Article

Not peer-reviewed version

Digital Requirements and Advertised Salaries in the Educational Labor Market of Kazakhstan: An Nlp-Based Analysis of Job Vacancy Data

Klara Sauanova , Zhanar Bydakhmet , Gulnar Omarova , Sholpan Sagyndykova *

Posted Date: 18 May 2026

doi: 10.20944/preprints202605.1145.v1

Keywords: educational labor market; digital skill requirements; NLP; text mining; computational social science; online job postings; employer sorting; quantile regression; Kazakhstan

doi:10.20944/preprints202605.1145.v1

Article

Digital Requirements and Advertised Salaries in the Educational Labor Market of Kazakhstan: An Nlp-Based Analysis of Job Vacancy Data

Klara Sauanova [1,2], Zhanar Bydakhmet [1,2], Gulnar Omarova [2] and Sholpan Sagyndykova [1,*]

[1] Institute of Information Technologies, Almaty University of Power Engineering and Telecommunications after G.Daukeyev, Baitursynuly street 126/1, 050013 Almaty, Kazakhstan

[2] International University of Information Technologies, Manas str. 34/1., 050040 Almaty, Kazakhstan

* Correspondence: sagyndykovasolpan58@gmail.com

Abstract

This paper examines whether digital skill requirements in educational job postings are associated with higher advertised salaries in the labor markets of Almaty and Astana, Kazakhstan, and whether the observed salary gap reflects genuine skill valuation or employer heterogeneity. Using 794 vacancies collected via the HH.kz API, we construct an analytical sample of 755 observations after data cleaning. A seven-category dictionary-based skill extraction pipeline is implemented and its integral lexical consistency is evaluated against a TF-IDF+Logistic Regression baseline (5-fold CV: F1=0.755, AUC=0.928). Vacancies specifying at least one digital requirement carry a median advertised salary that is 18.2% higher than non-digital postings. OLS with HC3- robust standard errors and occupational and city controls yields a coefficient of 0.200 (SE = 0.039, $p < 0.001$, approximately +22.2%). Adding an employer-type proxy reduces this estimate to +15.2% ($p < 0.001$). In the Almaty subsample, the effect is no longer statistically significant ($p = 0.075$). Part of the observed premium reflects a measurement problem. Higher-paying employers are also more likely to specify digital requirements, inflating the estimated association. The design does not identify firm effects and supports only an associational interpretation. Quantile regression shows that the salary-digital gradient increases from +15.0% at Q25 to +44.4% at Q90. This pattern is consistent with labor market segmentation rather than a uniform skill premium.

Keywords: educational labor market; digital skill requirements; NLP; text mining; computational social science; online job postings; employer sorting; quantile regression; Kazakhstan

1. Introduction

The requirements for educational staff are changing due to digitalization. Employers are increasingly specifying digital tools and platforms in job descriptions, in addition to traditional pedagogical qualifications. The salary difference between vacancies with digital skills and those without is large. At first glance, this gap suggests a market premium for digital skills. But this is not a clear-cut interpretation: higher-paying employers may also demand more digital skills, creating a measurement problem. This paper examines whether the observed salary gap reflects genuine skill valuation, employer heterogeneity, or a combination of both.

Kazakhstan is a pertinent and understudied case. Since 2017, the country has developed a national digital education agenda aligned with UNESCO's ICT Competency Framework for Teachers (ICT-CFT v.3) [1]. This process culminated in the introduction of a three-tier digital competency model for pedagogical staff in 2021 [2]. Despite this institutional framework, no vacancy-based study has examined whether these competencies are reflected in salaries or whether the observed salary differences are instead driven by employer characteristics. NLP applied to online job postings has become an established methodology for measuring skill demand in real time [3–5]. More broadly,

 doi:10.20944/preprints202605.1145.v1

this study contributes to the growing field of computational labor economics, which treats online vacancy postings as high-frequency indicators of employer demand [6]. Previous research shows that vacancy text captures emerging skill requirements earlier than conventional survey instruments [6], while qualitative job characteristics embedded in vacancy descriptions explain a substantial share of posted wage variation [9].

In labor economics, dictionary-based classification remains a widely used approach because it does not require labeled training data and produces interpretable and auditable output [7,8]. More recent studies have applied transformer-based models such as JobBERT [10] to English-language vacancy data, achieving higher recall but requiring domain-specific annotated corpora. Under these data conditions, a dictionary-based approach represents a defensible methodological choice for this dataset, which is Russian-language, domain-specific, limited in size (n = 755), and lacks an annotated corpus.

The technical contributions of the study comprise three components. First, it develops a seven-category skill taxonomy grounded in ICT-CFT and DigCompEdu for Russian-language educational vacancy text. Second, the internal consistency of the classification approach is evaluated against a TF-IDF + Logistic Regression baseline (AUC = 0.928). Third, the NLP output is integrated into a multi-specification regression framework that includes an employer-type proxy and quantile regression — a combination not previously applied to educational labor markets in Central Asia. The overarching analytical contribution is the decomposition of the observed salary–digital gap into a skill-demand component and an employer-sorting component, demonstrating that a substantial share of the apparent premium reflects employer heterogeneity rather than uniform market pricing of digital competencies.

However, no vacancy-based study of Central Asian educational labor markets has explicitly decomposed the observed digital salary gap into skill-demand and employer-sorting components, nor examined whether this association varies across the salary distribution using quantile regression methods. This paper addresses this gap.

This paper addresses three research questions:

RQ1: Are digital skill requirements in educational job postings associated with higher advertised salaries in the urban labor markets of Almaty and Astana?

RQ2: Which digital skill categories show statistically significant salary associations after controlling for occupation and employer-type proxy?

RQ3: Is the association robust across alternative model specifications, subsamples, and the exclusion of dominant employers?

The study makes both methodological and substantive contributions. Methodologically, it provides a replicable NLP pipeline for Russian-language vacancy text and evaluates its internal consistency against a TF-IDF + LR baseline (AUC = 0.928). It also applies a multi-specification regression framework to test the stability of the salary-digital association across specifications and subsamples.

Substantively, the findings point to a measurement problem: the observed digital salary gap reflects both skill demand and employer differences. Firm effects are not directly identified; however, the analysis provides suggestive evidence of employer-related confounding through subsample regressions and an employer-type proxy. The quantile results are particularly important because the salary-digital gradient increases from Q25 to Q90, a pattern more consistent with labor market segmentation than with a uniform skill premium. Such a pattern is unlikely to emerge without employer-type sorting. Similar concerns have also been noted in the broader vacancy-based wage literature.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the data and methodology. Section 4 presents the results. Section 5 discusses the findings and limitations.

 doi:10.20944/preprints202605.1145.v1

2. Materials and Methods

2.1. NLP for Job Market Analytics

Vacancy-based skill demand analysis grew substantially after Burning Glass Technologies [6] showed that job postings encode granular occupational requirements that rarely surface in official statistics. Deming and Kahn [8] argue that vacancy text captures emerging skill profiles before survey instruments can detect them; Squicciarini and Nachtigall [11] extend this insight to AI skill demand across OECD countries. In education specifically, Li and Huang [12] applied text mining to Chinese university postings, and Ilomäki et al. [13] documented digital competency gaps in European teacher labor markets. Recent studies extend NLP-based vacancy analysis to additional national contexts, including Lithuania and China, demonstrating the growing cross-national applicability of automated skill classification methods [14,15]. Dictionary-based NLP remains the standard in labor economics for its auditability [7,9]. By contrast, transformer models such as JobBERT [10] offer better recall but presuppose annotated training data that do not exist for our domain.

2.2. Employer Sorting and Salary Confounding

A less-discussed difficulty in vacancy-based analysis is that the same entity, the employer, sets both the advertised salary and the skill requirements. Mortensen and Pissarides [16] formalize this in a search-matching model, where observed salary dispersion partly reflects employer heterogeneity rather than worker characteristics. Sorkin [17] provides firm-level evidence showing that employer effects on salaries are large and persistent. In addition, recent studies using online job boards show that platform structure and employer concentration can influence observed wage distributions in vacancy data [18]. Recent vacancy-based research also documents measurable digital skill premia in posted wages. Using UK vacancy data, Garcia-Lazaro et al. [19] estimate a within-occupation digital wage premium of approximately +5.8%, with larger premia associated with advanced digital and AI-related skills. Their findings suggest that vacancy data can capture economically meaningful returns to digital competencies while also highlighting the importance of occupational and institutional controls in interpreting those returns. This reinforces the concern that salary–skill associations in online postings partly reflect employer composition rather than pure skill valuation. In the Kazakhstani education sector, the public-private salary differential is well documented [2]. Online and EdTech platforms often operate outside collective bargaining and set salaries more flexibly. As a result, any salary–digital-skill association in vacancy data risks conflating valuation with employer characteristics. The empirical strategy addresses this using an employer-type proxy and subsample regressions. This confound is not specific to Kazakhstan. Azar et al. [18] demonstrate, using US vacancy data, that employer concentration in local labor markets systematically suppresses wages independently of skill requirements — conceptually similar to the public-sector salary grid effect documented in the present study. More broadly, the posted wage literature has established that advertised salaries are strategic signals rather than neutral market prices: employers in competitive talent markets post higher wages to attract applicants, while monopsonistic employers post lower wages without necessarily losing candidates [20,22]. In digital labor markets, among EdTech platforms, this dynamic may be amplified. The World Economic Forum [21] and OECD [22] identify education as one of the sectors with the fastest-growing demand for digital skill. This is consistent with the higher salary associations observed for platform and CRM competencies in the present dataset.

2.3. Digital Competency Frameworks and Wage Returns

UNESCO's ICT-CFT (v.3) [1] and the European DigCompEdu [13] both organize digital teacher competencies into progressive levels and inform the taxonomy used here. The TPACK model [24,25] provides a theoretical account of why digital-pedagogical knowledge is multidimensional. In terms of salary returns, the general labor economics literature consistently finds positive returns to digital

 doi:10.20944/preprints202605.1145.v1

skills [26,27], though these returns are compressed in the public sector by collective wage-setting structures [17]. What the existing literature does not address is the role of employer type as a driver of digital stratification: whether the salary premium associated with digital requirements reflects worker skill or employer characteristics. This paper examines this issue within a broader empirical framework.

3. Materials and Methods

3.1. Data and Sampling

We collected data from HeadHunter (HH.kz) via its JSON API, targeting the educational labor markets of Almaty (region code 160) and Astana (159) over the twelve-month period from November 2024 to November 2025. This window captures a complete annual hiring cycle in the Kazakhstani educational sector, including peak recruitment periods at the start of the academic year (August–September) and mid-year replacement waves (January–February). Search terms included four occupational keywords: prepodavatel (instructor), vospitatel (childcare educator), metodist (methodologist), and tyutor (tutor). We retained only vacancies with voluntarily disclosed salary information. Non-KZT figures were converted at prevailing market rates. HTML markup was stripped from all description fields. After deduplication, the raw dataset comprised 794 vacancies; removing implausible salary values (below 50,000 or above 800,000 KZT, affecting less than 1% of records) left n = 755 for analysis.

API access details: the HH.kz JSON API is publicly accessible without authentication for vacancy search endpoints; results are paginated at 100 items per page and were iterated exhaustively across all pages for each keyword-region combination. Duplicate removal was performed by vacancy ID, retaining the earliest posting date for repeated listings. Non-KZT salary figures (USD, EUR) were converted to KZT using the National Bank of Kazakhstan mid-market rates for the respective posting date, retrieved via the NBK public API.

Sampling limitation. The dataset covers only postings with voluntarily disclosed salaries, a non-random subset of educational vacancies. If salary disclosure is systematically related to digital intensity or sector, estimates may not generalize to all vacancies. The dataset further reflects the platform-mediated and formally advertised segment of the educational labor market in Almaty and Astana rather than the full universe of educational hiring. Vacancies filled through internal recruitment, state allocation mechanisms, or informal referral networks are not observed. This likely leads to an overrepresentation of private schools, language centres, and EdTech employers on HH.kz.

3.2. NLP Pipeline

Given the size and language of the dataset, a dictionary-based approach was selected as the primary method. This approach is widely used in labor market research due to its transparency and interpretability [8,9]. A TF-IDF+Logistic Regression model was used as a consistency check to assess how well the dictionary captures patterns in the data (AUC = 0.928). No transformer-based model for Russian-language educational vacancy text is publicly available, and no annotated training corpus exists for this domain. This choice fits the data constraints and follows standard practice [18]. The pipeline comprises four stages. (1) Preprocessing: HTML was stripped via regex; text was lowercased; punctuation and whitespace were normalized. (2) Tokenization: whitespace tokenization for the dictionary classifier; sklearn TfidfVectorizer with unigram+bigram vocabulary (max_features = 5,000; min_df = 2) for ML baselines. (3) Feature extraction: each of the seven skill categories was coded as a binary indicator (1 if any pattern matched in the cleaned description, 0 otherwise). (4) Classification: dictionary classifier as the primary instrument; LR and RF as internal consistency benchmarks. The aggregate label has_digital = 1 if any category label equals 1.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 May 2026 doi:10.20944/preprints202605.1145.v1

Table 1. NLP Pipeline Architecture

Stage	Component	Implementation	Output
Preprocessing	HTML stripping; lowercasing; punctuation/whitespace normalization	re.sub(); str.lower(); Python 3.11	Clean plain text per vacancy
Tokenization	Whitespace tokenization; unigram+bigram construction	TfidfVectorizer(ngram_range=(1,2))	TF-IDF matrix 755×5,000
Feature extraction	Regex dictionary matching; TF-IDF for ML baselines	re.search(IGNORECASE); TfidfVectorizer(max_features=5000, min_df=2)	Binary 7-dim. vector; sparse TF-IDF matrix
Classification	Dictionary classifier (primary); LR+RF (consistency check)	Custom regex; LogisticRegression; RandomForestClassifier (sklearn 1.3)	has_digital label; 5-fold CV metrics

3.3. Digital Skill Taxonomy

Skill extraction was implemented using a dictionary-based approach built on regular expressions (regex) and TF-IDF methods. Recent studies increasingly rely on more complex NLP architectures. For example, Lukauskas et al. [14] use BERT embeddings combined with UMAP/HDBSCAN clustering to identify latent occupational groupings, while Chen et al. [15] apply LLM-based agents to extract structured information from healthcare vacancy postings with high reported accuracy.

In the present study, the choice of a dictionary-based approach reflects the need for interpretability and auditability in the context of educational policy research. Given the limited availability of annotated datasets for Russian- and Kazakh-language educational vacancy text, highly parameterized neural models are difficult to validate transparently. By contrast, regex-based classification allows direct inspection of each competency category (G1–G7) and provides a more reproducible framework for estimating salary associations.

Table 2. Digital Skill Taxonomy (n = 755)

Category	Regex patterns (representative)	n	%
G1: Office tools	excel|word|powerpoint|google docs|microsoft office	64	8.50%
G2: Comm. platforms	zoom|teams|google meet|webinar|distance	45	6.00%
G3: Social media	instagram|telegram|smm|seo|targeted	30	4.00%
G4: Educ. platforms	moodle|google classroom|lms|e-learning	22	2.90%
G5: Design & media	canva|figma|photoshop|videoediting	30	4.00%
G6: CRM & management	crm|1c|bitrix|trello	45	6.00%
G7: Programming & data	python|sql|javascript|machine learning|ai|autocad	20	2.60%

3.4. Internal Consistency Assessment

Methodological note: No human-annotated ground-truth corpus exists for this dataset. The supervised ML classifiers were therefore trained and evaluated against dictionary- derived labels, not against an independent gold standard. The reported metrics (F1 = 0.755, AUC = 0.928) reflect internal lexical consistency- the extent to which an independent data-driven model can reproduce the dictionary's decisions - not external classification validity. This limitation is typical for under-resourced NLP settings.

 doi:10.20944/preprints202605.1145.v1

An independent word-frequency based model reproduces the dictionary output with high fidelity. This suggests that the categories capture real lexical patterns in the vacancy text. Random Forest achieves AUC = 0.939 at lower recall (F1 = 0.615). Bootstrap resampling (B = 100, 80% fraction) confirms that the median salary differential is stable: 17.3% ± 2.6% (95% CI: [12.1%, 22.5%]).

External validation (manual audit, n = 30). To assess the external validity of the dictionary labels, a random sample of 30 vacancies was drawn using a fixed seed (seed = 42) and audited manually by a single researcher. Inter-rater reliability was not formally assessed; the audit should therefore be interpreted as an indicative consistency check rather than a validated inter-rater agreement measure. Comparing manual labels to dictionary-derived labels yields: Precision = 0.778, Recall = 0.824, F1 = 0.800, Accuracy = 0.767 (TP = 14, TN = 9, FP = 4, FN = 3). Error analysis identified two systematic error types. False positives (n = 4) arose from overly broad phrases: generic computer-literacy expressions (e.g., "office tools"), a soft-skill term ("communication skills") matched by a communication platforms pattern, an ambiguous design-related phrase ("design & media production"), and a social media mention in a student-outreach context that did not signal a genuine tool requirement. False negatives (n = 3) arose from paraphrase variation: platform-adjacent expressions such as "communication platforms" and "office tools" fell outside tool-name regex patterns, and one vacancy implied the use of educational platforms without naming them explicitly. These errors reflect typical dictionary-based NLP limitations: lower recall under paraphrase and ambiguity in terms. The manual F1 of 0.800 is modestly higher than the TF-IDF+LR internal consistency F1 of 0.755, suggesting the dictionary is somewhat more conservative than a data-driven classifier and suggests comparable external accuracy on this sample. Given the small sample size, these estimates should be treated as indicative rather than definitive.

3.5. Statistical Analysis and Robustness Strategy

Salary distributions are right-skewed (Breusch-Pagan tests, $p < 0.01$ in all OLS models). All regressions therefore use log(salary KZT) as the dependent variable with HC3 heteroskedasticity-robust standard errors. Medians serve as the primary descriptive measure. The Mann-Whitney U test is used for bivariate group comparisons. Welch's t-test and Cohen's d are reported as supplementary statistics. The employer-type proxy classifies companies by name-keyword matching into four categories: online/EdTech (n = 48), private school (n = 73), government/public (n = 60), and unclassified reference (n = 574). Manual validation on a random sample of 50 companies yields precision = 92% and a false positive rate of 8%. The proxy is used as a sensitivity check in M4; M2 remains the primary specification. Measurement error in the proxy may attenuate employer-type coefficients, so the reduction in has_digital under M4 represents a lower bound on the employer-sorting contribution.

A supplementary cluster analysis is reported in Appendix A.

Five robustness checks assess the stability and sources of the main association:

RC1: Excluding [top-5] employers by vacancy count (n = 689), to reduce the influence of dominant recruiters.

RC2: Instructors only (n = 572), the largest and most internally homogeneous occupational subgroup.

RC3/RC4: City-stratified regressions (Almaty, n = 384; Astana, n = 371) to assess geographic heterogeneity.

RC5: Employer-proxy model excluding [top-5] employers (n = 689), combining both adjustments.

A median (quantile, q = 0.5) regression is estimated as an alternative to OLS for the primary and employer-proxy specifications. Three OLS model specifications are compared: M2 (base controls: occupation + city), M4 (M2 + employer-type proxy), and M3 (all seven skill categories + controls). An employer-type limitation worth noting: full employer fixed effects- e.g., the Abowd, Kramarz and Margolis [1999] worker-firm decomposition- would be the econometrically appropriate tool for controlling employer heterogeneity. This is not feasible in a cross-sectional vacancy dataset where

 doi:10.20944/preprints202605.1145.v1

most employers appear only once or twice (median firm vacancy count = 1). The proxy approach is therefore a partial and limited correction, not a definitive control. Longitudinal matched employer-employee data would be needed to address this properly.

4. Results

4.1. Descriptive Statistics and Skill Demand Structure

Of the 755 vacancies analyzed, 179 (23.7%) specify at least one digital requirement. Disaggregated by occupational group: instructors 25.2%, methodologists 51.7%, childcare educators 3.6%, tutors 0.0%. By city: Almaty 22.7%, Astana 24.8%. The most commonly demanded categories are office tools (G1, 8.5%), CRM/management software (G6, 6.0%), and communication platforms (G2, 6.0%). Educational platform proficiency (G4, 2.9%) and programming/data skills (G7, 2.6%) appear least often- a noteworthy result, given that ICT-CFT Level 3 priorities emphasize precisely these skills.

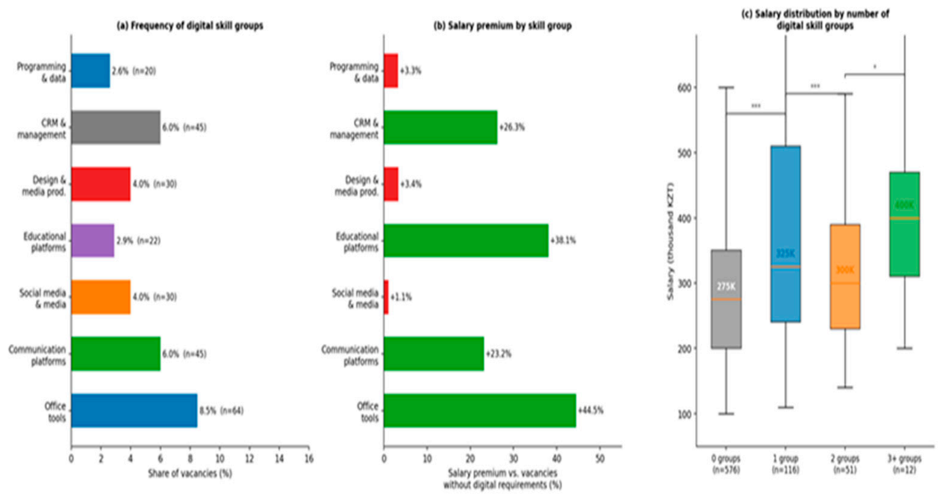

Figure 1. Digital skill demand structure (n = 755). (a) Category frequency; (b) Median salary difference vs. no-digital baseline; (c) Salary distribution by number of digital skill categories.

4.2. Bivariate Salary Association (RQ1)

Vacancies that list at least one digital requirement carry a median advertised salary of 325,000 KZT, against 275,000 KZT for postings with no digital requirements, a gap of 50,000 KZT, or +18.2%. The Mann-Whitney U test confirms this difference is statistically significant ($U = 65{,}298$, $p < 0.001$). Cohen's $d = 0.526$ signals a moderate effect. The mean gap is larger (+26.1%), reflecting right-skewed salaries. The median is therefore used as the primary measure throughout.

 doi:10.20944/preprints202605.1145.v1

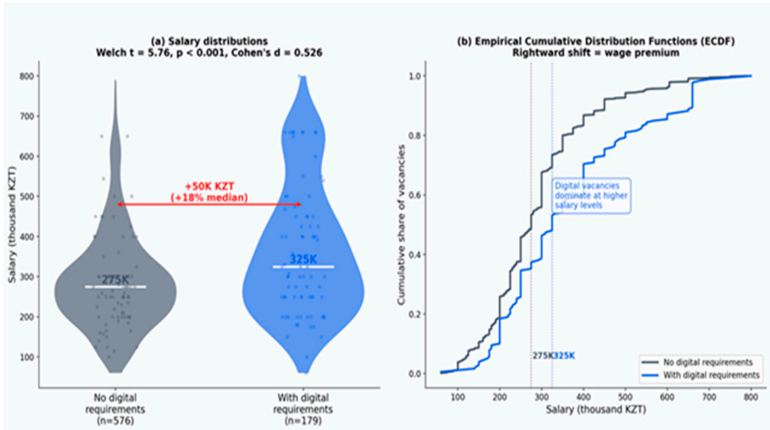

Figure 2. Salary distribution comparison. (a) Violin plots (Mann-Whitney U = 65,298, $p < 0.001$; Cohen's d = 0.526); (b) ECDF - rightward shift of the digital-requirement distribution is consistent across percentiles.

4.3. Regression Results (RQ1, RQ2)

Under M2 (occupation and city controls), the has_digital coefficient is 0.200 (HC3 SE = 0.040, $p < 0.001$), corresponding to approximately +22.2%. Adding an employer-type proxy in M4 reduces this to 0.141 (HC3 SE = 0.040, $p < 0.001$, +15.2%). The online/EdTech employer dummy is estimated at 0.348 (+41.6%, $p < 0.001$). This pattern is consistent with employer sorting. Part of the association is absorbed by employer controls. The adjusted R^2 across all specifications is 0.045–0.093, indicating that the available covariates capture a limited share of salary variation. Key variables such as contract type, experience, and firm size are not observed.

Table 3. OLS Regression Results: Dependent Variable = log(Salary KZT); HC3-Robust Standard Errors.

Variable	M2 Base	M4 +Employer proxy	M3 All 7 cat. +ctrl	Median reg. (M4 spec)
has_digital	0.200*** (0.040)	0.141*** (0.040)	-	0.087 [+9.1%]
G1: Office tools	-	-	0.331*** (0.059)	-
G2: Comm. platforms	-	-	0.127. (0.070)	-
G4: Educ. platforms	-	-	0.204* (0.098)	-
G6: CRM & management	-	-	0.148* (0.068)	-
G3,G5,G7 (ns)	-	-	ns (see Table 5)	-
emp_online (+41.6%)	-	0.348*** (0.086)	-	0.728 [+107%]
emp_private (−14.5%)	-	− 0.157** (0.054)	-	−0.202
city_bin (Astana)	−0.004 ns	0.000 ns	ns	+0.061 ns
prof_childcare educator	−0.103** (0.035)	−0.096** (0.035)	−0.107* (0.045)	−0.095
adj. R^2	0.046	0.093	0.07	Median regression (LAD)
n	755	755	755	755
Breusch-Pagan p	0.0009	0.0002	-	n/a

HC3- robust standard errors in parentheses for OLS models. Significance: *** p<0.001, **p<0.01, *p<0.05, .p<0.1, ns = not significant. Reference category: Almaty/instructor. Median regression (LAD; least absolute deviations) via linear programming; standard errors not reported (bootstrap SE available on request). Coefficients in percent = exp(b)−1.

Note on model fit: The adjusted R^2 of 0.046–0.093 across specifications is consistent with the vacancy-based wage modelling literature, where individual-level covariates such as experience, qualifications, and contract type are unavailable. Low explanatory power in this context reflects data constraints rather than model misspecification.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 May 2026 doi:10.20944/preprints202605.1145.v1

4.4. Category-Level Associations (RQ2)

Table 4. Salary Association by Skill Category: Descriptive and OLS (M3) Estimates

Skill Category	n	Median (KZT)	vs. baseline	M3 OLS coef. (HC3 SE) / significance
No digital (baseline)	576	275,000	-	Reference
G1	64	387,500	+40.9%	0.331 (0.059) *** / +39.2%
G2	45	350,000	+27.3%	0.127 (0.070) . / +13.5%
G3	30	300,000	+9.1%	−0.117 (0.083) ns
G4	22	400,000	+45.5%	0.204 (0.098) * / +22.6%
G5	30	262,500	−4.5%	−0.141 (0.084) ns
G6	45	325,000	+18.2%	0.148 (0.068) * / +15.9%
G7	20	250,000	−9.1%	−0.095 (0.099) ns [see note]

G7 Programming & data: n = 20; coef. = −0.095, p = 0.335. The negative coefficient should not be interpreted as evidence that programming skills are unrewarded. This is a composition-driven bias: in this sample, programming requirements appear almost exclusively in public-school IT teacher postings, where salaries are institutionally capped by state salary grids, independent of any skill premium. The results reflect sample composition rather than the broader market valuation of programming skills in the educational sector.

After occupational and city controls (M3), three categories retain statistically significant positive associations: office tools (G1, +39.2%, $p < 0.001$), educational platforms (G4, +22.6%, $p = 0.038$), and CRM/management tools (G6, +15.9%, $p = 0.031$). Communication platforms (G2) show a marginal positive association (+13.5%, $p = 0.068$). The programming and data category (G7) shows a small negative coefficient (−0.095, $p = 0.335$). This likely reflects a composition-driven artifact. G7 requirements in this sample are concentrated in public-school IT teaching positions subject to fixed state salary grids, which mechanically suppresses the coefficient. Including private-sector EdTech and university IT positions would likely yield different results for this category.

4.5. Non-Linearity and Threshold Effects

Table 5. Median Salary by Number of Digital Skill Categories

Categories	n	Median (KZT)	vs. 0-cat.	Mann-Whitney vs. baseline
0 (no digital)	576	275,000	-	Baseline
1 category	116	325,000	+18.2%	p < 0.001***
2 categories	51	300,000	+9.1%	p = 0.074 (ns)
3+ categories	12	400,000	+45.5%	p < 0.001*** (n=12, caution)

The two-category group (n = 51) shows no significant difference from baseline; the three-or-more group (n = 12) is too small for reliable inference.

4.6. Quantile Regression Analysis

OLS estimates the association at the conditional mean of log(salary). To assess whether the digital-requirement premium varies across the salary distribution, we report quantile regressions at five quantiles ($q = 0.10, 0.25, 0.50, 0.75, 0.90$) using the Koenker-Bassett estimator. Bootstrap standard errors are computed via stratified resampling ($B = 500$). Table 6 and Figure 3 present results.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 May 2026 doi:10.20944/preprints202605.1145.v1

Table 6. Quantile regression estimates of the has_digital coefficient across the salary distribution (M2 specification; bootstrap SE, B = 500).

Quantile	Salary (KZT, indicative values.)	Coef.	Bootstrap SE	t	p-value	% change in salary
Q10 (lowest 10%)	~180,000	0.223	0.093	2.39	0.017*	+25.0%
Q25	~230,000	0.140	0.060	2.32	0.020*	+15.0%
Q50 (median)	~287,500	0.167	0.056	2.96	0.003**	+18.2%
Q75	~350,000	0.251	0.076	3.29	0.001***	+28.6%
Q90 (highest 10%)	~500,000	0.368	0.061	5.99	<0.001***	+44.4%

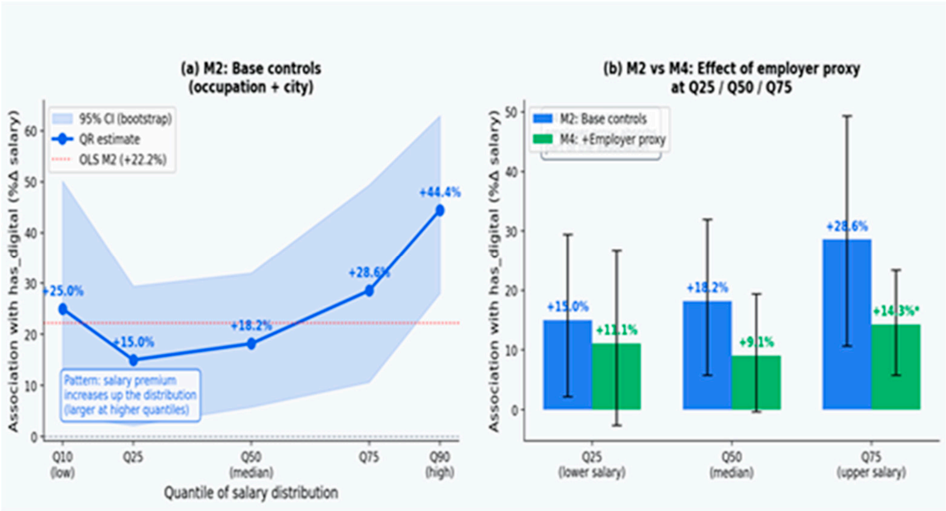

Figure 3. Quantile regression analysis. (a) has_digital coefficient at Q10–Q90 with 95% bootstrap CI and OLS M2 reference; (b) M2 vs. M4 estimates at Q25/Q50/Q75.

Figure 3 confirms this pattern. The coefficient increases across quantiles and declines under the employer-proxy specification. The results show an increasing gradient across the salary distribution. At Q25, the premium is +15.0% ($p = 0.020$). It rises steadily to +44.4% at Q90 ($p < 0.001$). Two mechanisms may explain this pattern. First, higher-paying employers, online platforms and premium EdTech providers – are more likely to specify digital requirements, concentrating the association at upper salary quantiles. Second, within a given employer, high-skill roles may command larger proportional premia. However, the design cannot distinguish between these two explanations. This pattern makes it unlikely that the aggregate association is driven purely by a floor effect at the low end of the distribution.

Applying the employer- proxy specification (M4) at Q25, Q50, and Q75 reduces the has_digital coefficient across all three quantiles. At Q25, the M4 coefficient (+11.1%, $p = 0.117$) is no longer statistically significant. The lower-salary segment is most sensitive to employer composition. At Q75, the M4 coefficient (+14.3%) remains statistically significant at $p < 0.001$. The digital–salary association at the upper end of the market holds after employer-type control. At the lower end, the evidence is less consistent.

4.7. Robustness Checks (RQ3)

Table 7. Robustness Check Summary: has_digital Coefficient across Specifications

Specification	n	Coef.	HC3 SE	p-value	%Δ	adj. R^2
M2: Base (occ. + city)	755	0.200	0.040	<0.001***	+22.2%	0.046
M4: + Employer proxy	755	0.141	0.040	<0.001***	+15.2%	0.093

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 May 2026 doi:10.20944/preprints202605.1145.v1

RC1: Excl. top-5 employers	689	0.121	0.040	0.003**	+12.8%	0.027
RC2: Instructors only	572	0.255	0.043	<0.001***	+29.0%	0.053
RC3: Almaty only	384	0.097	0.054	0.075 (ns)	+10.2%	0.031
RC4: Astana only	371	0.302	0.059	<0.001***	+35.2%	0.075
RC5: Proxy + excl. top-5	689	0.111	0.041	0.007**	+11.7%	0.033

HC3- robust standard errors; dependent variable = log(salary KZT). Significance: ***p<0.001, **p<0.01, ns = not significant. RC3 (Almaty only) is the only specification where has_digital is not significant at the 0.05 level.

The has_digital association is consistent in sign across all specifications. Point estimates range from +10.2% (Almaty only, not significant) to +35.2% (Astana only). The estimate contracts when the employer- type proxy is added (M4: +15.2%). It also contracts when the top-5 employers are excluded (RC1: +12.8%). Notably, the Almaty-only subsample (RC3: p= 0.075) is the only specification where the association does not reach conventional significance. This limits the generalizability of the aggregate results.

4.8. Occupational Heterogeneity

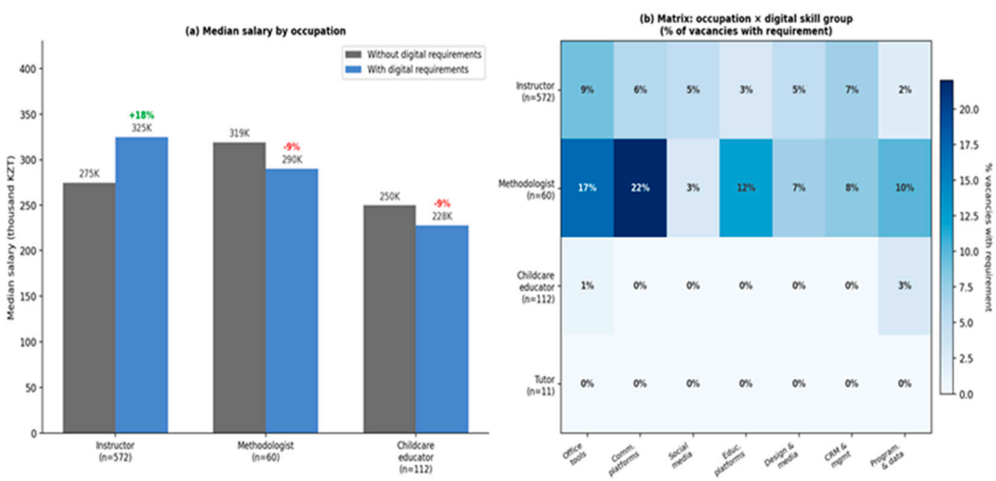

Figure 4. Salary associations by occupation. (a) Median salary with vs. without digital requirements; (b) Skill demand matrix by occupation and category.

The salary association is statistically significant only within the instructor group (RC2:+29.0%, p<0.001), which accounts for 75.8% of the sample (n=572). For methodologists, a compositional artifact appears to reverse the apparent sign. This interpretation is consistent with, though not directly confirmed by, the employer-level patterns discussed in Section 4.9. For childcare educators, only 4 vacancies specify digital requirements. This is too few to support reliable inference.

4.9. Employer Analysis

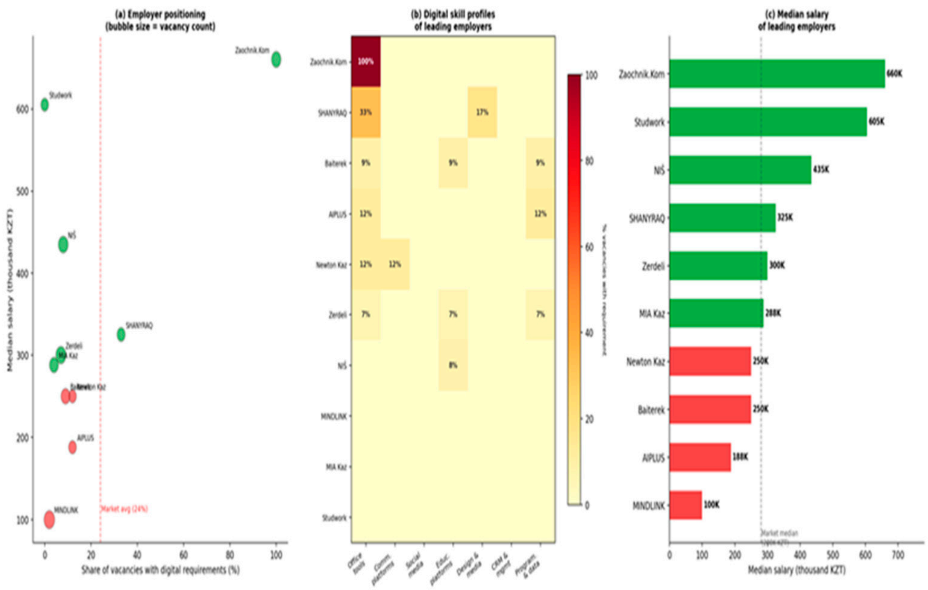

Figure 5. Leading employer analysis. (a) Employer positioning by digital rate vs. median salary; (b) Skill profiles heatmap; (c) Median salary ranking of employers.

The employer-level data provide evidence consistent with a sorting mechanism. Online academic services platforms (e.g., Zaochnik.com, Studwork) exhibit near-100% digital requirement rates and the highest median salaries in the sample (600,000–660,000 KZT). Nazarbayev Intellectual Schools offer competitive salaries (416,000 KZT) with relatively few digital requirements, consistent with institutional salary-setting decoupled from tool demands. This employer-level heterogeneity motivates the inclusion of the employer-type proxy in M4. It is discussed further in Section 5.

4.10. NLP Internal Consistency

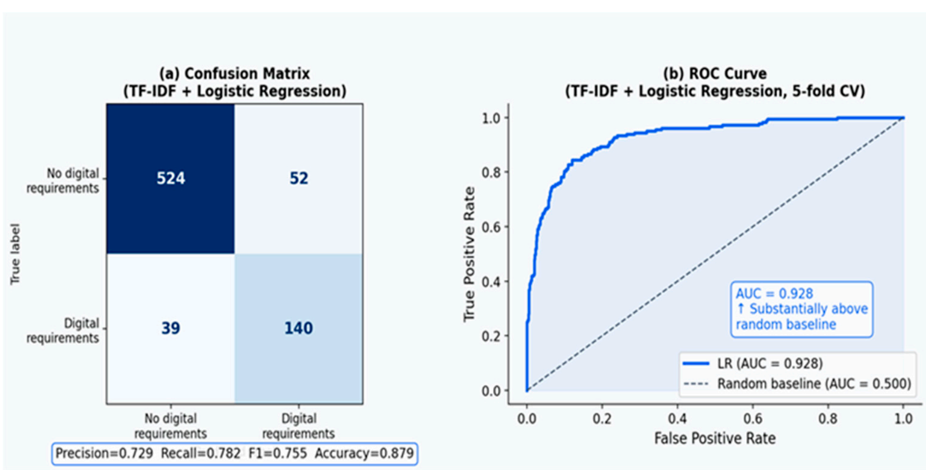

Figure 6. Internal consistency of the NLP pipeline (5-fold stratified cross-validation against dictionary labels). (a) Confusion matrix; (b) ROC curve (AUC = 0.928).

 doi:10.20944/preprints202605.1145.v1

Table 8. Internal Consistency Assessment: ML Baselines vs. Dictionary Labels (5-fold CV)

Method	Accuracy	Precision	Recall	F1	AUC
Dictionary classifier (primary)	-	-	-	-	-
TF-IDF+Logistic Regression	0.879	0.729	0.782	0.755	0.928
TF-IDF+Random Forest	0.862	0.912	0.464	0.615	0.939

No independent ground truth is available. ML metrics therefore reflect internal consistency against dictionary labels rather than external classification validity.

Across all specifications, the association is positive and statistically significant. However, the central contribution lies not in establishing the existence of this association, which is consistent with prior literature, but in clarifying its underlying structure.

5. Discussion

In every specification, the association between digital skill requirements and advertised salaries remains positive and statistically significant. The analysis focuses on the structure of this relationship rather than on its existence alone.

The magnitude of the estimated premium warrants comparison with findings from related studies. Li and Huang [12], analyzing Chinese university job postings, report a positive association between digital skill requirements and salary levels, which they interpret as a return to productive skills. In their study of teacher labour markets in Europe, Ilomäki et al. [13] observe that digital competencies increasingly function as baseline expectations rather than as direct salary differentiators in advanced digital economies.

The estimated associations in the present study are higher than the European benchmarks and also exceed the average premium reported in cross-national vacancy studies for lower-middle income countries [26]. The estimated coefficients in the present study are higher than those reported by Garcia-Lazaro et al. [19], who identify a within-occupation digital skill premium of approximately +5.8% using UK vacancy data. Several factors may explain this difference. First, the UK labour market is more institutionally mature and characterized by more standardized occupational wage structures, which may compress observable digital premia within occupations. Second, the Kazakhstani educational labour market appears substantially more segmented by employer type, particularly between public-sector institutions and flexible-wage private or EdTech employers. Third, the occupational controls used by Garcia-Lazaro et al. are considerably more detailed and likely absorb part of the employer-sorting variation that remains visible in the present dataset. In this sense, the reduction of the estimated coefficient from +22.2% in M2 to +15.2% in M4 brings the estimates closer to the magnitude reported in the UK literature rather than contradicting it. One possible explanation is scarcity-driven pricing: employers operating in flexible-wage segments of the educational market, particularly online platforms and EdTech providers, are both more likely to demand digital competencies and more likely to offer higher salaries.

However, the employer-type analysis suggests that this explanation is only partial. In the public sector, salary structures are often determined by standardized pay grids that weaken the relationship between skill requirements and compensation. As a result, part of the observed premium appears to reflect differences in employer composition rather than uniform market pricing of digital skills across the educational labour market.

5.1. Rethinking the "Digital-Requirement Premium"

At first glance, the results appear to support a standard human capital interpretation: vacancies requiring digital skills are associated with higher advertised salaries, with estimated premia ranging

 doi:10.20944/preprints202605.1145.v1

from +11.7% to +22.2% depending on the specification. Considered in isolation, these estimates are broadly consistent with evidence from OECD labour markets. However, the interpretation changes once employer characteristics are taken into account.

When the employer-type proxy is included in the regression, the estimated coefficient declines to +15.2%, and it decreases further when dominant employers are excluded from the sample. In the Almaty subsample, the association is no longer statistically significant ($p = 0.075$). This suggests that part of the observed relationship reflects employer heterogeneity rather than pure market pricing of digital skills.

Higher-paying employers, particularly online platforms and EdTech organizations, are also more likely to specify digital skill requirements in vacancy descriptions. As a result, wages and skill requirements are jointly shaped by employer characteristics, creating a measurement problem in vacancy-based data. The present design does not identify firm effects directly; instead, it provides suggestive evidence of employer-related confounding through subsample regressions and the employer-type proxy.

These findings are consistent with a broader literature emphasizing the role of firm heterogeneity in labour market inequality. In vacancy data, employers influence both salary-setting and the specification of skill requirements.

5.2. Heterogeneity Across the Salary Distribution

Further insight comes from the quantile regression results. The estimated digital premium increases steadily across the salary distribution, from +15.0% at Q25 to +44.4% at Q90. This pattern is difficult to reconcile with the idea of a uniform return to digital skills. Instead, it points to a segmented labour market in which digital requirements are concentrated in higher-paying organizational contexts associated with particular employer types.

Under this interpretation, digital skills function not only as indicators of productivity, but also as signals of access to better-paying institutional environments. The reduction of the estimated effect at lower quantiles after controlling for employer type further supports this interpretation.

5.3. Which Digital Skills Are Rewarded?

At the category level, the results reveal a non-trivial pattern. Skills associated with everyday operational tools — office software, educational platforms, and CRM or management systems — show statistically significant positive salary associations. By contrast, programming and data-related skills do not.

This finding differs from the hierarchy implied by frameworks such as ICT-CFT, where advanced technical competencies occupy the highest levels. As discussed in Section 4, the weak result for G7 is likely driven by sample composition. Programming-related requirements in the dataset are concentrated mainly in public-sector IT teaching positions, where salaries are constrained by standardized state pay schedules.

The absence of a positive association for G7 should therefore not be interpreted as evidence that programming skills lack market value. Rather, it reflects the institutional structure of salary-setting within the sampled educational labour market.

More broadly, the findings suggest that the labour market rewards digital skills embedded in the operational workflows of higher-paying organizations more consistently than advanced technical competencies alone. Operational digital tools such as LMS platforms, CRM systems, and platform-management software are more common in private educational institutions and EdTech organizations, where salary structures are also more flexible.

5.4. Implications for Digital Stratification

Taken together, the findings suggest a form of digital stratification shaped not only by workers' skills, but also by employers. The key divide in the labour market is not simply between workers who

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 May 2026 doi:10.20944/preprints202605.1145.v1

possess digital competencies and those who do not. It also reflects whether workers are employed by organizations that actively demand and reward those competencies.

This shifts the analytical focus away from individual skill acquisition alone toward the broader structure of labour demand. Policies focused exclusively on upskilling teachers — particularly in advanced areas such as programming or AI — may have limited effects on earnings unless accompanied by changes in the types of institutions employing them and, in the salary-setting systems under which those institutions operate.

5.5. Limitations and Interpretation Boundaries

Several limitations should be noted. The main challenge is endogeneity arising from the joint determination of wages and skill requirements by employers. Firms that offer higher salaries may also be more likely to demand digital skills, making it difficult to separate skill pricing from employer characteristics in cross-sectional vacancy data.

The employer-type proxy partially addresses this issue, but full employer fixed effects cannot be estimated in the present setting. Second, the sample is restricted to vacancies with disclosed salaries, limiting external validity. Third, the dataset represents a one-year cross-sectional snapshot (November 2024 to November 2025) and therefore cannot capture long-term trends or causal dynamics.

In addition, important individual-level variables such as experience, qualifications, and contract type are unavailable. The relatively low R^2 values (0.046–0.093) are therefore consistent with the broader vacancy-based wage literature rather than evidence of model failure. Finally, the NLP classification relies on a dictionary-based approach. Although the manual audit suggests acceptable performance (F1 = 0.800), external validation remains limited.

5.6. Contribution and Future Directions

Despite these limitations, the study contributes in two ways. Empirically, it provides one of the first studies to explicitly decompose the observed digital salary gap into skill-demand and employer-sorting components in the educational labour market of post-Soviet Central Asia. Conceptually, it suggests that a substantial share of the observed digital skill premium may reflect employer heterogeneity rather than pure market valuation of skills.

Future research could address these mechanisms more directly through longitudinal data containing employer identifiers. Expanding the dataset across additional regions and sectors would improve generalizability. The development of annotated corpora for Russian-language educational vacancy text would also support more advanced NLP methods and independent validation of skill classifications.

More broadly, digital skill requirements in vacancy postings function not only as indicators of employer demand for specific competencies, but also as signals of the organizational and salary-setting environments associated with those positions. Recognizing this dual role is important for understanding the relationship between skills and wages in contemporary labour markets.

6. Conclusions

The aim of this paper was to examine whether digital skill requirements in educational job postings are associated with higher advertised salaries in urban Kazakhstan. Using 755 validated vacancies from HH.kz together with a dictionary-based NLP pipeline cross-checked against a TF-IDF + Logistic Regression baseline, the analysis consistently identifies a positive association across model specifications.

The bivariate median gap is +18.2%. OLS models with controls estimate the association at +22.2%, while the inclusion of an employer-type proxy reduces the estimate to +15.2%. Excluding dominant employers further lowers the estimate to approximately +11.7–12.8%. The Almaty-only

 doi:10.20944/preprints202605.1145.v1

subsample is the only specification in which the association does not reach conventional statistical significance.

The central implication of the study lies not in the size of the estimated premium itself, but in its interpretation. The relationship between digital requirements and higher salaries in vacancy data appears to be shaped partly by employer heterogeneity. Higher-paying employers are also more likely to specify digital skill requirements in vacancy descriptions. As a result, skill valuation effects cannot be cleanly separated from employer-type effects in cross-sectional vacancy data. The findings should therefore be interpreted as associations in advertised vacancies rather than as estimates of causal wage returns.

The quantile regression results provide additional insight. The estimated salary-digital gradient increases from +15.0% at Q25 to +44.4% at Q90, consistent with labour market segmentation in which digital requirements are concentrated in higher-paying organizational settings. Together, these findings point to a measurement problem common to vacancy-based estimates of skill premia: when employer types differ systematically in both wage-setting and skill requirements, simple salary gaps are likely to overstate the return to skills.

Three skill categories remain statistically significant after controls. These competencies are more characteristic of private educational organizations and online employers than of the public sector. ICT-CFT-aligned upskilling programs prioritize advanced Level 3 competencies; however, these do not necessarily correspond to the current salary-reward structure observed in urban Kazakhstani vacancy data. This interpretation remains tentative given the cross-sectional and correlational nature of the evidence.

Several directions for future research follow from this study. Access to employer-level identifiers would allow the estimation of proper firm fixed effects and provide a more direct test of employer sorting. Longitudinal data could help track how salary-skill relationships evolve as the Kazakhstani labour market continues to digitalize. Human-annotated training corpora would improve external validation of the taxonomy and support the application of more advanced NLP methods. Future work could also employ causal inference designs, such as difference-in-differences or regression discontinuity approaches, to address the identification problem more directly. Finally, expanding the dataset beyond Almaty and Astana would improve understanding of regional differences in digital skill demand across Kazakhstan.

Supplementary Materials: No supplementary materials are associated with this article.

Author Contributions: Conceptualization, K.T.S. and Sh.N.S.; methodology, K.T.S. and Zh.B.B.; software, K.T.S.; validation, K.T.S., Sh.N.S. and Zh.B.B.; formal analysis, K.T.S.; investigation, K.T.S. and G.A.O.; resources, K.T.S., Sh.N.S; data curation, Zh.B.B.; writing-original draft preparation, K.T.S.; writing-review and editing, Sh.N.S., Zh.B.B., G.A.O.; visualization, K.T.S.; supervision, Sh.N.S.; project administration, K.T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The vacancy data were collected via the public HeadHunter (HH.kz) API and are not publicly redistributable due to platform restrictions. Aggregated data, the digital skill taxonomy (G1–G7 regex patterns), and the analysis code are available from the corresponding author upon reasonable request.

Acknowledgments: The authors thank the editors and anonymous reviewers for their constructive feedback.

Conflicts of Interest: The authors declare no conflicts of interest.

 doi:10.20944/preprints202605.1145.v1

Abbreviations

The following abbreviations are used in this manuscript:

NLP	Natural Language Processing
OLS	Ordinary Least Squares
ICT-CFT	ICT Competency Framework for Teachers (UNESCO)
HC3	Heteroskedasticity-Consistent Covariance Estimator (type 3)

Appendix A

Appendix A.1

Appendix A. Supplementary Cluster Analysis

To complement the regression analysis, a k-means cluster analysis was conducted using the three-level ICT-CFT skill matrix (Basic / Advanced / Specialized) to provide a descriptive segmentation of the vacancy dataset. The elbow method suggested an optimal solution at k = 4, corresponding to a 55% reduction in within-cluster sum of squares between k = 3 and k = 4 (75.5 to 34.3). The unusually high silhouette value (s = 0.916) should be interpreted cautiously. In sparse low-dimensional binary datasets, silhouette metrics can be mechanically inflated due to limited feature overlap rather than genuinely strong latent cluster structure. The clustering results should therefore be interpreted as descriptive rather than inferential evidence.

Cluster stability was assessed across 100 random initializations and indicated highly stable cluster assignments. The Davies–Bouldin index at k = 4 (0.445) was the lowest among all k ≤ 4 solutions. At k = 5 and k = 6, the Davies–Bouldin standard deviation increased (0.054 and 0.052), suggesting the emergence of unstable micro-clusters. The k = 4 solution was therefore retained on both statistical and interpretability grounds (Kaufman and Rousseeuw, 2005).

Four descriptive clusters were identified. Cluster 0 (n = 481, 60.9%) consists primarily of vacancies with limited or narrowly specified digital requirements and has a median salary of 300,000 KZT. Cluster 1 (n = 187, 23.7%) includes vacancies without explicit digital requirements and has a median salary of 250,000 KZT. Cluster 2 (n = 72, 9.1%) contains vacancies combining multiple digital competency categories and shows the highest median salary level (350,000 KZT). Cluster 3 (n = 50, 6.3%) is characterized by advanced and specialized digital requirements and also has a median salary of 350,000 KZT. Overall, the cluster structure is broadly consistent with the employer-type heterogeneity discussed in Section 5.2 and should be interpreted as descriptive rather than causal evidence.

References

1. UNESCO. ICT Competency Framework for Teachers, Version 3. Paris: UNESCO, 2018.
2. Ministry of Education and Science of the Republic of Kazakhstan. Order No. 472 on qualification requirements for pedagogical staff. Astana, 2021.
3. Hershbein B., Kahn L.B. Do recessions accelerate routine-biased technological change? Evidence from vacancy postings. American Economic Review. 2018. Vol. 108(7). P. 1737–1772. https://doi.org/10.1257/aer.20161570
4. Carnevale A.P., Jayasundera T., Repnikov D. Understanding Online Job Ads Data. Washington: Georgetown University CEW, 2014.
5. Colombo E., Mercorio F., Mezzanzanica M. AI meets labor market. Computers in Human Behavior. 2021. Vol. 117. https://doi.org/10.1016/j.chb.2020.106658
6. Deming D.J., Kahn L.B. Skill requirements across firms and labor markets. Journal of Labor Economics. 2018. Vol. 36(S1). P. S337–S369. https://doi.org/10.1086/694106
7. Marinescu I., Wolthoff R. Opening the black box of the matching function: the power of words. Journal of Labor Economics. 2020. Vol. 38(2). P. 535–568. https://doi.org/10.1086/705903

 doi:10.20944/preprints202605.1145.v1

8. Burning Glass Technologies. The Quant Crunch. Boston: BGT, 2017.
9. Lev-Aretz Y., Nygaard M. Skills-biased technological change and the labor market. Journal of Labor Research. 2020. Vol. 41. P. 141–163.
10. Decorte P. et al. JobBERT: Fine-tuning BERT for job title classification. arXiv:2109.09605. 2021.
11. Squicciarini M., Nachtigall H. Demand for AI skills in jobs. OECD STI Working Papers 2021/03. https://doi.org/10.1787/3ed32d94-en
12. Li Y., Huang J. Digital competency demand for university teachers in China. Education and Information Technologies. 2022. Vol. 27. P. 6959–6981. https://doi.org/10.1007/s10639-022-10923-x
13. Ilomäki L., Paavola S., Lakkala M., Kantosalo A. Digital competence as an emergent boundary concept. Education and Information Technologies. 2016. Vol. 21(3). P. 655–679. https://doi.org/10.1007/s10639-014-9346-4
14. Lukauskas M., Sarkauskaitė V., Pilinkienė V., Stundžienė A., Grybauskas A., Bruneckienė J. Enhancing skills demand understanding through job ad segmentation using NLP and clustering techniques. Applied Sciences. 2023. Vol. 13(10). 6119. https://doi.org/10.3390/app13106119
15. Chen Y., Zhan X., Yang W., Yan X., Du Y., Zhao T. A NLP analysis of digital demand for healthcare jobs in China. Scientific Reports. 2025. Vol. 15. 14518. https://doi.org/10.1038/s41598-025-98108-9
16. Mortensen D., Pissarides C. New developments in models of search in the labor market. Handbook of Labor Economics. Vol. 3B. 1999. P. 2567–2627.
17. Sorkin I. Ranking firms using revealed preference. Quarterly Journal of Economics. 2018. Vol. 133(3). P. 1331–1393. https://doi.org/10.1093/qje/qjy001
18. Azar J., Marinescu I., Steinbaum M., Taska B. Concentration in US labor markets: Evidence from online vacancy data. Labour Economics. 2022. Vol. 66. 101886. https://doi.org/10.1016/j.labeco.2020.101886
19. Garcia-Lazaro A., Mendez-Astudillo J., Lattanzio S., Larkin C., Newnes L. The Digital Skill Premium: Evidence from Job Vacancy Data. Economics Letters. 2025. Vol. 250. Article 112294. https://doi.org/10.1016/j.econlet.2025.112294
20. Manning A. Monopsony in Motion: Imperfect Competition in Labor Markets. Princeton: Princeton University Press, 2003. https://doi.org/10.1515/9781400850020
21. World Economic Forum. Future of Jobs Report 2023. Geneva: WEF, 2023.
22. OECD. OECD Skills Outlook 2023: Skills for a Resilient Green and Digital Transition. Paris: OECD Publishing, 2023. https://doi.org/10.1787/27452f29-en
23. Redecker C. European Framework for the Digital Competence of Educators: DigCompEdu. Luxembourg: Publications Office of the EU, 2017. https://doi.org/10.2760/178382
24. Koehler M.J., Mishra P. What is TPACK? Handbook of TPACK for Educators. New York: Routledge, 2012.
25. Mishra P., Koehler M.J. TPACK: Revisiting a teacher knowledge framework. Computers & Education. 2017. Vol. 108. https://doi.org/10.1016/j.compedu.2016.09.018
26. Acemoglu D., Autor D. Skills, tasks and technologies. Handbook of Labor Economics. Vol. 4B. 2011. https://doi.org/10.1016/S0169-7218(11)02410-5
27. Grundke R. et al. Who are the most sought-after? OECD STI Working Papers 2018/05. https://doi.org/10.1787/9a2bd220-en