known reliability are needed.

**Keywords:** Architectural Education, Student Characteristics, Academic Achievement, Studio Grade Reliability, Integrative Review

## Introduction

Architecture programmes make two decisions about students before they can know how those students will perform. They select applicants on school-leaving or national admission scores and, in some schools, on drawing or aptitude tests of their own (García-Escudero et al., 2021; Gawlak et al., 2021; Krishna & Yalavigi, 1993), and they then teach much of the

curriculum through the design studio. Both decisions rest on assumptions about which learner characteristics predict success. Studios are organised around design projects and run through desk critiques, design reviews and juries; Schön (1987) took the architectural studio as the prototype of a reflective practicum, in which students learn by doing under coaching. Which students do well there is a question for selection as much as for teaching.

A research literature has grown around that question, but it has grown one predictor at a time. Admission and school scores, spatial and cognitive tests, learning-style inventories, personality and emotional-intelligence questionnaires, and measures of background and study conditions have each been set against architecture students' grades, usually in a single school. Some authors drew practical conclusions from one cohort: that learning styles should be tested early in the studio (Kvan & Jia, 2005), that models built on school grades could inform admission (Aluko et al., 2016, 2018), or that metacognitive training belongs in studio curricula (Yazici & Dogan, 2024). Findings conflict between studies. Emotional intelligence, for example, was unrelated to studio scores in one programme and associated with design marks in another (Nazidizaji et al., 2014; Şenkal Sezer, Erbil, et al., 2016). Studies differ in country, instrument and outcome, so no single one can show whether such conflicts reflect the construct, the sample or the grades.

The grades are a second difficulty. Studio marks and jury grades are judgements by tutors and critics. A literature that does not examine its grades cannot tell a weak predictor from a noisy outcome, so an account of what predicts performance in architecture needs an account of how performance was measured.

This review addresses three questions. The first asks which learner-level variables have been tested as predictors of academic performance in architecture education, and with which designs and instruments. The second asks what direction and size of association those studies reported, and where and why their findings conflict. The third asks how academic performance was defined and measured, and how reliable that measurement was. The design is integrative (Whittemore & Knafl, 2005). The scope is limited to architecture programmes, to characteristics of individual students, and to outcomes that are either grades or design tasks scored by raters. Teaching interventions and the physical environment were excluded, because the question is what students bring to and develop within a programme rather than what the programme does to them.

After the Method and Findings, the Discussion argues that the reviewed studies established neither the reliability of studio grades nor the relative weight of predictors measured together, and sets out what follows for admission, studio assessment, accreditation and further research.

## Method

*Review design*

An integrative review design was adopted (Whittemore & Knafl, 2005) because the eligible literature mixes correlational, predictive, longitudinal, quasi-experimental and qualitative studies whose results cannot be pooled statistically. A systematic review with synthesis without meta-analysis was considered, but that reporting guideline addresses intervention effects (Campbell et al., 2020), whereas this review concerns associations and includes

qualitative reports. The five stages described by Whittemore and Knafl (2005) were followed: problem identification, literature search, data evaluation, data analysis and presentation. No protocol was registered in advance. Rules settled during screening are reported with the record that prompted each, so that every eligibility decision can be traced.

*Search strategy*

Scopus was searched with two strings, both limited to English-language records. The first combined outcome terms (academic performance, achievement and success; student and studio performance; studio grades; grade point, grade point average (GPA) and cumulative GPA (CGPA)) with phrases naming architecture students, programmes, schools, design studios or architectural education, and returned 154 records. The second was written after the first had missed known studies whose abstracts never used the phrase “academic performance”. It paired a similar set of architecture phrases with predictor terms (spatial ability, cognitive and learning style, learning approach, personality, self-efficacy, motivation, metacognition, admission, entry qualifications, prior attainment, creativity) and broad outcome terms. It returned 260 records, 56 of which duplicated the first search. Three studies held from earlier work were added by hand.

The two strings were entered as follows. Main string (154 records):

TITLE-ABS-KEY ( ( "academic performance" OR "academic achievement" OR "academic success" OR "student performance" OR "studio performance" OR "studio grade*" OR gpa OR cgpa OR "grade point" ) AND ( "architecture student*" OR "architectural student*" OR "architectural education" OR "architecture education" OR "architectural design studio" OR "architecture program*" OR "school* of architecture" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) )

Supplementary string (260 records):

TITLE-ABS-KEY ( ( "architecture student*" OR "architectural student*" OR "architectural education" OR "architecture education" OR "architectural design education" OR "architectural design studio" OR "architecture program*" OR "school* of architecture" ) AND ( "spatial abilit*" OR "cognitive style*" OR "learning style*" OR "learning approach*" OR "approaches to learning" OR personality OR "self-efficacy" OR motivation* OR metacogniti* OR admission* OR "entry qualification*" OR "prior attainment" OR creativ* ) AND ( performance OR grade* OR achievement OR success OR progression OR attainment ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) )

Two forms of citation searching followed. In backward searching, the reference lists of the 30 reports then included and held in full text were read, and the 30 references whose titles suggested eligibility were screened. In forward searching, all records identified in Scopus as citing the included reports were exported and screened: 446 records, of which 403 had not already been screened in the database searches; five of these duplicated records already found by hand or backward searching. Citation searching was subsequently closed, and records that could not be obtained by then are reported as not retrieved.

*Eligibility criteria*

Studies were eligible if they met all of the following criteria.

- Participants: students on architecture programmes at degree or sub-degree (diploma) level. Programmes titled architectural engineering were eligible where they were the institution's architecture degree. Interior architecture, landscape architecture and Technical Architecture (a building-engineering degree in Spain) were excluded. Mixed-discipline samples were eligible only where architecture results were reported separately; where those results were given in prose alone, the study was retained for the direction of its findings and contributed no statistics.
- Predictor: a learner-level variable, such as prior attainment or admission score, spatial ability, cognitive or learning style, learning approach, personality, self-beliefs or motivation, study behaviour, demographic or social background, perceived learning environment, or wellbeing. Physical-environment variables and teaching interventions were excluded unless an individual-difference association with performance was also reported.
- Outcome: academic performance measured and analysed against the predictor (course or studio grades, grade point average, degree class, progression or jury marks), or a design or drawing task whose product was scored by raters. General creativity tests and students' ratings of their own performance were not accepted as outcomes; grades reported by students were accepted and are identified as self-reported.
- Design: any empirical design. Qualitative studies were eligible where they addressed why some students performed better than others.
- Publication: peer-reviewed journal articles and peer-reviewed conference papers, with no date limit. Theses, monographs, book chapters and conference abstracts were excluded.
- Language: an English-language abstract. Full texts in other languages were eligible, although the database limit meant that they could enter only through citation searching.

Five rules were settled during screening, each prompted by a record: sub-degree programmes were admitted for a study of National Diploma students (Adewale & Adhuze, 2014); the language rule was widened to English abstracts; physical-environment predictors were excluded as not learner-level; task outcomes were restricted to scored design or drawing work; and architecture results reported in prose were admitted for a four-discipline cohort (Crowther & Briant, 2021). Statistics in the two non-English texts were taken from their tables, and their text was read with machine translation.

*Study selection*

Records were screened on title and abstract, and potentially eligible full texts were assessed. Of the 46 reports assessed in full from the database and hand searches (one from its abstract, the only text available), 34 met the criteria. Citation searching sent a further 23 reports to full-text assessment, of which 11 were included; two further reports could not be retrieved. The final corpus comprises 45 reports. Figure 1 gives the counts at each stage with the reasons for exclusion.

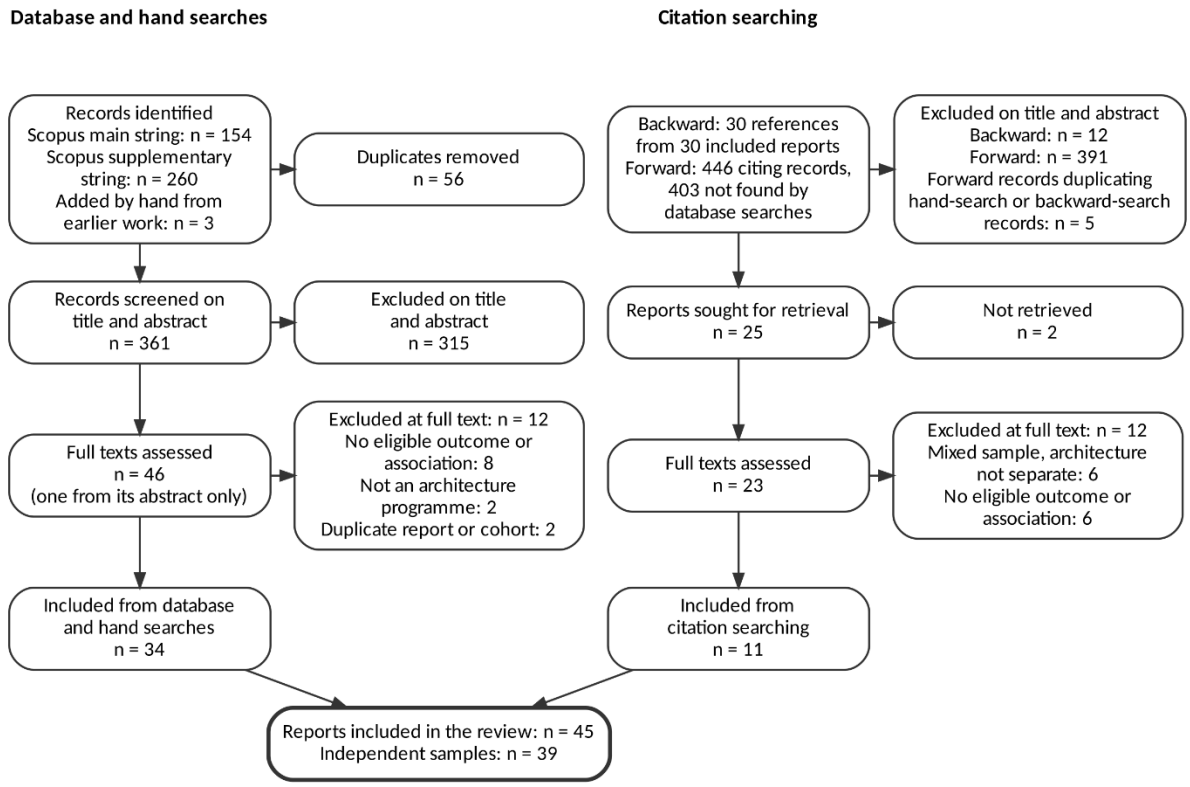



Figure 1 Flow of records through searching, screening and inclusion

Reports drawing on the same sample were counted once, which leaves 39 independent samples when probable overlaps are treated as shared. Aluko et al. (2016, 2018) analysed the same 101 graduates; Roberts (2006, 2007) drew on overlapping Cardiff cohorts; Peterson and Lansky (1980) followed the 1976 entering class of Peterson and Lansky (1977). Nazidizaji et al. (2014, 2015) sampled one population of 184 students, and Şenkal Sezer, Murat, et al. (2016) and Şenkal Sezer, Erbil, et al. (2016) one fourth-year cohort, and Crowther and Briant (2021, 2022) appear to have followed the same commencing cohort at one Australian university; these three overlaps are probable rather than confirmed. Yazici and Dogan (2026) appear to report on the cohort of Yazici and Dogan (2024): both describe 84 third-year students, and their pre-test regressions are almost identical (R = .861 and .862). It was not counted as a separate report but is cited for what it discloses about that cohort.

Screening, extraction and appraisal were conducted by the first author, who checked every inclusion decision, every statistic cited in the text and the extracted data for a sample of studies against the source documents. The second author, working independently of the first author's decisions and extraction, screened a stratified random sample of ten full-text reports (seven included and three excluded at first screening) and re-extracted data from seven included studies. Agreement on eligibility was complete (10/10; $\kappa = 1.00$). Both reviewers recorded the same predictor group, sample size and outcome-reliability status for all seven studies, and the same outcome type for six. Agreement was lower on individual items of the Joanna Briggs Institute (JBI) checklists (33/56, 58.9%; $\kappa = .41$; 33/49 and $\kappa = .50$ without item 4, later treated as not applicable) and on the overall quality band (4/7, 57.1%; $\kappa = .22$). Most

item-level disagreement concerned items whose application to grade-based outcomes was ambiguous. Decision rules were agreed for these items and applied to both reviewers' ratings and to all included studies, and the one remaining discrepancy was resolved by discussion, so agreement after the rules reflects consensus rather than independent rating.

*Data extraction and quality appraisal*

A standard form recorded the following for each study:

- bibliographic details;
- country, institutions, programme and any accrediting body named;
- year of study and sample size;
- design;
- the predictor and its instrument, with any reliability reported for the study's own sample;
- the outcome measure, who assigned it and any reliability reported for it;
- the analysis and key results as printed, with page numbers;
- the authors' conclusions and stated limitations.

Quality was appraised with the JBI checklist matching each design: analytical cross-sectional (2020 version, eight items; Moola et al., 2020) or qualitative (ten items; Lockwood et al., 2015). The cross-sectional checklist was also applied to the quasi-experimental study, whose predictor analysis was correlational, and to a case study, classed as qualitative, that compared grades across personality types. Three decision rules fixed how the cross-sectional checklist was applied. Item 4, on objective criteria for measuring the condition, was treated as not applicable, because item 7 already appraises how the outcome was measured. Item 3, on valid and reliable measurement of the exposure, was met when the predictor came from institutional records or its reliability was reported for the study's own sample, was unclear when an established instrument carried reliability from other samples only, and was not met for ad hoc measures. Item 6, on strategies for confounding, was met only where confounders were adjusted for in the design or the analysis. A study was rated High where at least 75% of applicable items were met, Moderate at 50–74% and Low below 50%. Ratings were used to qualify findings rather than to weight them formally, and no study was excluded on quality.

Two further checks, which formed the verification stage of the analysis, were applied to every included study. Where a paper printed enough to recompute a test, such as a chi-square value with its degrees of freedom, the p-value was recalculated and compared with the authors' interpretation. The conclusions stated in each abstract were compared with the results in the paper's own tables. Three studies could not be obtained as files and were read online (Avsec & Jagiełło-Kowalczyk, 2021; Ayalp & Erman, 2019; Elgazzar et al., 2019), and one was available only as an abstract (Tastan et al., 2026).

*Synthesis*

Analysis followed the sequence of data reduction, data display, data comparison, and conclusion drawing and verification described by Whittemore and Knafl (2005, p. 550). Each study was coded by the predictor it tested, in ten groups, and the groups were gathered into five themes: prior attainment and admission; spatial ability, intelligence and cognitive style; learning style and approach; personality, emotional intelligence, motivation and self-regulation; and background, study conditions and wellbeing. Studies testing several predictors were allocated by their principal predictor and cited elsewhere where relevant. A

condensed display matrix of the included reports is given in Appendix A, and Table 1 summarises it by theme.

Comparison ran in two directions. Within themes, agreement and divergence were located, and divergence was traced to sample, instrument or institution where the reports allowed. Across themes, the size of associations was compared wherever one study had measured more than one predictor. Effect sizes are reported as printed, not converted to a common metric, because many studies reported only significance tests. Task outcomes were reported separately from grades within each theme, and evidence reported in prose alone was used for the direction of findings only.

For the third review question, each study's outcome was coded for what was measured, who assigned the grade or score, whether any reliability was reported for it, and whether the accreditation framework governing the programme was stated.

**Findings**

*The evidence base*

The 45 reports span 1942 to 2026, but most of the literature is recent: 30 appeared from 2015 onwards and 21 from 2020. They come from 15 countries, led by Turkey (nine reports), the United States (seven), and Nigeria and Egypt (five each); the one report available only as an abstract did not name its country. The geographical spread is wide. The institutional spread is narrow. Thirty-nine reports drew on a single institution, two were unclear, and only four sampled more than one (Adewale & Adhuze, 2014; Elgazzar et al., 2019; Iqbal & Roberts, 2019; Sen et al., 2021), none of them more than four schools. Every association reported below therefore describes a handful of programmes at most, usually one cohort in one school. Designs were overwhelmingly observational. Twenty-seven reports were cross-sectional and 11 followed cohorts over time; two built predictive models, two combined methods, two were qualitative and one was quasi-experimental, although its predictor analysis was correlational (Porat & Ceobanu, 2024). Samples were small. Among the 43 reports that gave a sample size for architecture students, the median was 90 (range 17–3,910), 23 analysed fewer than 100 students and eight fewer than 50. The largest, 3,910 graduates of one Spanish school (García-Escudero et al., 2021), is more than eight times the size of the next (484; Peterson & Lansky, 1977), so the few stable estimates in the corpus come from very few sources.

Table 1 distributes the reports across the five themes. Forty used academic outcomes and five a scored design task. One study contributes narrative-only evidence, because its architecture results were reported in prose without a sample size (Crowther & Briant, 2021). No report reached a High quality rating; 10 were rated Moderate, 34 Low, and one could not be appraised. The Low ratings cluster on three checklist items. Of the 43 reports appraised with the cross-sectional checklist, eight identified confounding variables and one reported estimates adjusted for them; two measured the outcome in a way that could be judged valid and reliable; and 11 used analyses judged appropriate to their data. Because grade-based studies could rarely meet item 7, the ratings were recomputed without it: one report would then be rated High, 24 Moderate and 18 Low. The weakness is structural rather than incidental. Most reports set one predictor at a time against grades whose reliability was never

examined, so the synthesis that follows gives more weight to the direction and consistency of findings than to the size of any single coefficient.

Table 1
*Characteristics of the included reports by theme*

| Theme | Reports (samples) | Countries | Designs | Outcome type | Quality (Moderate/Low) | Median n (range) |
|---|---|---|---|---|---|---|
| Prior attainment and admission | 10 (9) | 8 | 4 cross-sectional, 4 longitudinal, 2 predictive | 10 academic | 1/9 | 110 (80–3,910)[a] |
| Spatial ability, intelligence and cognitive style | 11 (9) | 5[b] | 6 cross-sectional, 3 longitudinal, 1 mixed, 1 quasi-experimental | 10 academic, 1 task | 1/9[b] | 77 (17–484) |
| Learning style and approach | 5 (5) | 5 | 4 cross-sectional, 1 longitudinal | 4 academic, 1 task | ¼ | 75 (33–349) |
| Personality, emotional intelligence, motivation and self-regulation | 11 (11) | 5 | 8 cross-sectional, 1 longitudinal, 1 mixed, 1 qualitative | 8 academic, 3 task | 4/7 | 78 (18–237) |
| Background, study conditions and wellbeing | 8 (8) | 6 | 5 cross-sectional, 2 longitudinal, 1 qualitative | 8 academic | 3/5 | 118 (43–273)[c] |
| All reports | 45 (39) | 15[b] | 27 cross-sectional, 11 longitudinal, 7 other | 40 academic, 5 task | 10/34[b] | 90 (17–3,910) |

Note. Samples are counted within each theme. Three cohorts appear in two themes, so the theme totals (42) exceed the 39 independent samples. [a]Excludes Crowther and Briant (2021), which gave no architecture sample size. [b]One further report was available only as an abstract; its country was not stated and it was not appraised. [c]Excludes Iqbal and Roberts (2019), whose participants were 14 teachers.

*Prior attainment and admission*

School-leaving and admission scores were the most consistent predictors in the corpus, but their reach weakened the closer the outcome came to the studio. The clearest evidence is a census of 3,910 graduates of one Spanish school, in which the university admission score correlated at r = .534 with the overall architecture grade, between .319 and .470 with module grades, and at .192 with the final degree project (García-Escudero et al., 2021). Individual design courses sat lower still: Project Design I and II gave r = .17 and .19, against module coefficients of .470 for Technology and .467 for Theory. Aggregation may account for part of

this gradient, since a grade averaged over many courses is more reliable than a single course grade (Beatty et al., 2015), and at module level Project Design (.347) sat above Representation (.319). In four Nigerian polytechnics, O-level grades in physics and mathematics correlated at r = .459 and .423 with second-year CGPA (n = 128), although the authors read these coefficients as a weak relationship and an insignificant contribution despite t values of 5.80 and 5.24 (Adewale & Adhuze, 2014). Secondary-school examination points tracked graduates' overall grades in Poland at ρ = .47, .46 and .26 across three graduating years (Gawlak et al., 2021). Crowther and Briant (2021) stated that entrance rank explained about 10% of GPA variance in architecture but gave no coefficients.

Programme-specific admission tests did worse. Against the same Polish graduates, the drawing entrance examination correlated with the overall grade at ρ = .01, .27 and .09; the authors reported a mean "below 0.03", although the three yearly coefficients average .123 (Gawlak et al., 2021). Against first-semester drawing grades the examination correlated at about .3, falling to .11 by the third and fourth semesters, and about 30% of those admitted would not have been admitted on school results alone. An Indian departmental aptitude test was significantly associated with 12 of 20 course grades in a class of 19, but with four of 18 and two of 16 in the classes of 26 and 35 that followed (Krishna & Yalavigi, 1993). In Bratislava, the admission tasks overlapped with design-studio grades by between 19% and 43% on a non-standard "correlation rate" that matched grade levels, and the authors concluded that the tasks did not predict design performance (Hubinský et al., 2022). These results diverge by cohort more than by instrument. The number of significant Indian associations fell as class size grew, and the Polish and Slovak studies, like the Spanish census, analysed graduates only, which removes the weakest entrants and compresses the range a predictor can explain (Sackett & Yang, 2000). Gawlak et al. (2021) noted that two committees marked the drawing examination independently, yet no study reported agreement between the markers of its admission test, so the unreliability of the predictor cannot be separated from its lack of validity.

Early programme grades predicted later grades more strongly, for reasons that are partly mechanical. At one Indonesian school, the grade in a first-semester drawing course correlated with GPA at ρ = .75, .51 and .70 in three successive classes and at .54 when they were pooled (Lionar et al., 2025). The pooled coefficients with individual studios fell from .56 in Studio 2 to between .16 and .26 in Studios 5 to 8. The GPA contains both the predictor course and every studio grade, so part of the association is the same marks counted twice. An Egyptian study points to a related effect from a different angle (Rabboh, 2021). Overall sophomore studio scores correlated with overall graduation-project scores at ρ = .313 (n = 96), and coursework marks with later coursework marks at .354, while sophomore final-term marks bore no relation to graduation-project final-jury marks (ρ = −.085, p = .408). Rabboh (2021) taught and graded the sophomore studios that supplied the predictor. The cross-type coefficients were also positive (ρ = .224 and .256), so continuity of coursework assessment explains part of the pattern at most; the clearest contrast is the null between the two final assessments.

The two predictive-modelling studies, which analysed the same 101 Nigerian graduates, make claims that go beyond their results. Aluko et al. (2016, 2018) reported accuracies of 73.33% (κ = .318) and 76.67% for k-nearest-neighbour and support-vector models predicting pass or

fail from O-level grades. Twenty-two of the 30 test cases passed, so predicting a pass for every student would have scored 73.3%. Both reports nonetheless recommended their models for admission decisions.

Taken together, prior attainment was most closely associated with aggregate grades, less with module grades and least with studio, jury and final-project marks, while specialised admission tests gave small and cohort-dependent coefficients. Aggregate outcomes include courses that resemble school examinations, five of the nine samples were restricted to completers, and only one study was large enough to estimate a coefficient precisely. None tested whether the studio grades it predicted were reliable.

*Spatial ability, intelligence and cognitive style*

Spatial ability was the most consistently positive predictor outside prior attainment, but the size of its association depended on the sample more than on the construct. Among first-year grades, the smallest coefficients came from the largest study. Across three Egyptian schools (n = 353), two spatial tests correlated with first-year design grades at between r = .135 and .264, coefficients the authors described as strong (Elgazzar et al., 2019). The largest came from small groups. In one Israeli programme (n = 75, of whom 65 to 67 were analysed), a spatial test taken at entry correlated with first-year course grades at r = .37 to .47 and with final first-year GPA at .55, and with course grades and GPA at .46 to .67 among the 33 students who received no spatial training (Porat & Ceobanu, 2024). The test was not named, and the full-sample coefficients pool students who later received the training. Older studies fall between these poles. Among 75 male architecture freshmen at an American art school, the Minnesota Paper Form Board correlated at .33 with the average art grade, .25 with the design grade and .17 with the structural representation grade (Bryan, 1942). Among 17 graduating students, a cube-comparison test correlated at .49 with jury-graded independent design work, while a surface-development test correlated at .15 (Karlins et al., 1969). Against these, a mental rotation test gave r = .11 with third-year design rank in one Cardiff cohort (n = 47; Roberts, 2007).

Two features of the studies account for much of this spread. The first is timing: coefficients were mostly higher against first-year grades than against later marks, the jury-graded independent work in Karlins et al. (1969) being the exception, which suggests that spatial ability eases entry to studio work more than it orders students at the end of it. The second is the instrument. Cho (2012) reported that a studio grade among 21 freshmen was associated with an architecture-specific spatial test but not with three general spatial tests, although the table on which the claim rests appears to print significance levels in place of correlation coefficients. Karlins et al. (1969) found that two general tests diverged against the same outcome (.49 and .15), so even general measures are not interchangeable. With samples of 21 and 17, the domain-specific advantage is a hypothesis the corpus raises, not a finding it supports.

General intelligence added little. Bryan's (1942) general ability test correlated at .37 with the average art grade but at .22 with the design grade and .06 with the structural representation grade. A general intelligence test correlated at .07 to .11 with grades, class rank and independent design work in the graduating group (Karlins et al., 1969). In an Iranian programme, Raven's Progressive Matrices scores were unrelated to the mean of five design-

studio scores (r = .13) and to GPA (r = −.15), and correlated only with the final studio (r = .26; Nazidizaji et al., 2015). That single significant result was one of seven tested without correction. The authors' explanation, that intelligence matters more as design problems grow complex, was not tested.

Cognitive style, measured as a processing preference rather than an ability, had no stable relation with design marks. In three Cardiff cohorts, wholist-analytic and verbaliser-imager ratios correlated with third-year design rank at between .01 and .20, none significantly (Roberts, 2007). An earlier report on overlapping cohorts found that analytic students ranked higher in the first year in one of two cohorts, an advantage that had disappeared by the third (Roberts, 2006). In the later report, style was associated with leaving the programme, with verbalisers over-represented among the 46 leavers (Roberts, 2007), but that comparison was not adjusted for entry route or attainment. Roberts (2007) is also one of three studies, with Bryan (1942) and Karlins et al. (1969), to set several families of predictor against the same students' recorded marks. In the Cardiff cohorts, A-level points were as unrelated to design rank as style and rotation were (r = .16, .02 and .05 in three cohorts), although students who entered without A-levels left the programme more often (46% against 18%, $p < .01$).

Handedness entered the literature as a proxy for lateralised visual thinking. Of 484 students at one American school, 73.4% of left-handers graduated against 62% of right-handers ($p < .05$; Peterson & Lansky, 1977). In a later class, students who drew a cube in isometric form at the start of their second year had higher GPAs (2.91 against 2.67) and design grades (2.60 against 2.26) after six quarters (Peterson & Lansky, 1980), but the drawing followed a year of instruction, and the printed dropout statistic (G = 17.91) recomputes from the printed counts to 5.07.

The one task outcome in the theme, available only as an abstract, reported that spatial ability was more strongly associated with conceptual design tasks than with digital modelling (Tastan et al., 2026). Read as a whole, the theme supports a modest association between spatial ability and early studio grades and little association for general intelligence or cognitive style. Only one study adjusted spatial ability for school attainment, and it reported no coefficients for that model (Porat & Ceobanu, 2024), so whether spatial tests add anything to the grades already used in admission remains unknown.

*Learning style and approach*

This theme produced the least stable associations in the corpus. In each of its four grade studies, at least one printed statistic fails to support the conclusion the authors drew from it. Kvan and Jia (2005) reported a significant relation between Kolb styles and final studio marks in two programmes at one Chinese school, but their chi-square values (19.929 on 18 degrees of freedom; 24.715 on 21) give $p \approx .337$ and .260, and the style with the lowest marks differed between the two cohorts. Tucker (2009) found no difference in first-year design marks across the four Kolb styles. A recoded contrast favoured "Southerner" learners (M = 66.963 against 64.086; $t(106) = -1.827$, one-tailed $p = .0355$), yet it failed the Bonferroni threshold of .003 that Tucker had set, and third-year students completed the inventory two years after the marks it was compared with. Ayalp and Erman (2019) reported that convergers outperformed other styles in three construction management courses, but the paper prints p-values without test statistics. In one course the four style means lay between 3.00 and 3.33, with standard

deviations of 0.75 to 1.00, and no allocation of 55 students to four styles yields more than F = 0.52 (p = .67) from those figures, against the printed p = .000. The courses were technical, not studio, courses, 17 of the 72 students first surveyed did not take part throughout, and the authors argued that style itself changes through the programme in which they measured it. The one task-outcome study is no firmer. In a single third-semester section of 33 women at a Delhi university, accommodators received the highest ratings from three studio teachers for their design projects, as they did on a creativity test (Taneja et al., 2023). Agreement among the teachers was not reported. The analysis of variance (F = 167.48) does not state which score it analysed, and its critical value of 6.59 implies four error degrees of freedom, which suggests that it was run on aggregated scores rather than on students.

Learning approach, a deep or surface orientation to study distinct from style, produced the theme's most informative result. Across two Nigerian universities (n = 349), university type and gender explained 2.9% of the variance in CGPA, and adding approach scores raised this to 4.4% (Sen et al., 2021). The deep approach was positively related to CGPA (β = .229). The surface approach was not (β = −.134, p = .773, a p-value hard to reconcile with that coefficient at n = 349), although the abstract and conclusion report a significant negative effect. Students reported their own CGPA, in bands.

Kvan and Jia (2005) and Tucker (2009) both justified style assessment by appeal to matching instruction to learners. That claim can be tested only by assigning students of each style at random to matched and mismatched teaching and showing that each group does best under its own method (Pashler et al., 2008). No study in the corpus allocated students in that way. The pedagogical recommendation that motivates much of this literature has therefore not been examined in any of the studies reviewed here.

*Personality, emotional intelligence, motivation and self-regulation*

The dispositions in this theme were measured almost entirely by self-report, and they produced both some of the largest coefficients in the corpus and some of its clearest nulls. What separated the two was less the construct than whether predictor and outcome were independent.

Emotional intelligence was tested in three samples, with divergent results. Among 78 Iranian students, total emotional intelligence was unrelated to the mean of five design-studio scores (r = .05) and to GPA (r = .02; Nazidizaji et al., 2014). The one significant result among 30 correlations, self-management with the fifth studio (r = .26), is about what chance would yield. Among 35 fourth-year students in Turkey, emotional intelligence correlated with the four-year average of design-class marks (r = .375) but not with weighted GPA (r = .192; Şenkal Sezer, Erbil, et al., 2016). Students who had never failed a design class scored higher (t(33) = 2.515, p = .017). About 60 tests were run without correction, and several associations the authors treated as significant reached only p < .10. In a first-year class of 26, emotional intelligence correlated with term grades at r = .527 (Birer, 2012). Instruments and outcomes also differed, from a 28-item appraisal against single studios in Iran to two versions of the Bar-On inventory against a four-year design average and one term's grades in Turkey. The two positive results came from the two smallest samples.

Birer (2012) also illustrates the hazard that recurs across the theme. Achievement orientation, scored weekly by the project coordinators who assigned the grades, correlated with those grades at r = .836. The two measures share a source. The same problem appears in the theme's largest model. Among 84 third-year students at one Turkish school, two components of metacognitive awareness correlated with design-studio grades at r = .741 and .699, and with three goal orientations they explained 74.4% of the variance (Yazici & Dogan, 2024). A later report, apparently on the same 84 students, states that 58 of them had received a metacognitive intervention in the semester in which the grades were awarded, and that their sections outscored the control section ($F(1, 82) = 4.242$, $p = .043$; Yazici & Dogan, 2026). The 2024 paper reports no intervention, although the later report describes that study as using minimal prompts. If the cohorts are the same, its coefficients pool students who were and were not taught the strategies the inventory measures, with no term for the difference.

Traits, type classifications and self-rated creativity fared worst against grades. Tolerance of ambiguity was unrelated to design-project grades among 237 Egyptian first- and second-year students ($r = −.002$; Mahmoud et al., 2023) and in an earlier group of 18 from the same school ($r = −.036$; Mahmoud et al., 2020). Divergent thinking was also unrelated to grades in the larger group ($r = .054$). The earlier study had planned a juried design task as its outcome and abandoned it after judging that the task did not reflect the true performance of about 40% of students; its participants were also selected by prior design grade. An Enneagram study of 20 students, two of its eight types represented by one student each, claimed without any statistical test that personality type affected studio performance, and described grading both as blind to type and as conducted with type in view (Aydoğan & Özdemir, 2026). Self-rated creativity did not differ across self-reported GPA bands among 90 respondents, the lowest band holding one student (Bircek & Ayalp, 2022).

The three task-outcome studies include the only two in the corpus that reported agreement between the raters of an outcome. Four professors rated 71 Egyptian students' studio projects for creativity, and intuitive types were more often rated creative than sensing types ($p = .031$; Hemdan et al., 2023). That was one of five tests, it does not survive a Bonferroni threshold of .0125 for the four dichotomies, and the projects came from two different briefs. Among 90 first-year American students, extrinsic motivation correlated negatively with creativity ratings ($r = −.4714$; Watters & Spas, 2020). The printed p of .04 does not fit that coefficient ($p < .001$ at $n = 90$), and the regression degrees of freedom imply about 75 cases. In Poland, proactivity was the strongest positive predictor of creative design scores among 98 senior students, although the paper did not say who rated the designs or how far raters agreed, and the coefficients appear only in figures (Avsec & Jagiełło-Kowalczyk, 2021).

Across the theme, the largest associations arose where the predictor was rated by the graders or the grades followed an intervention aimed at the predictor. In the one sample of more than 200 in which an independent self-report met ordinary studio grades, the association was close to zero. No study measured a disposition alongside prior attainment, and the one that set a disposition beside spatial ability found it unrelated to grades (Bryan, 1942), so none can show that dispositions add anything to what is already known at admission.

*Background, study conditions and wellbeing*

This theme holds the corpus's weakest measurement. Two of its eight reports relied on grades that students reported themselves, one correlated group means, and one collected no student data at all. Its clearest finding concerns gender. In the one study to report both kinds of outcome, gender distinguished aggregate grades from design grades, as prior attainment did.

At one Turkish school, women had a higher weighted GPA than men (2.61 against 2.31; $t(53) = 2.714$, $p = .009$) but not a higher design average (70.49 against 68.89; $p = .405$), and age correlated with GPA ($r = −.306$) but not with design marks (Şenkal Sezer, Murat, et al., 2016). Men repeated design classes more often (12 of 17 against 14 of 38; $\chi^2(1) = 5.366$, $p = .021$), although the authors wrote that repetition was the same for men and women. In the smaller and probably overlapping sample of 35 fourth-year students, grades in design, structure, building and general classes did not differ by gender (Şenkal Sezer, Erbil, et al., 2016). No test statistics were given, so this result cannot be weighed against the GPA difference. In the architecture major of an Australian design degree ($n = 169$), women's first-year GPA advantage (4.91 against 4.64) had gone by graduation (5.24 against 5.09; Crowther & Briant, 2022). Gender was associated with graduation-project scores in an Egyptian sample ($F = 4.050$, $p = .047$), although Rabboh (2021) reported no difference, and studio scores did not differ by gender in an Iranian programme (Nazidizaji et al., 2014).

Evidence on social background is thin. Among 43 third-year students, Özyaba et al. (2013) named five socio-economic variables as affecting GPA, yet two of them (gender, $p = .123$; mother's education, $p = .701$) were non-significant in the authors' own table, and eight others with $p < .05$ went undiscussed. A chi-square of 4485.333 for birth year in a sample of 43 implies that the tests are not interpretable. In four Lahore schools, most of 14 design teachers held that students' social background shaped their studio performance, chiefly through communication skill and English, whereas the three teachers at the school with the most advantaged intake saw no such effect (Iqbal & Roberts, 2019). The study recorded teachers' beliefs, which cannot establish the authors' conclusion that cultural capital dictates students' learning paths.

Research on study conditions and wellbeing reported associations its analyses could not carry. Time spent in studio and in paid work was associated with lower self-reported GPA among 224 American students ($\beta = −.20$), but the paper states in successive sentences that self-efficacy was and was not associated with GPA (Bachman & Bachman, 2006). Self-reported weekly study hours correlated with final grades at $r = .95$ in a Spanish fourth-year course, a coefficient computed on four group means rather than on 118 students (Sirvent-Pérez et al., 2025). Satisfaction of basic needs differed across GPA bands among 82 Turkish first-year students, but not monotonically, in five uncorrected tests, with bands of 10 and 13 students, and with data that may predate the students' first GPA (Kilicaslan, 2020). In a Nigerian school, age ($\beta = −.32$) and perceptions of assessment and schedule were associated with self-reported CGPA, but the sign for fairness differs between text and table, and the results and discussion give opposite readings of the assessment effect (Oluwatayo et al., 2015).

The theme's value lies less in its findings than in what they imply for the others. Gender, age and social background were tested here as predictors, yet only two studies in the other themes entered any of them as covariates, and one reported no adjusted coefficients (Porat & Ceobanu, 2024; Sen et al., 2021). Associations reported elsewhere may therefore partly reflect who enrolled.

*How performance was measured*

The outcome side of this literature was taken as given. Of the 40 reports with academic outcomes, 25 analysed at least one design or studio grade, 12 used only aggregate grades such as GPA or the overall degree grade, two used grades in single non-studio courses, and one recorded teachers' perceptions without any grades. Four of the aggregate-grade reports relied on GPA reported by the students themselves (Bachman & Bachman, 2006; Bircek & Ayalp, 2022; Oluwatayo et al., 2015; Sen et al., 2021), in one case because records were hard to obtain, and none reported checking the figures against records.

Who assigned the grades was usually not reported. Of the 35 reports that used recorded grades, 15 said who had set them, and their procedures had little in common. A single instructor graded one house project (Cho, 2012) and one studio (Aydoğan & Özdemir, 2026). Kvan and Jia (2005) combined the instructor's mark (40%) with those of three external critics at a public jury (60%), after which a coordinator normalised the distribution, capping A and F grades at 10% of students each. Three teachers marked each assignment independently before moderation in Tucker (2009); boards of examiners confirmed marks set by panels of critics in Roberts (2006, 2007); and studio work in Bratislava was assessed by peer-to-peer review, in the authors' term (Hubinský et al., 2022). The weightings differed as much as the judges. A final presentation carried 50% of one design grade (Yazici & Dogan, 2026), and in-class activities 70% of another (Sirvent-Pérez et al., 2025). A "studio grade" therefore names a different judgement in almost every study.

None of the 40 academic-outcome reports gave any reliability for its grades. Several authors acknowledged the problem without measuring it. Roberts (2007) converted marks to within-cohort ranks, citing earlier evidence that design marks are unreliable; Tucker (2009) described the marking as subjective; and Nazidizaji et al. (2015) noted doubts about whether studio scores represent design ability. The one account of how graders' disagreement was handled came from a report excluded as a probable duplicate cohort, in which high variance between assessors prompted a consensus re-evaluation of the marks (Yazici & Dogan, 2026). Karlins et al. (1969) reported agreement between two faculty members' ratings of students' creativity (.63 and .88), but not for the jury grades they analysed.

The five task outcomes fared better, but only partly. Two reported agreement among raters: an average-measures intraclass correlation of .82 for four professors (Hemdan et al., 2023), and an alpha of .876 with a single-rater intraclass correlation of .542 for seven graduate students (Watters & Spas, 2020). Avsec and Jagiełło-Kowalczyk (2021) reported the internal consistency of their outcome's subscales (alpha .89 to .92) without saying who rated the designs. Taneja et al. (2023) reported no agreement for their three teachers, and the abstract of Tastan et al. (2026) did not describe scoring.

The missing reliability matters for interpretation more than for the size of coefficients. Classical test theory caps a predictor-outcome correlation at the square root of the outcome's reliability (Schmitt, 1996). At the single-rater value of .542, that ceiling is about .74. Schmitt (1996) also noted that even low criterion reliability, of around .50, does not seriously attenuate validity coefficients, so unless studio grades are far less reliable than that, unreliability alone cannot explain the near-zero results for dispositions. What unmeasured reliability prevents is comparison. In the Spanish census, the admission score correlated at .534 with the overall grade and at .192 with the final degree project (García-Escudero et al., 2021). Weaker prediction of design work and less reliable design marks would produce the same pattern, and no study in the corpus allows the two to be separated. Grades were also compressed. One studio grade ranged only from 72 to 88 (SD = 3.582; Cho, 2012), and design grades in Bratislava had a standard deviation of 0.26 on a scale running from 1.0 to 3.0 (Hubinský et al., 2022). Compression of this kind restricts the variance any predictor can explain (Sackett & Yang, 2000).

The accreditation context of the programmes was almost never stated. Two reports named a quality-assurance body that had accredited the programme (Aluko et al., 2016; Gawlak et al., 2021), one described its schools as accredited without naming the body (Iqbal & Roberts, 2019), and others cited national curricula or regulators (Adewale & Adhuze, 2014; Nazidizaji et al., 2014; Sen et al., 2021) or discussed professional accreditation in general terms (Hubinský et al., 2022). None named the professional body that validates the programme for registration. None of the seven American reports stated whether its programme was accredited by the National Architectural Accrediting Board, and Roberts (2007) cited the Royal Institute of British Architects only as disciplinary context. No study related its outcome to a validating body's criteria. The corpus therefore cannot say whether the grades it analysed reflect what the programmes were accredited to teach.

**Discussion**

The most consistent finding in the corpus is that earlier marks forecast later marks, most strongly when both are aggregates and least for studio and final-project marks, although nine of the ten reports behind this finding were rated Low. It holds across higher education generally (Richardson et al., 2012); what this corpus adds concerns the studio grade. Beyond that the evidence is modest, weak or null. The more consequential finding concerns what the studies did not measure. No study estimated the reliability of the grades it set out to predict. Only three reported coefficients for more than one family of predictor against the same students' recorded marks (Bryan, 1942; Karlins et al., 1969; Roberts, 2007). None of the three adjusted one predictor for another, and only Bryan (1942) included a disposition, a neuroticism score unrelated to grades. The relative weight of entry attainment, spatial ability and dispositions in studio performance therefore remains unestimated.

The gradient from aggregate to studio grades admits three readings that the corpus cannot separate. Studio work may draw on capacities that school examinations do not capture, in which case better predictors are worth seeking. Studio grades may instead be less consistent than course grades, in which case any predictor will look weaker against them. Aggregates may also look stronger simply because they average many grades. Juries have long been contested by those who take part in them: most of the students, faculty and practising architects whom Anthony (1987) questioned believed that juries needed improvement. Yet

none of the 45 predictor studies reviewed here, spanning 84 years, asked how far two juries would agree on the same work.

For admission, the evidence favours school-leaving and admission scores as predictors of overall progress and gives no support for screening applicants by learning style, personality type or emotional intelligence. Programme-specific drawing and aptitude tests gave small coefficients that varied by cohort, and the evidence supports neither keeping nor dropping them. No study tested portfolios or interviews, and none of the nine Turkish reports examined prior attainment. Many of the favourable coefficients came from graduates or completers, and the validity of an admission measure can be judged only on everyone it admitted.

For studio assessment, the implication is direct. A school cannot know what predicts studio performance until it knows how consistent its studio grades are, and the methods exist. Two task-outcome studies in the corpus estimated rater agreement with intraclass correlations (Hemdan et al., 2023; Watters & Spas, 2020), and research on studio grading outside this literature has used a Rasch model to evaluate tutors' grading (Hassanpour et al., 2012) and compared comparative judgement with rubric marking in admission (van den Heuvel & Bohm, 2023). Recording independent marks from more than one assessor before moderation would let a school estimate the agreement of its own juries, although juries are teaching events as well as assessments and can reach marks by consensus or normalisation (Kvan & Jia, 2005; Yazici & Dogan, 2026), so doing so would change practice rather than simply document it. For accreditation, the corpus offers almost nothing. Studies across schools under different accreditation regimes would show whether the associations reported here depend on how programmes are validated, for example whether external examining yields more consistent studio grades.

Further research needs multi-institution cohorts that measure prior attainment, spatial ability and at least one disposition at entry, so that each can be estimated net of the others; outcomes reported separately for studio and non-studio courses, with studio-grade reliability estimated from independent marks; analyses that retain leavers, correct for multiple tests and report effect sizes with confidence intervals; and a statement of each programme's accreditation regime and grading procedure. Testing the matching of teaching to learning style also needs random allocation to matched and mismatched teaching (Pashler et al., 2008).

*Limitations*

Only Scopus was searched, and its English-language limit removed non-English documents before screening. Eleven of the 45 reports were found only by citation searching, several on predictors the strings did not name, such as emotional intelligence and intelligence, which points to limited sensitivity in the strings; citation searching was a single round. Screening, extraction and appraisal were conducted by the first author, and exclusions at title and abstract were not checked by a second person. The first author checked every inclusion decision and every statistic cited, and the second author, working independently, agreed fully on eligibility but less well on quality bands before decision rules were agreed ($\kappa = .22$). Three studies were read online, and one only as an abstract. The 2020 JBI cross-sectional checklist was retained, rather than its revised risk-of-bias successor (Barker et al., 2026), because appraisal and the agreement check had been completed with it; several of its items fit grade-

based outcomes imperfectly, which is why decision rules were needed. The rule for item 3 credited predictors drawn from institutional records, while the same records used as outcomes could rarely meet item 7, so the appraisal favours grade-based predictors, and the conclusion on prior attainment rests on reports rated Low. Effect sizes were reported as printed rather than pooled, so the review describes the direction and consistency of associations rather than estimating their size.

**Conclusion**

Across 45 reports from 39 samples, the learner characteristic most consistently associated with architecture students' grades was earlier attainment, although nine of the ten reports behind this finding were rated Low; its association weakened from aggregate grades to studio and final-project marks. Spatial ability had a modest association with early studio grades. Programme-specific admission tests, learning styles, intelligence, personality types and emotional intelligence gave weak, null, cohort-dependent or statistically unsupported results, and the largest coefficients for dispositions arose where predictor and outcome were not independent. The evidence is thin. Most of it came from single cohorts in single schools, the median sample was 90 architecture students, and only two studies adjusted for confounders, one without reporting the adjusted estimates.

Performance was defined mostly as whatever grade the institution recorded. Graders were usually unnamed, the composition of studio grades varied from one study to the next, and no study reported the reliability of the grades it analysed. The accreditation context of the programmes was almost never stated, so the corpus cannot say whether those grades reflect what accreditation requires.

The literature can therefore say that early marks forecast later marks. It cannot yet say which characteristics of students underlie performance in the studio, because this literature has not examined the studio grade itself. The next step is studies that measure several predictors at entry, in more than one school, against studio grades whose reliability is known.

**Acknowledgements**

This research was supported by the University of Technology Sarawak (UTS) Research Grant (UTS/RESEARCH/2/2023/11).

**References**

Adewale, P. O., & Adhuze, O. B. (2014). Entry qualifications and academic performance of architecture students in Nigerian polytechnics: Are the admission requirements still relevant? *Frontiers of Architectural Research, 3*(1), 69–75. https://doi.org/10.1016/j.foar.2013.11.002

Aluko, R. O., Adenuga, O. A., Kukoyi, P. O., Soyingbe, A. A., & Oyedeji, J. O. (2016). Predicting the academic success of architecture students by pre-enrolment requirement: Using machine-learning techniques. *Construction Economics and Building, 16*(4), 86–98. https://doi.org/10.5130/AJCEB.v16i4.5184

Aluko, R. O., Daniel, E. I., Oshodi, O. S., Aigbavboa, C. O., & Abisuga, A. O. (2018). Towards reliable prediction of academic performance of architecture students using data mining techniques. *Journal of Engineering, Design and Technology, 16*(3), 385–397. https://doi.org/10.1108/JEDT-08-2017-0081

Anthony, K. H. (1987). Private reactions to public criticism: Students, faculty, and practicing architects state their views on design juries in architectural education. *Journal of Architectural Education, 40*(3), 2–11. https://doi.org/10.1080/10464883.1987.10758454

Avsec, S., & Jagiełło-Kowalczyk, M. (2021). Developing a framework to predict factors significant for creative architectural design performance of freshmen and senior architecture students, by adopting and validating the CEDA. *International Journal of Engineering Education, 37*(3), 594–607.

Ayalp, G. G., & Erman, O. (2019). Learning styles of architecture students and performance in construction management courses: A case study. *International Journal of Engineering Education, 35*(4), 1250–1262.

Aydoğan, G., & Özdemir, M. (2026). Evaluations of architectural students performance in design studio related with their personal type: A view from Enneagram. *International Journal of Art & Design Education, 45*(3), 754–768. https://doi.org/10.1111/jade.70025

Bachman, L., & Bachman, C. (2006). Student perceptions of academic workload in architectural education. *Journal of Architectural and Planning Research, 23*(4), 271–304.

Barker, T. H., Hasanoff, S., Aromataris, E., Stone, J. C., Leonardi-Bee, J., Sears, K., Klugar, M., Tufanaru, C., Moola, S., Liu, X.-L., & Munn, Z. (2026). The revised JBI critical appraisal tool for the assessment of risk of bias for analytical cross-sectional studies. *JBI Evidence Synthesis, 24*(3), 401–408. https://doi.org/10.11124/JBIES-24-00523

Beatty, A. S., Walmsley, P. T., Sackett, P. R., Kuncel, N. R., & Koch, A. J. (2015). The reliability of college grades. *Educational Measurement: Issues and Practice, 34*(4), 31–40. https://doi.org/10.1111/emip.12096

Bircek, N., & Ayalp, G. G. (2022). Impact of demographic characteristics on architecture students' creativity. *AIP Conference Proceedings, 2574*, Article 100006. https://doi.org/10.1063/5.0105276

Birer, E. (2012). An evaluation of the effect of emotional intelligence on achievement orientation in architectural design education. *Gazi University Journal of Science, 25*(2), 541–553.

Bryan, A. I. (1942). Grades, intelligence and personality of art school freshmen. *Journal of Educational Psychology, 33*(1), 50–64. https://doi.org/10.1037/h0060928

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ, 368*, Article l6890. https://doi.org/10.1136/bmj.l6890

Cho, J. Y. (2012). Spatial ability, creativity, and studio performance in architectural design. In T. Fischer, K. De Biswas, J. J. Ham, R. Naka, & W. X. Huang (Eds.), *Beyond codes and pixels: Proceedings of the 17th International Conference on Computer-Aided Architectural Design Research in Asia* (pp. 131–140). CAADRIA.

Crowther, P., & Briant, S. (2021). Predicting academic success: A longitudinal study of university design students. *International Journal of Art & Design Education, 40*(1), 20–34. https://doi.org/10.1111/jade.12329

Crowther, P., & Briant, S. (2022). Gender-based differences in academic achievement in a university design program. *International Journal of Art & Design Education, 41*(4), 631–643. https://doi.org/10.1111/jade.12429

Elgazzar, E. M., Helmy, S. A., & Ibrahim, R. M. (2019). Assessing the capability of spatial ability in predicting success in the beginning design studio. *Journal of Engineering and Applied Science, 66*(1), 1–23.

García-Escudero, D., Bardí-Milà, B., Fayos, F., & Valls Dalmau, F. (2021). Relationship between the university admission score and academic performance in architecture studies. The case of Plan 94 at the Barcelona School of Architecture. *ACE: Architecture, City and Environment, 16*(47), Article 10375. https://doi.org/10.5821/ace.16.47.10375

Gawlak, A., Pruszewicz-Sipińska, E., & Bonenberg, W. (2021). Drawing skills of candidates for architectural studies vs. learning outcomes of graduates. Comparative research based on the example of the Faculty of Architecture, Poznan University of Technology. *Education Sciences, 11*(7), Article 339. https://doi.org/10.3390/educsci11070339

Hassanpour, B., Utaberta, N., & Zaharim, A. (2012). Necessity of tutor training in architecture education using Rasch measurement model. *American Journal of Applied Sciences, 9*(11), 1884–1890. https://doi.org/10.3844/ajassp.2012.1884.1890

Hemdan, J. T., Taha, D. S., & Cherif, I. A. (2023). Relationship between personality types and creativity: A study on novice architecture students. *Alexandria Engineering Journal, 65*, 847–857. https://doi.org/10.1016/j.aej.2022.09.041

Hubinský, T., Legény, J., & Špaček, R. (2022). STEM and HASS disciplines in architectural education: Readiness of FAD-STU bachelor students for practice. *Education Sciences, 12*(5), Article 294. https://doi.org/10.3390/educsci12050294

Iqbal, M., & Roberts, A. (2019). Teachers' perception of students' performance in the architectural design studio in the light of Bourdieu. *British Journal of Sociology of Education, 40*(8), 1154–1169. https://doi.org/10.1080/01425692.2019.1647093

Karlins, M., Schuerhoff, C., & Kaplan, M. (1969). Some factors related to architectural creativity in graduating architecture students. *The Journal of General Psychology, 81*(2), 203–215. https://doi.org/10.1080/00221309.1969.9711286

Kilicaslan, H. (2020). The investigation of the level of fulfilment: The basic needs of architecture students. *Eurasian Journal of Educational Research, 87*, 29–46. https://doi.org/10.14689/ejer.2020.87.2

Krishna, K. S. A., & Yalavigi, S. (1993). Architectural aptitude tests and course performance: A case study. *Architectural Science Review, 36*(3), 131–138. https://doi.org/10.1080/00038628.1993.9696748

Kvan, T., & Jia, Y. (2005). Students' learning styles and their correlation with performance in architectural design studio. *Design Studies, 26*(1), 19–34. https://doi.org/10.1016/j.destud.2004.06.004

Lionar, M. L., Pranowo, L., Zagi, N. Z., Kardina, K. I. D., Kurniawan, H., & Putri, D. D. (2025). Evaluating relationships between hand-drawing ability and design performance: Case of Architecture Study Program UGM. *ARTEKS: Jurnal Teknik Arsitektur, 10*(1), 1–10. https://doi.org/10.30822/arteks.v10i1.2841

Lockwood, C., Munn, Z., & Porritt, K. (2015). Qualitative research synthesis: Methodological guidance for systematic reviewers utilizing meta-aggregation. *International Journal of Evidence-Based Healthcare, 13*(3), 179–187. https://doi.org/10.1097/XEB.0000000000000062

Mahmoud, N. E., Kamel, S. M., & Hamza, T. S. (2020). The relationship between tolerance of ambiguity and creativity in architectural design studio. *Creativity Studies, 13*(1), 179–198. https://doi.org/10.3846/cs.2020.9628

Mahmoud, N. E., Kamel, S. M., & Hamza, T. S. (2023). The correlation between architecture students' ambiguity tolerance and their creativity: Negative capability inside the design studio. *Creativity Studies, 16*(2), 479–495. https://doi.org/10.3846/cs.2023.17461

Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., Currie, M., Qureshi, R., Mattis, P., Lisy, K., & Mu, P.-F. (2020). Chapter 7: Systematic reviews of etiology and risk. In E. Aromataris & Z. Munn (Eds.), *JBI manual for evidence synthesis*. JBI. https://doi.org/10.46658/JBIMES-20-08

Nazidizaji, S., Tomé, A., & Regateiro, F. (2014). Search for design intelligence: A field study on the role of emotional intelligence in architectural design studios. *Frontiers of Architectural Research, 3*(4), 413–423. https://doi.org/10.1016/j.foar.2014.08.005

Nazidizaji, S., Tomé, A., & Regateiro, F. (2015). Does the smartest designer design better? Effect of intelligence quotient on students' design skills in architectural design studio. *Frontiers of Architectural Research, 4*(4), 318–329. https://doi.org/10.1016/j.foar.2015.08.002

Oluwatayo, A. A., Aderonmu, P. A., & Aduwo, E. B. (2015). Architecture students' perceptions of their learning environment and their academic performance. *Learning Environments Research, 18*(1), 129–142. https://doi.org/10.1007/s10984-015-9172-7

Özyaba, M., Polat, S., Erbil, Y., & Yurtkuran, S. (2013). Mimarlık lisans öğrencilerinin sosyo-ekonomik durumlarının eğitim sürecindeki başarılarına etkisi [Socio-economic status and undergraduate success: A case study in architectural education]. *Uludağ Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 18*(2), 1–6. https://doi.org/10.17482/uujfe.30702

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9*(3), 105–119. https://doi.org/10.1111/j.1539-6053.2009.01038.x

Peterson, J. M., & Lansky, L. M. (1977). Left-handedness among architects: Partial replication and some new data. *Perceptual and Motor Skills, 45*(3, Suppl.), 1216–1218. https://doi.org/10.2466/pms.1977.45.3f.1216

Peterson, J. M., & Lansky, L. M. (1980). Success in architecture: Handedness and/or visual thinking. *Perceptual and Motor Skills, 50*(3, Suppl.), 1139–1143. https://doi.org/10.2466/pms.1980.50.3c.1139

Porat, R., & Ceobanu, C. (2024). The role of spatial ability in academic success: The impact of the Integrated Hybrid Training program in architecture and engineering higher education. *Education Sciences, 14*(11), Article 1237. https://doi.org/10.3390/educsci14111237

Rabboh, E. H. (2021). Examining the impact of students' ratings of architectural design studio on graduation project before and midst COVID-19: An exploratory study on Egyptian sophomore architecture students. *International Design Journal, 11*(4), 267–281. https://doi.org/10.21608/idj.2021.180936

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

Roberts, A. (2006). Cognitive styles and student progression in architectural design education. *Design Studies, 27*(2), 167–181. https://doi.org/10.1016/j.destud.2005.07.001

Roberts, A. S. (2007). Predictors of future performance in architectural design education. *Educational Psychology, 27*(4), 447–463. https://doi.org/10.1080/01443410601104361

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*(1), 112–118. https://doi.org/10.1037/0021-9010.85.1.112

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.

Sen, G., Adeboye, A., & Alagbe, O. (2021). The influence of architecture students' learning approaches on their academic performance in two Nigeria universities. *International Journal of Learning, Teaching and Educational Research, 20*(2), 137–151. https://doi.org/10.26803/ijlter.20.2.8

Şenkal Sezer, F., Erbil, Y., & Murat, D. (2016). Mimarlık eğitiminde duygusal zekâ–akademik başarı ilişkisi üzerine bir araştırma [A study in the correlation of emotional intelligence and academic achievement in architecture education]. *TMD International Refereed Journal of Design and Architecture*, (7), 165–183. https://doi.org/10.17365/TMD.2016716515

Şenkal Sezer, F., Murat, D., & Erbil, Y. (2016). Impact of the demographics on academic achievement in architecture education. *The Turkish Online Journal of Educational Technology*, Special Issue for INTE 2016 (November), 1178–1182. https://www.tojet.net/special/2016_11_1.pdf

Sirvent-Pérez, C. D., Pérez-Carramiñana, C., Saura-Gómez, P., González-Avilés, Á. B., & Ruiz-Cáceres, J. Á. (2025). Evaluating active learning: The role of non-presential workload monitoring in academic achievement and student satisfaction in architecture programs within the European Higher Education Area. *Education Sciences, 15*(1), Article 41. https://doi.org/10.3390/educsci15010041

Taneja, R., Taneja, P., & Goel, M. (2023). Impact of learning styles on student's creativity: Insights from India. *ISVS e-journal, 10*(10), 143–153. https://doi.org/10.61275/ISVSej-2023-10-10-10

Tastan, H., Heyik, M. A., & Er, İ. E. (2026). From conceptual design to 3D modelling: Integrating spatial ability, creativity, and technological tools in the architecture curriculum.

*International Journal of Technology and Design Education*. Advance online publication. https://doi.org/10.1007/s10798-026-10130-6

Tucker, R. (2009). Getting old and heading south: The academic success of Southerner learners in design cohorts. *Higher Education Research & Development, 28*(2), 195–207. https://doi.org/10.1080/07294360902725066

van den Heuvel, L., & Bohm, N. L. (2023). Exploring the reliability, time efficiency and fairness of comparative judgement in the admission of architecture students. In G. Reilly, M. Murphy, B. V. Nagy, & H.-M. Järvinen (Eds.), *SEFI 2023: 51st Annual Conference of the European Society for Engineering Education: Engineering education for sustainability, proceedings*. European Society for Engineering Education (SEFI). https://doi.org/10.21427/WEY8-JZ69

Watters, P., & Spas, J. (2020). Finding creativity in the extrinsically motivated environment of architecture students. *Innovations in Education and Teaching International, 57*(2), 231–241. https://doi.org/10.1080/14703297.2019.1572525

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing, 52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Yazici, G., & Dogan, F. (2024). An inquiry into the relationships among design learning, metacognitive awareness, and academic goal orientation. In C. M. Gray, E. Ciliotta Chehade, P. Hekkert, L. Forlano, P. Ciuccarelli, & P. Lloyd (Eds.), *DRS2024: Boston*. Design Research Society. https://doi.org/10.21606/drs.2024.319

Yazici, G., & Dogan, F. (2026). Reflection on designing: Metacognitive interventions to enhance metacognitive awareness, motivation, and performance in design learning. *International Journal of Technology and Design Education, 36*(3), 1347–1373. https://doi.org/10.1007/s10798-025-10044-9