Information Sciences and Technological Innovations

www.isti.reapress.com

Inf. Sci. Technol. Innov. Vol. 2, No. 4 (2025) 261–274.

Paper Type: Original Article

# From Predictive Validation to Bounded Educational Decision Support: An Uncertainty-Aware Analytics Framework for Perceived Employability among Vocational Students in Aceh, Indonesia

**Marzieh Jamili[1,*], Yeganeh Sadat Mansourian[2], Seyedeh Esmat Rasouli[1]**

[1] Department of Educational Sciences, Sar.C., Islamic Azad University, Sari, Iran; mahtabany12@gmail.com; erasouli@mail.iau.ac.ir.
[2] Department of Mathematics and Physics, University of Campania "Luigi Vanvitelli", Caserta, Italy; yeganehsadat.mansourian@studenti.unicampania.it.

**Citation:**

| | |
|---|---|
| *Received: 11 June 2025*<br>*Revised: 07 August 2025*<br>*Accepted: 18 October 2025* | Jamili, M., Mansourian, Y. S,. & Rasouli, S. E. (2025). From predictive validation to bounded educational decision support: An uncertainty-aware analytics framework for perceived employability among vocational students in Aceh, Indonesia. *Information sciences and technological innovations*, 2(4), 261-274. |

## Abstract

Educational decision support requires more than an accurate model: it must separate predictive usefulness from causal interpretation, quantify uncertainty, and connect model evidence to defensible actions. This study evaluates repeated out-of-fold prediction of self-reported perceived employability and develops a bounded, uncertainty-aware analytics framework for vocational education. The analysis used the version-4 public workbook containing 255 Grade XII students from 17 schools in seven districts of Aceh, Indonesia. Mean perceived employability was predicted from mean internship experience, 21st-century competence, and psychological capital. Data integrity, response quality, reliability, and construct separation were audited. A mean-only baseline, ordinary linear regression, Elastic Net, and Random Forest (RF) were compared using five-fold cross-validation repeated 20 times; tunable models used nested four-fold inner validation. Linear regression achieved Mean Absolute Error (MAE) = 0.254, RMSE = 0.350, and $R^2$ = 0.643, reducing MAE by 38.8% relative to the baseline. Elastic Net was effectively equivalent, whereas RF had 9.6% higher MAE. Repeated out-of-fold drop-column analysis showed that competence supplied the clearest unique information: removing it increased MAE by 0.0355 (13.8%; paired-bootstrap 95% CI [0.0152, 0.0552]). Psychological capital provided smaller supportive evidence, while internship experience had limited unique contribution. Bootstrap item-ranking stability identified project execution and leadership as the strongest curriculum-review priorities. The contribution is an auditable data-to-decision protocol that integrates nested predictive validation, conditional construct evidence, measurement diagnostics, ranking uncertainty, calibration audit, and explicit decision boundaries. The framework supports curriculum and programme review, not causal prescriptions, automated placement, or high-stakes individual decisions.

**Keywords:** Vocational education, Perceived employability, Educational analytics, Repeated cross-validation, Uncertainty-aware decision support, 21st-century competence.

## 1 | Introduction

Vocational education is expected to connect school learning with occupational practice while preparing students for changing labour-market demands. Employability is not a single technical attribute. Contemporary frameworks describe it as a multidimensional and contextual capacity involving occupational expertise,

Corresponding Author: mahtabany12@gmail.com
https://doi.org/10.48314/isti.vi.52

adaptive competence, personal resources, career identity, and access to opportunities [1–3]. Educational institutions therefore face two distinct information problems: estimating whether students report readiness for work and deciding which domains merit curriculum, student-development, or internship-programme review.

Internships are a central mechanism in vocational education, but participation alone does not guarantee high-quality learning. Reviews emphasise authentic tasks, workplace support, alignment with industry demand, and opportunities to convert experience into transferable competence [4], [5]. Alongside work-based learning, 21st-century competence and psychological capital are relevant because students must coordinate projects, communicate across media, diagnose problems, sustain confidence, and recover from setbacks [6–8].

The public Aceh dataset contains internship experience, 21st-century competence, psychological capital, and self-reported employability for 255 vocational students [9], [10]. A previous article using the same respondents estimated a moderated-mediation model and reported that competence helped connect internship experience with employability, while psychological capital strengthened the internship–competence relationship [11]. That explanatory contribution does not answer a different information-science question: which constructs retain stable, unique out-of-fold predictive information, and how can that evidence be translated into auditable but bounded educational decision support?

This distinction is substantive rather than semantic. Correlated constructs can appear influential in descriptive or structural models while adding little unique predictive information. Flexible algorithms can appear innovative without improving generalisation. Feature importance can also be misleading when model selection, measurement overlap, uncertainty, subgroup errors, and the intended decision context are not examined together [12–14]. The problem is therefore a data-to-decision problem: transforming noisy and correlated survey information into evidence that is predictive, uncertainty-aware, interpretable, and appropriately governed.

The study addresses four research questions:

I. RQ1: which candidate model provides the best repeated out-of-fold prediction of perceived employability?

II. RQ2: which educational constructs contribute unique predictive information after conditioning on the others?

III. RQ3: are model selection and construct ordering robust to demographic adjustment and response-quality sensitivity?

IV. RQ4: which item-level development needs are sufficiently stable to inform cautious curriculum and programme review?

## 1.1 | Contributions and Scope

The study makes four bounded contributions.

I. Predictive validation rather than a second explanatory model. The existing survey is reframed as an unseen-case prediction problem. Transparent, regularised, and nonlinear regressors are compared under repeated outer validation, with nested tuning where required.

II. Conditional information rather than marginal association. Standardised coefficient stability is combined with repeated out-of-fold drop-column analysis to estimate whether each construct adds predictive information after conditioning on the others.

III. Uncertainty-aware prioritisation. Measurement diagnostics, paired bootstrap intervals, bootstrap support probabilities, and item-ranking stability are integrated rather than reporting a single feature-importance ranking.

IV. Governed evidence-to-action translation. Predictive evidence is separated into primary, supportive, and contextual tiers, with a proximal assessment and an explicit evidence boundary attached to each candidate response.

The contribution is not a new causal mechanism, a new machine-learning algorithm, or external validation. It is an auditable analytical protocol for educational decision support under limited, cross-sectional survey data. *Fig. 1* summarises the conceptual and analytical layers.

# 2 | State of the Art and Research Gap

## 2.1 | Employability as a Contextual Educational Outcome

Competence-based employability frameworks emphasise occupational expertise, anticipation and optimisation, flexibility, collaboration, and balance [1]. Capital-based perspectives broaden this view by incorporating human, social, identity, and psychological resources together with labour-market context [2], [3]. Self-reported perceived employability should therefore not be equated with verified employment, wages, job retention, or supervisor-assessed performance. It is better interpreted as a readiness and resource profile that may precede labour-market outcomes.

## 2.2 | Vocational Learning, Internships, and Competence Development

Recent vocational-education literature favours demand-driven and experiential designs connecting learning with local industries and authentic work requirements [15]. Pianda et al. [4] found broad support for internship–employability relationships but also substantial heterogeneity in programme design and measurement. Longitudinal evidence indicates that employability resources can change during internships [15], suggesting that internship quality may operate through the competences and resources developed rather than placement participation alone.

Research on 21st-century skills in secondary education is extensive but uneven. A scoping review found that implementation and stakeholder perceptions receive more attention than rigorous assessment [6]. Authentic-assessment reviews recommend realistic tasks, transparent performance standards, and feedback linked to work-like outputs [16], [17]. Project-based learning is consequently a plausible design option for project execution, collaboration, and problem solving, but individual studies do not establish that one pedagogy improves employability in every vocational context [18].

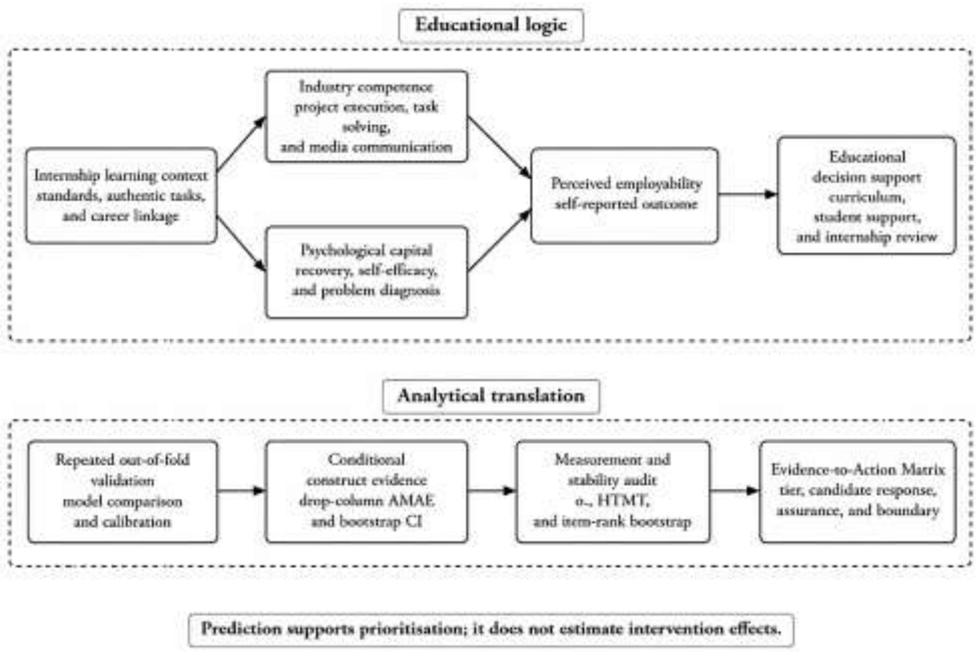



**Fig. 1. Framework for translating predictive evidence into bounded educational decision support. Arrows in the upper layer are conceptual relationships, not tested causal pathways.**

## 2.3 | Psychological Capital and School-to-Work Readiness

Psychological capital integrates hope, efficacy, resilience, and optimism as potentially developable resources [7]. Student studies associate it with employability skills, career engagement, perceived employability, and vocational readiness [8], [19], [20]. Psychological capital nevertheless overlaps conceptually with competence and may add less unique information after both constructs are modelled together.

## 2.4 | Predictive and Explainable Educational Analytics

Explainable educational analytics requires more than a feature-importance plot. Explanations should align with the user, decision, and consequences of error [12]; mispredictions can be especially consequential at the low and high ends of an outcome distribution [13], [14]. Tuning and evaluation must be separated to reduce selection bias [21], [22]. Repeated cross-validation can describe resampling variability, but its fold estimates are dependent and do not provide a universally unbiased variance estimator [23]. These constraints motivate paired out-of-fold comparisons, explicit uncertainty statements, and conservative decision boundaries.

## 2.5 | Research Gap

The Aceh data have already supported an explanatory moderated-mediation analysis [11]. Four gaps remain: rigorous out-of-fold model comparison, unique conditional construct contribution, integration of measurement and ranking uncertainty, and translation into bounded educational decision support. The present analysis addresses those gaps while explicitly withholding causal and external-validity claims.

# 3 | Methodology

## 3.1 | Study Design, Data Source, and Provenance

This study is a secondary cross-sectional analysis of the open workbook and bilingual questionnaire archived in Mendeley Data, version 4 [10]. The version-4 workbook was treated as the computational source of truth: all row counts, demographic summaries, construct scores, and model inputs were recomputed directly from that file rather than copied from published summary tables. It contains 255 Grade XII vocational students from 17 schools across seven districts in Aceh province, Indonesia. The source article reports proportional stratified sampling, data collection from September to December 2022, written informed consent, anonymisation, and approval by the Universitas Padjadjaran Ethics Committee (Protocol No. 7y/UN5.KEP/EC/2025) [9]. No new participants were recruited. School identifiers were absent, preventing school-grouped validation and multilevel modelling.

Workbook structure, numeric conversion, missingness, duplicate records, coding ranges, and item-response ranges were audited. No missing item values, duplicate respondents, invalid demographic codes, or responses outside the 1–5 scale were detected. Three respondents selected the same option for every substantive item; they were retained in the primary analysis and removed only in sensitivity analysis. The Supplementary Information documents the provenance audit and discrepancies between the archived workbook and selected summary counts in the source article.

## 3.2 | Variable Construction and Analytical Roles

Let $x_{ikq}$ denote respondent $i$'s response to item $q$ of construct $k$, and let $m_k$ be the number of items in that construct. Construct scores were arithmetic item means,

$$s_{ik} = \frac{1}{m_k}\sum_{q=1}^{m_k} x_{ikq}. \tag{1}$$

So all scores remained on the original 1–5 response scale. The outcome $y_i$ was the mean of 12 perceived-employability items. The three primary predictors were mean internship experience (14 items), 21st-century competence (10 items), and psychological capital (14 items). Outcome items were excluded from the feature matrix, preventing direct target leakage. Gender, age, and accreditation were included only in sensitivity models. School status was not included because one category contained seven respondents and the English source label for the other category ("country") is ambiguous.

**Table 1. Variable construction and analytical role.**

| Role | Construct or Field | Items | Operational Definition | Analytical Use |
|---|---|---|---|---|
| Outcome | Perceived employability | 12 | Arithmetic mean on the 1–5 scale | Continuous prediction target |
| Primary predictor | Internship experience | 14 | Arithmetic mean on the 1–5 scale | Work-based learning context |
| Primary predictor | 21st-century competence | 10 | Arithmetic mean on the 1–5 scale | Competence information |
| Primary predictor | Psychological capital | 14 | Arithmetic mean on the 1–5 scale | Personal-resource information |
| Sensitivity covariates | Gender, age, Accreditation | – | Source-coded demographic fields | Robustness check only |
| Response-quality Flag | Complete straightlining | 50 | Identical response across all Substantive items | Exclusion sensitivity only |

Notes: Construct scoring was prespecified in the archived design manifest. No employability item entered the predictor matrix. School status was excluded from adjusted models because of sparse and ambiguous coding.

Before model fitting, pairwise correlations, variance-inflation factors, and the standardised design-matrix condition number were inspected. The maximum variance-inflation factor was 2.73, and the condition number was 3.10. These values do not remove the substantive concern about content overlap, which was assessed separately through Heterotrait–Monotrait Ratio (HTMT).

## 3.3 | Measurement and Response-Quality Audit

Internal consistency was assessed using Cronbach's alpha, corrected item–total correlations, and McDonald's omega. Omega was estimated from a one-factor maximum-likelihood model fitted to sample-standardised items. Construct separation was examined using the HTMT computed from absolute Pearson item correlations [24]. Respondent-level percentile bootstrap intervals used 1,000 resamples for omega and 5,000 for HTMT. Because the items are five-point ordinal self-reports, these analyses are supplementary diagnostics and do not replace ordinal confirmatory factor analysis. Exact formulae, loadings, intervals, and convergence checks are provided in the Supplementary Information.

## 3.4 | Candidate Prediction Models

All preprocessing was estimated within training data. For predictor $j$ in a training set T, standardisation was:

$$\mathrm{z_{ij}} = \frac{\mathrm{x_{ij}} - \hat{\mathrm{x}}_{\mathrm{j},\mathcal{T}}}{\mathrm{x}_{\mathrm{j},\mathcal{T}}}, \tag{2}$$

where $x^{`}_{j,\mathrm{T}}$ and $s_{j,\mathrm{T}}$ are the training-fold mean and standard deviation. The same transformation was then applied to the corresponding test fold.

Four regressors were compared. The mean-only baseline predicted the training-fold outcome mean for every test observation. Ordinary least squares estimated:

$$(\hat{\beta}_{\circ}, \hat{\beta}) = \arg\min_{\beta_0,\beta} \sum_{i\in\mathcal{T}} \left(y_i - \beta_0 - z_i^{\top}\beta\right)^2. \tag{3}$$

Elastic Net combined $L_i$ and $L_z$ penalties [25]:

$$(\widehat{\beta_0}, \bar{\beta}) = \arg\min_{\beta,\beta} \left\{\frac{1}{2n_{\mathcal{T}}} \sum_{i\in\mathcal{T}} \left(y_i - \beta_0 - z_i^{\top}\beta\right)^2 + \alpha\left[\rho\,\Big\|\beta\|_i + \frac{1-\rho}{2}\Big\|\,\beta\|_2^2\right]\right\}, \tag{4}$$

where $\alpha \geq 0$ controls total regularisation and $\rho \in [0, \mathrm{i}]$ controls the $L_i$–$L_z$ mixture. The grid contained nine $\alpha$ values from 0.0001 to 1.0 and six $\rho$ values from 0.05 to 1.0.

Random Forest (RF) represented a nonlinear benchmark [25]. With $B = 3oo$ regression trees, its prediction was:

$$\widehat{f_{RP}}(x) = \frac{1}{B}\sum_{B=1}^{B} T_B(x), \tag{5}$$

where $T_b$ is a bootstrap-grown regression tree with random feature selection. Inner tuning considered tree depth {3, None}, minimum leaf size {2, 3, 10}, and feature fractions {0.55, 1.0}.

## 3.5 | Repeated Nested Validation and Performance Measures

The outer evaluation used shuffled five-fold cross-validation repeated $R$ = 20 times with seed 42. The procedure produced 100 train–test evaluations and exactly one out-of-fold prediction per respondent in each repeat. Elastic Net and RF were tuned separately within each outer training set using shuffled four-fold inner cross-validation and Mean Absolute Error (MAE); the outer test fold was never used for tuning. This separation was used to limit selection bias [21], [22].

For model $m$, the repeated out-of-fold loss estimate was:

$$\hat{L}_m = \frac{1}{R}\sum_{r=1}^{R} \frac{1}{n}\sum_{i=1}^{n} \ell\left(y_i, \hat{y}_{ir}^{(m)}\right), \tag{6}$$

where $y^{(m)}$ is respondent $i$'s out-of-fold prediction in repeat $r$. The primary loss was absolute error. Secondary metrics were:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \tag{9}$$

Values reported in the model table are means across the 20 outer repeats. Repeat-level standard deviations are descriptive summaries of resampling variation, not confidence intervals, because repeated cross-validation estimates are dependent [23].

The prespecified selection rule prioritised the lowest mean outer MAE and then parsimony when predictive performance was effectively equivalent. Repeat-wise paired comparisons were used to describe consistency; no independent-fold $t$ test was interpreted as valid.

## 3.6 | Conditional Construct Evidence and Uncertainty

The selected linear model was examined through standardised coefficient stability across the 100 outer fits. Coefficients were not interpreted causally; their purpose was to show sign and ordering stability after training-fold standardisation.

For construct $j$, unique predictive contribution was defined through repeated out-of-fold drop-column analysis:

$$\Delta\mathrm{MAE}_j = \mathrm{MAE}_{-j} - \mathrm{MAE}_{full}\,. \tag{10}$$

Where the reduced and full models used identical outer splits. Positive values indicate deterioration after removing construct $j$. Respondent-level paired bootstrap resampling with 5,000 resamples produced 95% percentile intervals for the difference in repeated out-of-fold absolute error. Bootstrap support was summarised as

$$\widehat{\mathrm{P}}_{\mathrm{j}} - \frac{1}{\mathrm{B}_{\mathrm{b}\times 10}} \sum_{\mathrm{b}=1}^{\mathrm{n_b}} \mathrm{I}\left(\mathrm{AMAE}_{\mathrm{j}}^{(\mathrm{b})} > 0\right). \tag{11}$$

with $B_{\text{boot}}$ = 3,000. The resulting value is a bootstrap support probability, not a frequentist $p$ value.

### 3.7 | Item-Ranking Stability and Bounded Action Translation

Within each predictor construct, item means and ascending ranks were recomputed in 5,000 respondent-level bootstrap samples. Rank 1 denoted the lowest item mean and hence the largest descriptive room for development. For item $q$,

$$\bar{\mathrm{P}}_{\alpha}(\mathrm{Top}_{-3}) = \frac{1}{\mathrm{B}_{\mathrm{b,DON}}} \sum_{\mathrm{b}=1}^{\mathrm{B}_{\mathrm{b}-1}} \mathrm{I}(\mathrm{r}_{\alpha \mathrm{b}} \leq 3). \tag{12}$$

where $r_{qb}$ is the within-construct rank in bootstrap sample $b$. Items entered the Evidence-to-Action Matrix when $P_q$ (Top-3) ≥ 0.60. This threshold is an explicit operational rule, not a validated universal standard. Item-ranking probabilities are descriptive stability measures and are not item-level predictive importance.

The translation rule combined two evidence levels. Construct-level drop-column evidence determined the tier: primary, supportive, or contextual. Item-level stability then identified candidate content within that tier. Each candidate response was paired with a proximal assessment and a boundary statement. No response was treated as an intervention with an established effect.

### 3.8 | Sensitivity, Calibration, Subgroup Audit, and Software

Sensitivity models added gender, age, and accreditation and separately excluded the three complete straightliners. Each respondent's mean of 20 repeated out-of-fold predictions, $y_i$, was used for descriptive calibration:

$$\mathrm{y_i} = \mathrm{a} + \mathrm{b}\hat{\bar{\mathrm{y}}}_{\mathrm{i}} + \varepsilon_{\mathrm{i}}, \tag{13}$$

where ideal descriptive calibration corresponds to $a$ = 0 and $b$ = i. Prediction ranges, outcome-quartile errors, and subgroup MAE were audited. Subgroup results were not treated as formal fairness tests because groups were not prespecified for inferential comparison and several relevant contextual variables were unavailable. Analysis was conducted in Python 3.12.13 using pandas and scikit-learn 1.5.1 [26]. The random seed, split manifests, model grids, derived tables, and figures were retained in the analysis archive. Yeganeh Sadat Mansourian conducted the computational analyses, implemented the analytical workflow, and verified the numerical outputs against the archived tables and manifests. Marzieh Jamili and Seyedeh Esmat Rasouli reviewed the theoretical framework, educational interpretation, and pedagogical implications. All authors critically revised the manuscript and approved the final analytical interpretation.

## 4 | Results

### 4.1 | Data Integrity and Measurement Quality

The version-4 workbook contained 255 analysable records. No missing item responses, duplicate records, invalid demographic codes, or values outside the 1–5 response range were detected. Three complete straightliners were retained in the primary analysis. *Table 2* reports independently recomputed sample and construct summaries.

Internal consistency was high. The minimum corrected item–total correlation was 0.548, and deleting any item reduced rather than improved the corresponding alpha. The competence–employability HTMT estimate was borderline, so shared questionnaire method or content overlap may contribute to the predictive signal.

## 4.2 | Predictive Performance and Robustness

Candidate-model and sensitivity results are summarised in *Table 3*; *Fig. 2* provides the corresponding graphical comparison.

Linear regression reduced MAE by 38.8% relative to the mean-only baseline. Elastic Net differed from linear regression by 0.00007 MAE on average. RF had 9.6% higher MAE and did not outperform linear regression in any of the 20 paired repeats. Adding demographics increased MAE by approximately 1.0% and was worse in 19 of 20 repeats. Excluding the three straightliners also produced a small deterioration and did not change construct ordering.

Across the 100 outer linear-model fits, all three construct coefficients were positive. Mean standardised coefficients were 0.294 for competence, 0.131 for psychological capital, and 0.102 for internship experience. Their ranges were [0.225, 0.357], [0.081, 0.214], and [0.056, 0.139], respectively. The same ordering was retained in all sensitivity specifications.

## 4.3 | Conditional Construct Evidence and Educational Priorities

Removing competence increased out-of-fold MAE by 0.0355 (13.8%; 95% CI [0.0162, 0.0562]; bootstrap support $P^+$ = o.9998). Psychological capital produced a smaller increase of 0.0066 (95% CI [-0.0027, 0.0159]; $P^+$ = 0.9170), while internship experience increased MAE by only 0.0013 (95% CI [-0.0072, 0.0095]; $P+$ = 0.5970). Competence therefore supplied the clearest unique predictive information conditional on the other constructs.

**Table 2. Sample characteristics, construct descriptives, and measurement quality.**

| **Panel A: Sample Characteristics** | | | | | |
|---|---|---|---|---|---|
| Variable | Category | n | % | | |
| Gender | Male | 111 | 41.7 | | |
| | Female | 155 | 58.3 | | |
| Age | 15 years | 22 | 8.3 | | |
| | 17 years | 184 | 69.2 | | |
| | 18 years | 60 | 22.6 | | |
| School status | Code 1 (Source label: "country")[a] | 259 | 97.4 | | |
| | Private | 7 | 2.6 | | |
| Accreditation | A (Excellent) | 131 | 49.2 | | |
| | B (Good) | 135 | 50.8 | | |
| **Panel B: Construct Descriptives and Reliability** | | | | | |
| Construct | Items | Mean | SD | $\alpha$ | $\omega$ |
| Internship experience | 14 | 3.999 | 0.611 | 0.919 | 0.921 |
| 21st-century competence | 10 | 3.830 | 0.588 | 0.907 | 0.907 |
| Psychological capital | 14 | 4.119 | 0.573 | 0.926 | 0.927 |
| Perceived employability | 12 | 3.970 | 0.588 | 0.917 | 0.918 |

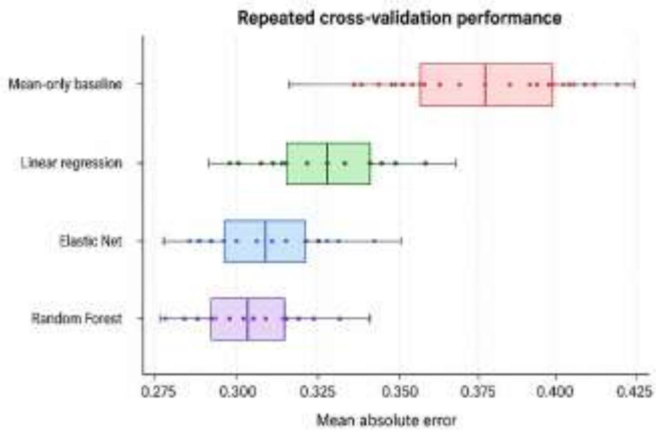
[a] The wording is preserved from the English codebook because the substantive meaning of "country" could not be independently verified. School status was not used in adjusted prediction models.

Notes: N = 266. All values were recalculated from the archived version-4 workbook. Omega intervals, the data-provenance audit, and one-factor diagnostics are reported. Competence–employability HTMT = 0.854, bootstrap 95% CI [0.774, 0.915].

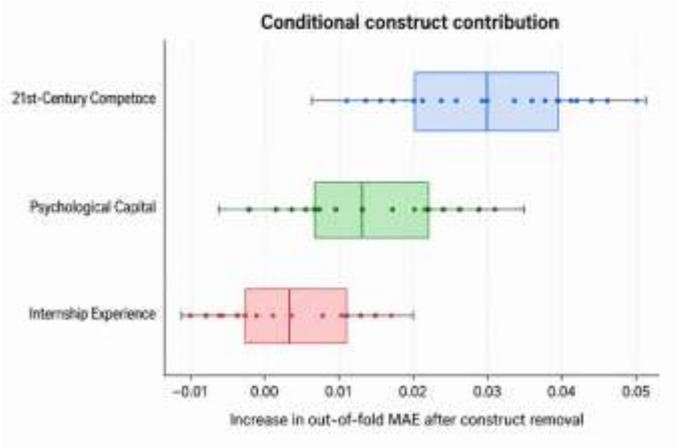
**Table 3. Repeated cross-validation performance and sensitivity analyses.**

| **Panel A: Candidate-Model Performance Model** | **MAE** | **RMSE** | $R^2$ | **MAE SD** |
|---|---|---|---|---|
| Mean-only baseline | 0.432 | 0.590 | -0.009 | 0.002 |
| Linear regression | 0.264 | 0.350 | 0.543 | 0.002 |
| Elastic Net | 0.254 | 0.350 | 0.544 | 0.002 |
| RF | 0.290 | 0.381 | 0.578 | 0.004 |
| **Panel B: Linear-Model Sensitivity Specifications Specification** | **MAE** | **RMSE** | $R^2$ | $n$ |
| Primary, full sample | 0.264 | 0.350 | 0.543 | 255 |
| Demographically adjusted, full sample | 0.257 | 0.354 | 0.535 | 255 |
| Primary, excluding straightliners | 0.257 | 0.352 | 0.535 | 253 |
| Adjusted, excluding straightliners | 0.270 | 0.355 | 0.527 | 253 |

Notes: Elastic Net and RF were tuned in nested inner folds. Values are means across 20 outer repeats; MAE SD is descriptive and is not a confidence interval. Percentage comparisons were calculated from unrounded estimates. Linear regression had lower MAE than Elastic Net in 15 of 20 repeats. Their aggregate performance was effectively indistinguishable; the prespecified primary metric and parsimony favoured linear regression.




**a.**




**b.**

**Fig. 2. Predictive performance and conditional construct contribution: a. mean repeated-CV MAE with descriptive SD across 20 repeats, and b. increase in out-of-fold MAE after construct removal, with paired-bootstrap y5% intervals.**

The item-level bootstrap identified stable or moderately stable lower-scoring items within each predictor construct. *Table 4* translates these findings into bounded priorities.

**Table 4. Evidence-to-action matrix for curriculum and programme review.**

| Tier and Item | Educational Target | *P* (Top-3) | Candidate Educational Response | Proximal Assessment |
|---|---|---|---|---|
| Primary (C3) | Project planning, organisation, and execution | 0.999 | Authentic team projects with milestones, role allocation, progress reviews, and delivery responsibility | Planning/execution rubric, milestone completion, and project log |
| Primary (C4) | Leadership and group inspiration | 0.990 | Rotating leadership roles, shared goal setting, peer feedback, and structured leadership debriefs | Peer/instructor leadership rubric and team-process reflection |
| Primary (C10) | Media-information analysis and communication | 0.610 | Workplace media-literacy portfolio using realistic vocational communication tasks | Portfolio rubric covering information appraisal, audience fit, clarity, and production quality |
| Supportive (PsyCap10) | Constructive recovery after failure | 1.000 | Structured failure debriefs, recovery planning, and reflection after setbacks | Recovery-plan quality, reflective response, and follow-up engagement |
| Supportive (PsyCap12) | Confidence in completing difficult tasks | 1.000 | Progressively challenging mastery tasks with scaffolding and formative feedback | Challenge completion, task-specific self-efficacy check, and feedback-response record |

**Table 4. Continued.**

| Tier and Item | Educational Target | $P$ (Top-3) | Candidate Educational Response | Proximal Assessment |
|---|---|---|---|---|
| Supportive (PsyCap8) | Problem identification and analysis | 0.846 | Case-based diagnosis before solution generation and root-cause discussion | Problem-diagnosis rubric and quality of evidence use |
| Contextual (SIE3) | Understanding workplace standards | 1.000 | Joint school–workplace standards orientation with examples of acceptable performance | Standards-alignment checklist and student–supervisor agreement |
| Contextual (SIE12) | Connecting internship work to career opportunities | 1.000 | Career-pathway mapping and a post-placement debrief linking tasks, skills, and future pathways | Career map, evidence-based reflection, and next-step skill goals |

Notes: $P$ (Top-3) is the proportion of 5,000 respondent-bootstrap samples in which an item appeared among the three lowest-scoring items within its construct. Primary items inherit bootstrap-supported construct evidence, although competence–employability separation is borderline. Supportive items come from a construct whose drop-column interval includes zero. Contextual items come from a construct with limited unique contribution. Item ranks are descriptive; none of the candidate responses were evaluated as an intervention. SIE13 was excluded because $P$(Top-3) = 0.460, close to SIE4 at 0.364.

### 4.4 | Calibration and Model Audit

Mean repeated out-of-fold predictions ranged from 1.590 to 4.930 and remained within the 1–5 outcome scale. Descriptive calibration was close to ideal (intercept 0.056; slope 0.986), but the model regressed towards the mean. The lowest observed employability quartile was overpredicted by 0.302 points, whereas the highest quartile was underpredicted by 0.278 points. Female respondents had higher descriptive MAE than male respondents (0.288 versus 0.231), but their respondent-bootstrap MAE intervals overlapped, and demographic adjustment worsened overall generalisation. These findings are model-audit signals, not evidence of a stable demographic effect.

## 5 | Discussion

### 5.1 | Principal Interpretation

The contribution is not the rediscovery of a positive competence–employability association. That association is consistent with employability theory and the previous moderated-mediation analysis [1], [11]. The added value is that competence retained a substantial unique contribution under repeated out-of-fold validation; its coefficient remained positive in every outer fit, and a more flexible tree model did not improve generalisation.

The model comparison should not be overinterpreted. RF's weaker performance does not prove that the underlying relationship is linear; it shows only that, with 255 observations and three construct-level predictors, additional nonlinear flexibility did not improve repeated out-of-fold accuracy. Similarly, the large $R^2$ should not be interpreted as verified labour-market prediction. All substantive measures were collected through one self-report questionnaire, and competence–employability separation was borderline.

From an information-science perspective, the main result is the behaviour of the full pipeline. Generalisation testing favoured a simple model; conditional information analysis separated competence from weaker constructs; measurement and item-rank uncertainty tempered interpretation; calibration exposed regression to the mean; and the action matrix preserved the distinction between predictive priorities and intervention effects. This combination is more defensible than selecting a complex algorithm and translating an unqualified importance plot directly into policy.

### 5.2 | Primary Curriculum-Review Priority: Competence

Project execution and leadership showed both meaningful descriptive development gaps and very high ranking stability. Authentic projects are a plausible response because they make planning, coordination, revision, and delivery observable. Reviews emphasise that implementation and assessment quality remain major gaps [6], [14], [17]. The proposed responses therefore include proximal rubrics and work products instead of assuming that participation automatically improves employability.

C10 was retained cautiously. Its probability of appearing among the three lowest-scoring competence items was 0.510, so media-information communication is a secondary component rather than an equally certain target.

## 5.3 | Supportive Priority: Psychological Capital

Recovery after failure, confidence with difficult tasks, and problem diagnosis were stable lower-scoring items. These themes align with efficacy and resilience [7] and previous student-employability research [8], [19], [20]. However, the construct-level drop-column interval crossed zero. Psychological-capital activities should therefore complement rather than displace competence-focused curriculum development.

## 5.4 | Contextual Priority: Internship Programme Quality

Internship experience contributed little unique information after competence and psychological capital were included. This finding does not imply that internships are unimportant. Reviews and longitudinal evidence suggest that internships can build employability resources when tasks, supervision, and reflection are meaningful [4], [5]. The evidence supports reviewing workplace standards and career linkage, not claiming an independent employability effect.

## 5.5 | Implications for Analytics-Supported Management

The framework suggests a disciplined management sequence: verify generalisation, select the simplest competitive model, estimate conditional information, quantify measurement and ranking uncertainty, attach proximal assessments, and audit prediction failures. It is appropriate for curriculum and programme review, not individual labelling, automated placement, or high-stakes student screening. Before implementation, local vocational educators, workplace supervisors, and students should review the candidate responses for feasibility and contextual relevance.

A practical side effect of this conservative approach is that decision-makers receive fewer and weaker claims than a conventional association study might offer. That restraint is a benefit for governance but a limitation for immediate action: the framework identifies where to investigate and assess, not what intervention will work.

## 5.6 | Limitations and Future Research

The design is cross-sectional and cannot establish temporal ordering, mediation, or intervention effects. All substantive variables were self-reported in one questionnaire, and the outcome was perceived employability rather than verified employment, wages, job retention, or supervisor-assessed performance.

The sample contains 266 students from 17 schools in one Indonesian province. School identifiers were absent from the public workbook, so school-grouped cross-validation and multilevel modelling were impossible. Random folds may share unobserved school context across training and test sets; the study must not be described as externally validated.

The archived workbook and selected demographic summaries in the source article are not fully concordant. This analysis treats the version-4 workbook as the reproducible source of truth and documents the discrepancy in the Supplementary Information. The school-status label is also ambiguous and was not interpreted or used in adjusted models.

The measurement diagnostics do not replace ordinal confirmatory factor analysis. Competence–employability separation was borderline, and approximate one-factor average variance extracted values were slightly below 0.50. The 0.60 item-stability threshold is operational, not a validated decision standard. Bootstrap stability makes descriptive ranks less sample-specific under respondent resampling, but it does not turn items into causal treatment targets.

Future work should preserve school identifiers, use school- or district-grouped validation, collect longitudinal and externally assessed labour-market outcomes, and evaluate the framework in an independent cohort. Intervention or quasi-experimental studies are required before candidate responses can be interpreted as effective educational actions. Stakeholder review should precede local implementation.

# 6 | Conclusions

A transparent linear model provided the best balance of accuracy, stability, and interpretability in this dataset. Competence was the only construct with a large and bootstrap-supported unique contribution. Psychological capital supplied smaller supportive evidence, while internship experience was more defensible as programme-quality context than as an independent predictive lever.

The information-science contribution is an uncertainty-aware data-to-decision protocol linking nested generalisation assessment, conditional construct information, measurement diagnostics, item-rank stability, calibration audit, and explicit action boundaries. Project execution and leadership are the strongest curriculum-review targets; psychological resources can complement that work; and internship standards and career linkage merit programme review. These are priorities for assessment and local deliberation, not causal prescriptions.

## Acknowledgements

The authors acknowledge the original researchers for making the anonymised dataset and bilingual questionnaire publicly available.

## Authors' Contributions

Conceptualisation and overall study design: M. J., Y. S. M., and S. E. R. Methodology, software, validation, formal analysis, data curation for secondary analysis, and visualisation: Y. S. M. Theoretical framing, literature synthesis, educational-science and human-sciences interpretation, pedagogical contextualisation, and development of the educational implications: M. J. and S. E. R. Writing (original draft of the methodology and results sections): Y. S. M. Writing (original draft of the introduction, theoretical background, discussion, and conclusions): M. J. and S. E. R. Writing (review and editing): All authors. All authors have read and approved the submitted version and accept responsibility for their respective contributions and for the integrity of the final manuscript.

## Funding

No funds, grants, or other support were received during the preparation of this manuscript. No funder had a role in the design, analysis, interpretation, writing, or decision to submit.

## Data Availability

The source workbook and bilingual questionnaire are publicly available through Mendeley Data, version 4: https://doi.org/10.17532/hmn3b4c5c4.4. Derived analysis outputs and design manifests accompany the submission archive. No new participant data were generated.

## Code Availability

Executable analysis code is available from the corresponding author upon reasonable request, in coordination with Y. S. M., who conducted the computational analyses. The accompanying archive includes model-design manifests, split specifications, derived tables, and figure outputs but should not be described as a complete executable reproduction repository until the code is deposited.

## Conflict of Interest

The authors have no relevant financial or non-financial interests to disclose.

## Consent for Publication

Not applicable; the analysis used anonymised public data and contains no identifiable participant information.

# Ethics Approval and Consent to Participate

This secondary analysis involved no new recruitment or contact with participants. The source article reports ethics approval from the Universitas Padjadjaran Ethics Committee (Protocol No. 7y/UN5.KEP/EC/2025), written informed consent, and anonymisation [9].

# Use of AI-Assisted Tools

A large language model assisted with code organisation, language editing, and manuscript structuring. It was not listed as an author and was not used as an independent source of evidence or interpretation. Y. S. M. executed and verified the computational analyses and checked the numerical outputs against the archived results. M. J. and S. E. R. reviewed the theoretical framework, educational-science and human-sciences interpretation, pedagogical implications, and relevant literature. All authors critically revised the AI-assisted material, verified the content relevant to their contributions, and accept responsibility for the accuracy, integrity, and originality of the final manuscript.

# References


[1] Heijde, C. M. Van Der, & Van Der Heijden, B. I. J. M. (2006). A competence-based and multidimensional operationalization and measurement of employability. *Human resource management: Published in cooperation with the school of business administration, the university of michigan and in alliance with the society of human resources management*, *45*(3), 449–476. https://doi.org/10.1002/hrm.20119

[2] Clarke, M. (2018). Rethinking graduate employability: The role of capital, individual attributes and context. *Studies in higher education*, *43*(11), 1923–1937. https://doi.org/10.1080/03075079.2017.1294152

[3] Lo Presti, A., Ingusci, E., Magrin, M. E., Manuti, A., & Scrima, F. (2019). Employability as a compass for career success: Development and initial validation of a new multidimensional measure. *International journal of training and development*, *23*(4), 253–275. https://doi.org/10.1111/ijtd.12161

[4] Pianda, D., Hilmiana, H., Widianto, S., & Sartika, D. (2024). The impact of internship experience on the employability of vocational students: A bibliometric and systematic review. *Cogent business & management*, *11*(1), 2386465. https://doi.org/10.1080/23311975.2024.2386465

[5] Gupta, S. L., Mittal, A., Singh, S., & Dash, D. N. (2024). Demand-driven approach of vocational education and training (VET) and experiential learning: A thematic analysis through systematic literature review (SLR). *Asian education and development studies*, *13*(1), 45–63. https://doi.org/10.1108/AEDS-07-2023-0083

[6] Kain, C., Koschmieder, C., Matischek-Jauk, M., & Bergner, S. (2024). Mapping the landscape: A scoping review of 21st century skills literature in secondary education. *Teaching and teacher education*, *151*, 104739. https://doi.org/10.1016/j.tate.2024.104739

[7] Luthans, F., & Youssef-Morgan, C. M. (2017). Psychological capital: An evidence-based positive approach. *Annual review of organizational psychology and organizational behavior*, *4*, 339–366. https://doi.org/10.1146/annurev-orgpsych-032516-113324

[8] Sulistiobudi, R. A., & Kadiyono, A. L. (2023). Employability of students in vocational secondary school: Role of psychological capital and student-parent career congruences. *Heliyon*, *9*(2), e13214. https://doi.org/10.1016/j.heliyon.2023.e13214

[9] Pianda, D., Hilmiana, H., Widianto, S., & Sartika, D. (2026). A stratified survey dataset on internship experience, competencies, psychological capital, and employability of vocational students in Aceh, Indonesia. *Data in brief*, *67*, 112992. https://doi.org/10.1016/j.dib.2026.112992

[10] Pianda, D., Hilmiana, H., Widianto, S., & Sartika, D. (2026). *Dataset of student internship experience, competence, psychological capital and employability's vocational student (Original data)*. https://doi.org/10.17632/hmn3b4c6c4.4

[11] Pianda, D., Hilmiana, Widianto, S., & Sartika, D. (2025). The influence employability of vocational students through internship experiences and 21st-century competencies: A moderated mediation model. *Cogent education*, *12*(1), 2476285. https://doi.org/10.1080/2331186X.2025.2476285

[12] Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J., ... & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and education: Artificial intelligence*, *3*, 100074. https://doi.org/10.1016/j.caeai.2022.100074

[13] Saqr, M., & López-Pernas, S. (2024). Why explainable AI may not be enough: Predictions and mispredictions in decision making in education. *Smart learning environments*, *11*(1), 52. https://doi.org/10.1186/s40561-024-00343-4

[14] Hlosta, M., Herodotou, C., Papathoma, T., Gillespie, A., & Bergamin, P. (2022). Predictive learning analytics in online education: A deeper understanding through explaining algorithmic errors. *Computers and education: Artificial intelligence*, *3*, 100108. https://doi.org/10.1016/j.caeai.2022.100108

[15] Lo Presti, A., Costantini, A., Akkermans, J., Sartori, R., & De Rosa, A. (2023). Employability development during internships: A three-wave study on a sample of psychology graduates in Italy. *Journal of career development*, *50*(6), 1155–1171. https://doi.org/10.1177/08948453231161291

[16] Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in educational evaluation*, *83*, 101425. https://doi.org/10.1016/j.stueduc.2024.101425

[17] Zhan, Y., Boud, D., & Du, Z. (2025). Designing for authentic assessment: A scoping review. *Higher education*, 1–18. https://doi.org/10.1007/s10734-025-01588-9

[18] Rehman, N., Huang, X., Mahmood, A., AlGerafi, M. A. M., & Javed, S. (2024). Project-based learning as a catalyst for 21st-Century skills and student engagement in the math classroom. *Heliyon*, *10*(23), e39988. https://doi.org/10.1016/j.heliyon.2024.e39988

[19] Ayala Calvo, J. C., & Manzano García, G. (2021). The influence of psychological capital on graduates' perception of employability: The mediating role of employability skills. *Higher education research & development*, *40*(2), 293-308. https://doi.org/10.1080/07294360.2020.1738350

[20] Baluku, M. M., Mugabi, E. N., Nansamba, J., Matagi, L., Onderi, P., & Otto, K. (2021). Psychological capital and career outcomes among final year university students: The mediating role of career engagement and perceived employability. *International journal of applied positive psychology*, *6*(1), 55–80. https://doi.org/10.1007/s41042-020-00040-w

[21] Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, *7*(1), 91. https://doi.org/10.1186/1471-2105-7-91

[22] Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *The journal of machine learning research*, *11*, 2079–2107. https://www.jmlr.org/papers/volume11/cawley10a/cawley10a.pdf

[23] Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *Journal of machine learning research*, *5*(Sep), 1089–1105. https://www.jmlr.org/papers/volume5/grandvalet04a/grandvalet04a.pdf

[24] Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

[25] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society series b: Statistical methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00527.x

[26] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *The journal of machine learning research*, *12*, 2825-2830. https://doi.org/10.5555/1953048.2078195