ISSN 3079-7780 (Print)
ISSN 3079-7799 (Online)

BUKETOV BUSINESS REVIEW

2025
Volume 30
Issue 4 (120)

Promoting Innovation

bbr.buketov.edu.kz

ISSN 3079-7780 (Print)
ISSN 3079-7799 (Online)

BUKETOV BUSINESS REVIEW

2025

Volume 30, No. 4 (120)

Founded in 1996

Published 4 times a year

Karaganda
2025

Publisher: **NLC "Karaganda National Research University named after academician Ye.A. Buketov"**
Postal address: 28, Universitetskaya Str., Karaganda, 100024, Kazakhstan
E-mail: businessreview@buketov.edu.kz. Web-site: https://bbr.buketov.edu.kz

Buketov Business Review. — 2025. — Vol. 30, Iss. 4(120). — 49 p. — ISSN 3079-7780 (Print). ISSN 3079-7799 (Online).

Proprietary: NLC "Karaganda National Research University named after academician Ye.A. Buketov".

Registered by the Ministry of Culture and Information of the Republic of Kazakhstan. Rediscount certificate No. KZ16VPY00103920 dated 22.10.2024.

Signed in print 30.12.2025. Format 60×84 1/8. Photocopier paper. Volume 6,1 p.sh. Circulation 200 copies. Price upon request. Order № 184.

Printed in the Publishing house of NLC "Karaganda National Research University named after academician Ye.A. Buketov".

28, Universitetskaya Str., Karaganda, 100024, Kazakhstan. E-mail: printed@karnu-buketov.edu.kz

CONTENTS

https://doi.org/10.31489/2025Ec4/1211

JEL: D14
ORIGINAL RESEARCH

Received: 09.06.2025 | Accepted: 07.10.2025

Analysis of the impact of socio-economic factors on household credit behavior

Bayan Kupeshova[1], **Gulbahyt Zholdasbekova**[2]*, **Murat Isabayev**[3]

Abstract

Purpose of the research is to identify the socio-economic determinants of loan delinquencies among households in Kazakhstan, considering the specifics of the regional context and the spread of digital lending.

Methodology is based on the analysis of cross-sectional survey data from 256 households conducted in 2023 in the Kapshagay region. To test two hypotheses, logit and probit regressions were applied: H1 — the number of loans is positively associated with the risk of default; H2 — low per capita household income increases the probability of delinquency.

Originality / value of the study lies in the use of microdata that reflect actual borrowing practices in an emerging financial market, as well as in the inclusion of demographic characteristics that are rarely considered in borrower assessments. Unlike traditional models employed by banks for automated credit decision-making, this study additionally analyzes variables such as per capita household income. This makes it possible to more accurately capture social vulnerability and potential insolvency risks among different population groups.

Findings show that the number of active loans significantly increases the probability of default, whereas per capita household income does not have a substantial effect. Social status and credit accessibility also proved to be significant factors.

The study contributes to the literature on financial vulnerability and may be useful for regulators and microfinance organizations in developing sustainable lending policies.

Keywords: Default risk, household debt, credit behavior, Kazakhstan, logit-probit models, socioeconomic determinants, digital lending.

Introduction

The relevance of studying household over-indebtedness is of particular importance in the context of implementing the United Nations Sustainable Development Goals (SDGs). In particular, this concerns the objectives of reducing poverty and improving well-being through inclusive economic growth [1]. Excessive debt burden may undermine these goals, as it leads to an increase in the number of socially vulnerable population groups.

According to the World Bank report (December 2024), nearly 19 % of Kazakhstani households are already in a state of financial vulnerability, experiencing difficulties in meeting basic living needs [2]. Moreover, consumer loans—issued without adequate income growth—are increasingly becoming the primary source of financing for these groups. This situation indicates the emergence of systemic risk: under conditions of high inflation and stagnant real incomes, combined with costly borrowing, the probability of widespread defaults increases, threatening both the sustainability of household budgets and the stability of the financial sector as a whole. An additional factor is the technological accessibility of credit, in particular online loans and mobile banking applications, which lowers the entry threshold for borrowing and contributes to rising indebtedness without proper assessment of repayment capacity. While the digitalization of financial services promotes borrowing, it simultaneously amplifies the risk of default, especially among financially vulnerable population groups [3].

At the same time, commercial banks are oriented toward the retail market, since loans to individuals generate higher returns than corporate lending. This institutionally reinforces the banks' motivation to stimulate consumer lending, including among vulnerable groups. Financial vulnerability is most pronounced among households with low incomes and limited financial literacy [4].

The purpose of this article is to conduct a comprehensive study of the factors underlying household over-indebtedness, combining a systematic review of academic literature with empirical analysis using logistic modeling. Particular attention is paid to identifying the socio-economic determinants influencing the

[1]NPJSC "Narxoz University", Almaty, Kazakhstan, bayan.kupeshova@narxoz.kz
[2]NPJSC "Narxoz University", Almaty, Kazakhstan, gulbakhyt.zholdasbekova@narxoz.kz (corresponding author)
[3]NPJSC "Narxoz University", Almaty, Kazakhstan, murat.isabayev@narxoz.kz

probability of loan delinquencies, as well as to developing recommendations aimed at strengthening the financial resilience of the population.

Literature Review and Research Positioning

The analysis of recent scientific publications makes it possible to identify key socio-demographic and economic characteristics that influence household credit behavior. Most researchers agree that variables such as age, gender, education, income, family size, and employment are significant predictors of loan delinquencies. However, approaches to their study vary considerably in both methodological and empirical aspects.

1. Comparison of Methodologies and Variables

The studies of Xiao and Yao, Dempere, and Malik use aggregated household survey data to examine the relationship between family structure and the ownership of different types of debt [5,6,7]. These works apply logit modeling with a focus on mortgage and student loans. In contrast, the present study relies on microdata that include specific cases of delinquencies, which makes it possible to model not only the probability of holding debt but also the actual risk of default.

The studies by Fernández-López S. [8, 9] and Chen F. [10] focus on the role of financial literacy, emphasizing that it can amplify or mitigate the impact of income and employment on borrower behavior. These works employ survey-based financial literacy indices and their interaction with behavioral variables, whereas our model applies an objective indicator—per capita income—which increases sensitivity to social vulnerability without relying on respondents' self-assessment.

Białowolski R. [4] applies a comparative approach between subjective and objective over-indebtedness, revealing that the type of loan and the level of education influence the perception of debt burden. In contrast, our study focuses not on perception but on the actual presence of delinquencies, confirmed by specific payment status.

The study by Xidonas P. [11] uses EU microdata and examines access to credit, including the probability of rejection. It applies binary choice models with an emphasis on employment and housing status. Although the structure of this model is close to that employed in the present study, the key distinction lies in the context: our model is based on data from an emerging financial market—Kazakhstan—where access to digital loans is not accompanied by a centralized borrower assessment system.

2. Financial Literacy and Behavior

The studies by Mutsonziwa and Fant [12], as well as De Oliveira Santini [13], highlight the impact of cross-borrowing and low financial literacy on the propensity for delinquency. However, their focus is on African and Latin American countries, and they rely on survey-based methods. In contrast, our study incorporates variables that reflect the structure of current indebtedness and models the actual behavior of borrowers at the household level.

Diba, Abrantes-Braga, and Veludo-de-Oliveira [14] emphasize "reborrowing" because of distorted perceptions of credit conditions. While we do not directly examine cognitive biases, our findings show that multiple loans (Numlo) significantly increase the risk of default, thereby empirically confirming the consequences described in these behavioral models.

3. Debt Burden, Assets, and Digitalization

Madeira [15] and Białowolski [4] examine the structure of debt and the sensitivity of households to interest rates. These studies rely on macroeconomic panels or aggregated data, whereas our analysis employs individual-level observations, which allows for a more precise assessment of the relationship between the number of loans and delinquency at the microdata level.

Agarwal and Chua [3] raise the important issue of digital lending and its dual impact—on the one hand, expanding access, and on the other, increasing the risk of excessive borrowing. Our study addresses the indirect effects of digitalization, pointing to the lack of systematic information sharing among lenders, which enables borrowers to obtain loans from multiple institutions within a single day.

4. The Kazakhstani Context

In the Kazakhstani context, the study by Mukan M. et al. [16] raises questions about the influence of financial literacy on the choice of microcredit products but does not analyze socio-demographic factors such as age, income, or family structure. Ishuova Zh., Daribayeva M., and Boluspayev Sh. [17,18] focus on macro-level aspects—consumption smoothing and market regulation—whereas the present study concentrates on the individual level, modeling the probability of delinquency as a function of borrowers' socio-economic characteristics.

Thus, the main distinction of this study from the existing literature lies in:

- the use of micro-level data on actual delinquencies, rather than relying solely on self-reported or aggregated indicators;
- the integration of demographic variables into risk assessment models, including rarely used variables such as per capita income;
- the focus on the regional Kazakhstani context, where the specifics of the financial infrastructure (absence of centralized credit scoring, weak digital transparency, active microcredit expansion) require adapted approaches to risk analysis.

This makes it possible to refine existing findings and provide an empirical basis for developing effective scoring tools tailored to the realities of emerging markets.

The empirical base of the study was formed from a cross-sectional sample of 256 households surveyed in 2023 in the Kapshagay region of Almaty oblast. Data collection was carried out through a targeted questionnaire survey, considering the socio-demographic diversity of respondents (age, gender, education, employment, family composition). The resulting data make it possible to quantitatively assess the impact of socio-economic characteristics on credit behavior and the level of default risk.

The current socio-economic situation in Kazakhstan is characterized by the growth of consumer lending and an increasing debt burden, particularly in vulnerable regions. Against this backdrop, a key question arises: which household characteristics increase the risk of loan delinquencies? To address this, the study formulates the main research question: which socio-economic factors have a significant impact on the probability of delinquency?

For empirical verification, two hypotheses are formulated: H1: An increase in the number of active loans is associated with a higher probability of default; H2: Low per capita household income increases the risk of delinquency. Testing these hypotheses will not only confirm the findings presented in the literature but also adapt them to the regional context. Thus, the study aims to refine the factors of financial vulnerability and to provide an evidence base for corrective policies in the field of consumer lending.

Main part of the study

Macroeconomic Context: Inflation and Rising Indebtedness

A comparison of household lending dynamics with persistently high interest rates and inflationary spikes indicates a deterioration in household debt sustainability. Despite rising living costs, borrowing volumes continue to grow, reflecting the escalating problem of over-indebtedness. The situation appears particularly concerning in 2023, when an inflationary surge did not curb credit expansion but, rather, coincided with its acceleration.

Figure 1 presents the dynamics of consumer lending, inflation, and household income in Kazakhstan for the years 2023–2025.

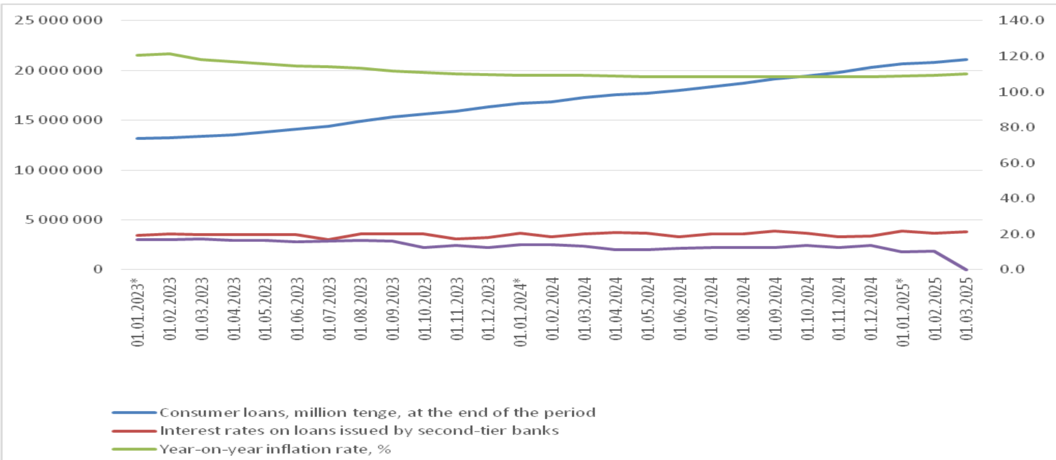

Figure 1 presents the dynamics of consumer lending, inflation, and household income in Kazakhstan for the years 2023–2025.

Note — compiled by the author based on data from the National Bank of the Republic of Kazakhstan [19,20].

As can be seen, the growth of household debt burden significantly outpaces income growth, which exacerbates the problem of over-indebtedness and reduces households' financial resilience. This illustrates how the macroeconomic situation helps explain the micro-level factors identified. An analysis of macroeconomic data for 2023–2025 reveals a pronounced imbalance between the growth of indebtedness and the dynamics of household income. The total volume of loans issued to individuals increased by approximately 70 %—from 13 to 22 trillion tenge—whereas the nominal annual growth of per capita monetary income fluctuated within the range of 10–14 %, with some months even showing a slowdown. Taking into account high inflation, real household income remained almost stagnant or grew only marginally, which reinforced dependence on borrowed funds [19,20].

Of particular concern is that, against the backdrop of inflationary pressure and limited income growth, average interest rates on consumer loans remained at the level of 20–25 % per annum, while in certain segments (e.g., unsecured microfinance) they could exceed 40 %. This makes the debt burden especially heavy for vulnerable household groups. Households are forced to increase borrowing in order to offset the effects of inflation and maintain consumption, while the share of income allocated to debt servicing continues to rise. As a result, their ability to cope with additional expenses or economic shocks deteriorates.

Research Methodology and Statistical Data

To analyze the factors influencing the probability of loan delinquencies among Kazakhstani households, the logistic regression method was applied. This approach is a standard tool for modeling events with binary outcomes, such as the presence or absence of a loan delinquency [21,22]. Logit and probit models make it possible to capture the nonlinear relationship between predictors and the probability of an event, as well as to interpret the results through odds ratios and marginal effects.

The study is based on a cross-sectional sample of 256 households surveyed in 2023 in several localities of Almaty oblast (Kapshagay, Zarechnoye, Kerbulak, and others) [23]. Data collection was carried out within the framework of the *“Society Without Debt”* project of the Uly Dala Association for Rural Business Development, using a stratified questionnaire survey that covered the main socio-demographic characteristics: age, gender, education, employment, family composition, income, and parameters of debt burden (Table 1).

Dependent Variable:

days — a binary variable that takes the value 1 if a loan delinquency of more than 90 days is present, and 0 otherwise;

Key Explanatory Variables:

Numlo — the number of active loans, reflecting the degree of debt burden;

IncPer — per capita household income, calculated as total income (Inc) divided by the number of household members (Perhouse), measured in tenge;

City — type of settlement (urban/rural);

Age — age of the respondent;

Gen — gender of the respondent (male/female);

Edu — level of education (primary, secondary, higher);

Sst — social status (employed, retired, student, etc.);

Monpay — total monthly payment on all loans (in tenge).

Table 1 — Description of Variables Used in the Study

№	Variable	Symbol	Description and Measurement
1	Place of residence of the respondent	City	urban / rural
2	Age	Age	Age of the respondent, years
3	Gender	Gen	Gender of the respondent: male/female
4	Education	Edu	Level of education of the respondent (primary, secondary, higher, etc.)
5	Social Status	Sst	Social status of the respondent (e.g., employed, retired, student, etc.)
6	Large Family	Lfam	Large family (yes/no)
7	How many people live in the household, including the respondent?	Perhouse	Number of people living together in one household
8	Do you have any loans or debts?	Avlo	Does your household currently have any loans or debts? (yes/no)
9	Number of loans	Numlo	Number of active loans in the household
10	Monthly loan payment	Monpay	Amount of total monthly payment on all loans, tenge
11	Are there any loans delinquent for more than 90 days?	days	Presence of overdue loan payments of more than 90 days (yes = 1, no = 0)
12	Total monthly household income	Inc	Total monthly household income, tenge

Note — compiled by the authors based on [23]

The logit model has the following form:

$$\Pr(\text{days}_i = 1) = \frac{e^{\beta_0+\beta_1\cdot\text{Numlo}_i+\beta_2\cdot\text{IncPer}_i+\beta_3\cdot\text{Age}_i+\beta_4\cdot\text{X}_i}}{1+e^{\beta_0+\beta_1\cdot\text{Numlo}_i+\beta_2\cdot\text{IncPer}_i+\beta_3\cdot\text{Age}_i+\beta_4\cdot\text{X}_i}} \tag{1}$$

where:
$\Pr(\text{days}_i = 1)$ — probability that household i has a loan delinquency,
Numlo_i — number of active loans,
IncPer_i — per capita household income,
X_i — vector of socio-demographic variables (gender, education, status, place of residence),
β — estimated model coefficients.

$$\Pr(\text{days}_i = 1) = \Phi(\beta_0+\beta_1\cdot\text{Numlo}_i+\beta_2\cdot\text{IncPer}_i+\beta_3\cdot\text{Age}_i+\beta_4\cdot\text{X}_i) \tag{2}$$

where Φ denotes the standard normal distribution function.

H1: An increase in the number of active loans is associated with a higher probability of delinquency ($\beta_1 > 0$).
H2: Low per capita household income increases the risk of default ($\beta_2 < 0$).

The application of logit and probit models is determined by the binary nature of the dependent variable and is supported by academic practice in credit risk assessment [24]. Both models allow for capturing the probabilistic nature of financial behavior, in contrast to linear regression, which is inadequate for modeling binary outcomes [22].

The proposed empirical model differs from those previously discussed in the literature both in methodological approach and in the context of application. The main distinctions and novelty can be summarized as follows:

1. Regional context and specifics of digital lending.

In contrast to most foreign studies focused on stable financial systems [10,11], the present research relies on data from Kazakhstan—a country with a rapidly developing microcredit market and a relatively high share of vulnerable households. A distinctive feature of the Kazakhstani context is the widespread practice of simultaneously obtaining multiple loans through online applications, often without proper assessment of the borrower's solvency. This circumstance necessitates a specific model that takes into account both the number of loans and demographic characteristics.

2. Per capita income (IncPer)

Unlike most studies that use total household income as an explanatory variable [21], the present research employs the derived variable IncPer—per capita household income. It was constructed from survey data by dividing the total monthly household income (Inc) by the number of household members (Perhouse). Such normalization provides a more accurate measure of financial vulnerability, reflecting the resources available per individual within the household. This is particularly relevant for large families and extended households, which are common in rural areas of Kazakhstan. This approach offers a more sensitive assessment of repayment capacity than using gross household income.

3. Extended use of socio-demographic characteristics.

In contrast to models focused on behavioral indicators of financial vulnerability and their relationship with credit availability and debt burden [14,16], the present study analyzes structural socio-demographic variables: age, gender, education, social status, and household composition. Examining variables such as age, gender, level of education, and family structure makes it possible to better understand hidden factors of social vulnerability that are not directly observable but significantly affect the risk of delinquency. This enhances the accuracy of models under conditions of diverse borrower socio-economic backgrounds.

4. Methodological robustness: logit and probit modeling.

To enhance the reliability of the results, both logit and probit estimations were applied. This dual approach makes it possible to test the robustness of the findings with respect to the specification of the error distribution [22,25]. Both types of models produced consistent results, which strengthens the credibility of the empirical conclusions.

1. Kazakhstani context: original survey microdata are used, reflecting local practices of digital lending, including multiple borrowing and microfinance.

2. Per capita income (IncPer): unlike the traditional approach of using total household income, a more accurate indicator of repayment capacity is proposed, one that is adapted to large families.

3. Integration of socio-demographic variables: the model incorporates factors not typically included in most banking data—age, gender, education, and status—allowing for a more comprehensive assessment of default determinants.

Results and conclusions

The results of the logit and probit regressions are presented in Table 2 and Table 3. The main objective of the analysis is the empirical testing of hypotheses H1 and H2 regarding the impact of the number of active loans and per capita income on the probability of loan delinquencies. The estimations were carried out using robust standard errors.

H1: An increase in the number of loans raises the risk of default ($\beta_1 > 0$);

H2: Low per capita income increases the risk of default ($\beta_2 < 0$).

Table 2 — Results of logit and probit regressions for estimating the probability of delinquency

	Characteristic	Logit	Probit
1	2	3	4
2	Number of observations	252	252
3	Wald chi2(9)	42.21	46.25
4	Prob > chi2	0.0000 (model statistically significant)	0.0000 (model statistically significant)
5	Pseudo R2	0.1964	0.1977
6	Numlo (coefficient)	0.541216 (statistically significant, p = 0.005)	0.3071444 (statistically significant, p = 0.002)
7	Other variables	Age, IncPer, City, Gen, Edu, Sst: not statistically significant (p > 0.05)	Age, IncPer, City, Gen, Edu, Sst: not statistically significant (p > 0.05)
8	Constant	-2.604637 (statistically significant, p = 0.010)	-1.56204 (statistically significant, p = 0.001)
Note — compiled by the authors based on [23]			

In both models, the variable Numlo is a significant predictor with a positive effect on the probability of default, fully confirming hypothesis H1.

Table 3 — Results of Logit and Probit Regressions for Hypothesis Testing

	Variable	Logit: Coef. (p-value)	Probit: Coef. (p-value)	Hypothesis outcome
1	2	3	4	5
2	Numlo (number of loans)	0.541 (0.005)	0.307 (0.002)	H1 confirmed
3	IncPer (per capita income)	–0.521 (0.455)	–0.305 (0.342)	H2 not confirmed
4	Age	0.181 (0.563)	0.116 (0.469)	Not significant
5	City	–0.688 (0.482)	–0.301 (0.505)	Not significant
6	Gen	–1.252 (0.300)	–0.476 (0.319)	Not significant
Note — compiled by the authors based on [23]				

In both models, the variable Numlo is a significant predictor with a positive effect on the probability of default, fully confirming hypothesis H1.

The variable IncPer shows a negative sign in both models, which is consistent with hypothesis H2; however, in both cases the coefficients are statistically insignificant.

The results of the logistic and probit regressions showed that the number of open credit lines (Numlo) has a statistically significant positive effect on the probability of loan delinquency. This confirms the hypothesis that borrowers with a larger number of open credit lines are more prone to delinquencies. However, other socio-demographic variables, such as age, per capita income, place of residence, gender, education, and social status, did not show a statistically significant impact on the probability of delinquency. Both models (logistic and probit) demonstrated overall statistical significance ($p < 0.0001$), but relatively low pseudo R^2

values (around 0.19), indicating that other important factors influencing the probability of delinquency are not captured in the model.

Table 4 — Marginal Effects (margins, dy/dx)

	Variable	Logit: dy/dx (p)	Probit: dy/dx (p)
1	2	3	4
2	Numlo	0.0192 (0.036)	0.0256 (0.014)
3	IncPer	–0.0185 (0.476)	–0.0255 (0.387)
Note — compiled by the authors based on [23]			

The effect of the variable Numlo on the probability of default is significant in both the logit and probit models: an additional loan increases the risk of delinquency by approximately 2–2.5 %.

The effect of income (IncPer) is negative but statistically insignificant.

The analysis of marginal effects for the logit model showed that the number of open credit lines (Numlo) has a statistically significant positive impact on the probability of loan delinquency. An increase of one active loan raises the probability of delinquency by an average of 1.92 percentage points ($p = 0.036$), all else being equal.

A similar result was obtained in the probit model: the variable Numlo also demonstrated a positive and statistically significant effect. An additional credit line increases the probability of default by an average of 2.56 percentage points ($p = 0.014$) (Table 4).

Other socio-demographic characteristics—age, per capita income, type of settlement, gender, level of education, and social status—did not show a statistically significant effect on the probability of delinquency in either model ($p > 0.1$).

To estimate the marginal effects, mean values of the predictors were used, along with robust standard errors, which account for potential heteroskedasticity and enhance the reliability of interpretation.

Table 5 — Classification Quality

	Metric	Logit Model	Probit Model
1	2	3	4
2	Classification accuracy	93.65 %	93.65 %
3	Sensitivity	11.76 %	11.76 %
4	Specificity	99.57 %	99.57 %
5	PPV (precision)	66.67 %	66.67 %
6	NPV	93.98 %	93.98 %
Note — compiled by the authors based on [23]			

Both models—the logistic and probit regressions—demonstrated high overall classification accuracy (93.65 %), indicating their ability to correctly identify the majority of borrowers. However, a more detailed analysis of performance metrics revealed a substantial imbalance between sensitivity and specificity.

In particular, the sensitivity of both models was only 11.76 %, indicating an extremely low ability to detect borrowers with an actual risk of default. In other words, the models correctly classify only about one out of nine borrowers who became delinquent. This critically limits their practical value for credit institutions, as a significant number of high-risk clients remain unidentified.

At the same time, the specificity of the models turned out to be very high—99.57 %, indicating a high accuracy in recognizing reliable payers. Such asymmetry in model performance points to a bias toward the majority class (clients without delinquencies), which may be a consequence of strong class imbalance in the sample.

Low sensitivity is a key limitation: it increases the likelihood of granting loans to borrowers with a high risk of delinquency, which, in turn, may lead to substantial financial losses.

The estat classification command showed:

Overall classification accuracy: 93.65 %

Sensitivity (true positive rate): 11.76 %

Specificity (true negative rate): 99.57 %

This means that the model predicts the absence of default well, but performs poorly in predicting its occurrence, which is typical for imbalanced samples with a low share of defaults (Table 5).

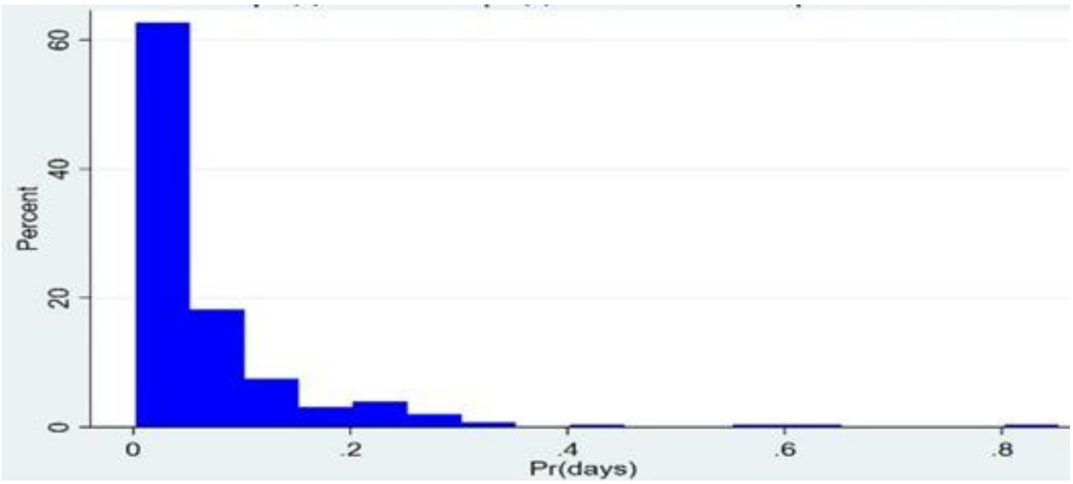

Figure 2 — Distribution of Predicted Probabilities of Default

Note — visualization and analytical processing of data were carried out using artificial intelligence tools based on [23].

The distribution of predicted default probabilities in Figure 2 shows that most households are concentrated in the low-probability zone (<10 %). However, there is a clearly defined “risk group” with probabilities above 30 %. This indicates the existence of vulnerable households characterized by a high debt burden and weak financial resilience.

5. Goodness-of-Fit (Pearson Test)
Model χ^2 (df = 152), Logit: 241.08
Model χ^2 (df = 152), Probit: 214.37
p-value Logit = 0.0000
p-value Probit = 0.0006

The probit model shows a better fit to the data (smaller deviation), although both models formally fail the goodness-of-fit test—an indication of possible model misspecification.

6. Multicollinearity (VIF)
Logit model: Mean VIF = 5.31
Probit model: Mean VIF = 2.87

Both models show no signs of serious multicollinearity (VIF < 10). The probit model demonstrates a more stable predictor structure.

The results of the analysis confirmed the significance of the variable “number of active loans” in predicting the risk of delinquency, thereby supporting hypothesis H1. An increase in the number of active loans statistically significantly raises the probability of default in both the logit and probit models, with similar coefficients and marginal effects.

At the same time, hypothesis H2, which assumed a negative effect of income level on the risk of delinquency, did not receive statistical support. Although the coefficients had the expected sign, they did not reach statistical significance, which may indicate the influence of unobserved factors—such as income instability, type of credit product, or borrowers’ behavioral characteristics.

Despite high overall classification accuracy (93.65 %), both models demonstrated extremely low sensitivity (11.76 %), which substantially limits their practical applicability. The probit model showed slightly better robustness according to the goodness-of-fit criterion (χ^2), as well as lower multicollinearity (VIF) and a somewhat stronger marginal effect for the key variable Numlo.

In addition, the Pearson test revealed a statistically significant deviation of both models from the observed data, which may be associated with omitted predictors, unaccounted nonlinear relationships, or specification errors.

Recommendations for model improvement:

Use class balancing methods (SMOTE, weighted loss functions);

Include additional behavioral and macroeconomic variables;

Lower the classification threshold;

Apply modern machine learning algorithms (XGBoost, Random Forest).

Conclusions

The conducted study made it possible to empirically confirm a significant relationship between the level of household indebtedness and the risk of delinquency, which is consistent with the findings of a number of foreign studies [26]. It was established that each additional loan obligation increases the probability of default, making this indicator a key factor in the construction of scoring models.

At the same time, per capita household income did not demonstrate statistical significance, despite the negative direction of its effect. This may be explained by the fact that formal income does not reflect the actual repayment capacity of borrowers, particularly under conditions of digital and parallel borrowing. These findings are consistent with Madeira [15], who emphasizes that income instability and unpredictability are more important than its nominal level.

The analysis also revealed the influence of social status and credit accessibility, highlighting the need to account for behavioral and institutional factors in credit risk assessment models. The obtained results demonstrate the potential of logistic models when relevant variables are included; however, they also indicate the limited predictive power of such models without addressing class imbalance and behavioral aspects.

Practical Recommendations

- Development of scoring tools that are sensitive to the number of active loans, taking into account the structure of current debt rather than relying solely on formal repayment capacity.
- Digitalization of credit data through a unified platform that allows for tracking parallel loan applications submitted by borrowers across different MFIs and banks.
- Creation of early warning systems for borrowers with multiple loans, including referral to debt management programs.
- Expansion of financial literacy programs, with a focus on socially vulnerable groups (pensioners, students, the unemployed).
- Introduction of restrictions on multiple borrowing for clients with a high debt burden, in order to reduce systemic risk.

Directions for Further Research

Expansion of the sample and use of panel data to analyze the dynamics of credit behavior.

Application of machine learning techniques (Random Forest, XGBoost) to identify non-trivial patterns of default behavior.

Inclusion of psychological and behavioral factors, including analysis of borrowers' digital activity (frequency of online loans, behavior on lenders' websites).

Evaluation of the effectiveness of implemented recommendations based on experimental or quasi-experimental data.

References

THE 17 GOALS (2025). Sustainable Development Goals. *sdgs.un.org/goals.* Retrieved from https://sdgs.un.org/goals.

World Bank (2025). Doklad Vsemirnogo banka o sotsialno-ekonomicheskom blagopoluchii za dekabr 2024 goda [World Bank Report on Socio-Economic Well-being for December 2024]. *thedocs.worldbank.org.* Retrieved from https://thedocs.worldbank.org/en/doc/802d52e685aec7e83bc350a84e4b2dba-0080062025/original/L2Kaz-Brief-December-2024-ru.pdf [in Russian].

Agarwal, S., & Chua, Y.H. (2020). FinTech and household finance: A review of the empirical literature. *China Finance Review International, 10,* 4, 361–376. DOI: 10.1108/CFRI-03-2020-0024.

Białowolski, P., Kowalski, R., Węga, A., & Węga, G. (2024). To explore the divergence between subjective and objective debt burden measures and establish an optimal DSTI threshold. *International Journal of Consumer Studies, 48,* 3, 245–263. DOI: 10.1111/ijcs.12890.

Xiao, J.J., & Yao, R. (2020). How family structure affects debt ownership and burden. *Journal of Family and Economic Issues, 41,* 1, 87–104.

Dempere, J., & Malik, S. (2021). Explanatory power of demographics, behavior, and literacy on fraud victimization. *International Journal of Consumer Studies, 45,* 2, 165–182. DOI: 10.1111/ijcs.12698.

Fernández-López, S., Álvarez-Espinosa, M., Castro-González, S., & Rey-Ares, L. (2022). Financial capability and financial vulnerability. *International Journal of Bank Marketing, 40,* 4, 755–773. DOI: 10.1108/IJBM-04-2021-0406.

Fernández-López, S., Daoudi, J., & Rey-Ares, L. (2022). Social interactions and borrowing decisions among Europeans 50+. *International Journal of Bank Marketing, 40,* 2, 135–152. DOI: 10.1108/IJBM-02-2021-0270.

Chen, F., Jiang, G., & Gu, M. (2024). Financial knowledge and credit card behavior. *International Journal of Bank Marketing, 42,* 5, 779–795. DOI: 10.1108/IJBM-05-2022-0245.

Xidonas, P., et al. (2025). Household credit behavior and constraints in France. *Journal of Economic Studies, 53,* 2, 110–128. DOI: 10.1108/JES-07-2023-0387.

Mutsonziwa, K., & Fant, A. (2019). Drivers of over-indebtedness and poverty in Southern Africa. *International Journal of Social Economics, 46,* 3, 359–378. DOI: 10.1108/IJSE-03-2018-0224.

De Oliveira Santini, F., Ladeira, W.J., Mette, F.M.B., & Ponchio, M.C. (2019). Antecedents and consequences of financial literacy: Meta-analysis. *International Journal of Bank Marketing, 37,* 6, 1273–1290. DOI: 10.1108/IJBM-11-2018-0330.

Diba, F., Abrantes-Braga, M.A., & Veludo-de-Oliveira, T. (2019). Behavioral dimensions of financial well-being. *Journal of Consumer Affairs*, *53,* 2, 238–257. DOI: 10.1111/joca.12181.

Madeira, C. (2017). Cyclical volatility of consumer debt risk in Chile. *Journal of Financial Stability, 28,* 1–10. DOI: 10.1016/j.jfs.2017.03.003.

Mukhan, M.M., Malik, G., & Mukametkaliyeva, M.A. (2023). The impact of financial literacy on microloan volumes among Kazakh students. *Central Asian Economic Review,* 2, 129–140. DOI: 10.52821/2789-4401-2023-2-129-140

Ishuova, Zh., Daribaeva, M., & Boluspaev, Sh. (2024). Financial stability and consumption adjustment under employment shocks. *Journal of Economic Research and Business Administration, 1*(147), 68–80. DOI: 10.26577/jerba202414716 [in Russian].

Ishuova, Zh.Sh., Daribaeva, M.Zh., & Boluspaev, Sh.A. (2023). Consumer spending smoothing and credit demand in Kazakhstan. *Vestnik Evraziiskoi Nauki — Bulletin of Eurasian Science,* 15, 4. DOI: 10.15862/50ECVN423 [in Russian].

National Bank of Kazakhstan (2024). Otchet o finansovoi stabilnosti Kazakhstana 2023 [Report on Financial Stability of Kazakhstan 2023]. *nationalbank.kz.* Retrieved from https://www.nationalbank.kz [in Russian].

National Bank of Kazakhstan (2025). Statisticheskii biulleten No. 3, 2025 [Statistical Bulletin No. 3, 2025]. *nationalbank.kz.* Retrieved from https://www.nationalbank.kz [in Russian].

Das, T. (2019). Credit access and poverty reduction in rural Assam. *International Journal of Social Economics, 46,* 10, 1113–1131. DOI: 10.1108/IJSE-11-2018-0273.

Gujarati, D.N., & Porter, D.C. (2009). *Basic Econometrics* (5th ed.). New York: McGraw-Hill Education.

Ismawati, I.Y., & Faturrohman, T. (2023). *To develop a credit scoring model using borrower characteristics.* Comparative Economic Studies on Trade and Finance in Emerging Economies.

Dougherty, C. (2011). *Introduction to Econometrics* (4th ed.). Pearson Education.

Fuinhas, J.A., Silva, N., & Duarte, J. (2023). Debt contagion among credit types. *Studies in Economics and Finance, 40,* 3, 366–386. DOI: 10.1108/SEF-11-2020-0514.

https://doi.org/10.31489/2025Ec4/1204

JEL: C53, I21, J11
ORIGINAL RESEARCH

Received: 30.06.2025 | Accepted: 07.10.2025

Career Self-Efficacy of School and University Students: A Systematic Review of Individual and Contextual Antecedents (1995–2025)

Assylbek Nurgabdeshov[1] *, **Amina Gassanova**[2]

Abstract

This article presents a systematic review of research on the antecedents of career self-efficacy among school and university students. Drawing on Social Cognitive Career Theory (SCCT), the review focuses on how individual characteristics and contextual conditions jointly shape students' beliefs in their ability to explore options, make decisions and pursue preferred career pathways. A PRISMA-guided search of Scopus, Web of Science, Emerald and EBSCO identified 48 empirical studies published between 1995 and 2025 that examined career self-efficacy or closely related constructs among secondary and higher education students. Data were extracted using a structured coding template and synthesized thematically.

The findings show that individual antecedents encompass both relatively stable traits (e.g., curiosity, persistence, openness to experience, emotional stability) and malleable psychological resources (e.g., emotional regulation, self-esteem, career adaptability, perceived person–environment fit). Contextual antecedents include social support from parents, teachers, counsellors and peers; structured career interventions such as courses, workshops and experiential programmes; and wider socio-demographic and structural factors related to social class, gender, ethnicity and migration background. Across studies, career self-efficacy is unevenly distributed, reflecting broader patterns of opportunity and constraint.

The review reinforces and refines SCCT by demonstrating that contextual supports and barriers operate as proximal determinants of self-efficacy alongside individual resources, and by highlighting psychological resources as mechanisms linking structure and agency. It identifies major gaps in the literature, including limited longitudinal and intersectional research, narrow geographical coverage and short-term evaluations of interventions. The article concludes by outlining implications for theory and practice and by calling for multi-level strategies that combine individual-focused support with efforts to address structural inequalities in education and career guidance.

Keywords: Career self-efficacy, Secondary Education, Career decision-making, Social Cognitive Career Theory, Higher Education

Note: The authors used AI tools for grammar correction, language refinement, and stylistic adjustments to enhance the clarity and coherence of the manuscript.

Introduction

Career self-efficacy has become a central construct in educational psychology and career development research because it shapes how young people approach choices about their future. Building on Bandura's notion of self-efficacy as people's beliefs in their capabilities to organise and execute courses of action (Bandura, 1993), career self-efficacy refers to students' confidence in performing tasks such as exploring options, making decisions and pursuing preferred pathways. Prior studies show that these beliefs are closely linked to academic engagement, willingness to explore non-traditional routes and persistence in the face of setbacks, as well as eventual transitions into further study, training or the labour market (Betz, 2000; Jantzer, Stalides & Rottinghaus, 2009; Moldashev et al., 2019; Umirzakov et al., 2019). For adolescents in school and emerging adults in higher education, career self-efficacy is therefore a key psychological resource that underpins how they navigate increasingly complex educational and occupational environments.

Social Cognitive Career Theory (SCCT) has provided a powerful lens for understanding how such beliefs develop. SCCT conceptualises career self-efficacy as arising from the interplay of personal characteristics, learning experiences and contextual supports and barriers (Lent, 2008). Within this framework, students' confidence in their career-related capabilities is not a fixed trait, but a dynamic system of beliefs shaped over time by feedback from family, teachers, peers, institutions and the wider social structure. Empirical studies illustrate, for example, how emotional stability and personality characteristics relate to career

[1]Edinburgh Business School, Heriot-Watt University, Edinburgh, UK, nurgabdeshov@gmail.com (corresponding author)
[2]Business School, SDU University, Kaskelen, Kazakhstan, Amina.gassanova@sdu.edu.kz

self-efficacy (Bubić & Ivanišević, 2016), how emotional intelligence supports career-related decision-making (Santos, Wang & Lewis, 2018) and how career adaptability can help young adults maintain confidence in the face of barriers (Creed, 2018; Duffy, Douglass & Autin, 2015). Together, this work points to a complex set of individual and contextual influences that operate across secondary and higher education.

However, despite broad agreement that career self-efficacy plays a decisive role in students' career development, the literature remains fragmented. Studies are scattered across disciplines such as educational psychology, counselling, vocational behaviour and higher education research; they often focus either on intra-individual variables (e.g. personality, self-concept, emotional resources) or on selected contextual influences (e.g. parental support, religious or community support, specific interventions) without integrating these strands (Lent, 2008; Jantzer et al., 2009). The evidence base is dominated by cross-sectional survey research conducted in single institutions, relies heavily on convenience samples and frequently treats socio-economic status, gender, ethnicity and migration background as background controls rather than as central objects of inquiry. As a result, we have useful but partial insights and lack the kind of synthesis needed to understand how individual and contextual antecedents work together across secondary and higher education.

This study addresses these gaps by conducting a systematic review of research on the career self-efficacy of school and university students published between 1995 and 2025. Using a PRISMA-guided approach, we identified, screened, and analysed forty-eight empirical studies that examine career self-efficacy or closely related constructs among secondary and higher education populations. Each article was coded for publication outlet, national context, educational level, methodological approach, measurement strategy, and substantive findings. The coded material was then synthesised thematically to identify recurring patterns in individual and contextual antecedents, as well as notable gaps and imbalances in the evidence base.

The review shows that students' career self-efficacy is shaped by two broad groups of antecedents. Individual factors include dispositional tendencies (such as curiosity, persistence and emotional stability) and malleable psychological resources (such as emotional regulation, self-esteem, career adaptability and perceived person–environment fit). Contextual factors encompass social support from parents, teachers, counsellors, peers and religious or community figures; structured career interventions such as courses, workshops and experiential programmes; and wider socio-demographic and structural conditions linked to class, gender, ethnicity and migration background. By bringing these strands together, the review extends existing knowledge in three ways: it consolidates a fragmented literature into an integrated framework of antecedents, refines SCCT-based understandings of how person–context transactions shape career self-beliefs in adolescence and emerging adulthood, and highlights underexplored areas, including developmental trajectories, intersectional disadvantage and the longer-term outcomes of career self-efficacy.

The remainder of the paper is structured as follows. The next section outlines the methodology of the systematic review, including search strategy, inclusion and exclusion criteria and data coding procedures. The Findings section then synthesises the reviewed studies by organising antecedents into individual and contextual domains and identifying key thematic clusters within each. The Discussion interprets these findings in light of SCCT, drawing out theoretical implications and methodological limitations and outlining directions for future research. The paper concludes by summarising the main contributions and offering practical recommendations for educators, counsellors and policymakers seeking to strengthen the career self-efficacy of school and university students.

Methodology

Data Collection

This study adopts a systematic literature review approach to ensure transparency, consistency, and rigour in synthesising prior work (Tranfield et al., 2003; Page et al., 2021). The review was guided by the PRISMA framework, which structures the process into four stages — identification, screening, eligibility, and inclusion — and allows readers to trace how the final sample of articles was derived.

The focus of the review is on studies that examine career self-efficacy or closely related constructs such as career decision self-efficacy among students in secondary and higher education. To capture the development of this research over time, the search covered the period from 1995 to 2025. This time frame spans three decades during which both SCCT and the study of self-efficacy in education have expanded significantly.

Searches were conducted in four major databases widely used in education, psychology, and social sciences research: Scopus, Web of Science, Emerald, and EBSCO. Boolean operators were applied to combine keywords and synonyms related to the construct, population, and educational level. Typical search strings included combinations such as “career self-efficacy” OR “career decision self-efficacy” AND “students” AND (“secondary school” OR “high school” OR “college” OR “university” OR “undergraduate”). The search was restricted to peer-reviewed journal articles written in English. In addition to database searches, manual checks of key journals in career development and educational psychology, as well as backward and forward citation tracking, were conducted to ensure that influential studies not captured by the initial search terms were included (Sarzhanova & Nurgabdeshov, 2025; Umirzakov et al., 2019).

Screening and eligibility

The initial search yielded a larger pool of publications, from which duplicates were removed. Titles and abstracts were then screened to exclude clearly irrelevant items. Full texts were obtained for the remaining articles and assessed against predefined inclusion and exclusion criteria.

To be included, studies had to: (i) be published in peer-reviewed journals; (ii) empirically examine career self-efficacy, career decision self-efficacy, or closely related career-specific efficacy beliefs as a focal construct; (iii) focus on student populations in secondary schools, high schools, colleges, or universities; and (iv) report primary data based on quantitative, qualitative, or mixed-method designs. Studies that focused exclusively on general self-efficacy without a clear career component, or that addressed broader aspects of career development without measuring self-efficacy, were excluded. Non-journal sources such as books, dissertations, conference papers, and reports were also excluded in order to maintain a consistent quality threshold.

Following this procedure, a final sample of forty-eight empirical studies published between 1995 and 2025 was retained for detailed analysis. These articles represent work conducted in a range of national contexts and educational levels, but all examine how individual or contextual factors shape students’ confidence in dealing with career-related tasks.

Data coding and analysis

Data from each article were extracted using a structured coding template. For every study, the following information was recorded: year of publication, journal outlet, country or region, educational level of participants (secondary school, high school, college, or university), sample characteristics, research design (quantitative, qualitative, or mixed-method), measurement instruments used to assess career self-efficacy, and key findings related to antecedents and outcomes of career self-efficacy.

On the basis of this coding, descriptive statistics were generated to characterise the evidence base in terms of publication outlets, temporal trends, and methodological approaches. Across the forty-eight studies, approximately nineteen per cent employed qualitative designs, fifty per cent used quantitative survey-based methods, twenty-seven per cent adopted mixed-method approaches, and four per cent were conceptual or theoretical contributions linked closely to empirical work. These patterns confirm that the field is dominated by cross-sectional survey research, with a smaller but growing body of qualitative and mixed-method studies.

The substantive analysis proceeded in two steps. First, findings from individual studies were grouped into broad categories according to whether they focused primarily on individual antecedents (for example, personality traits, psychological resources, academic self-concept) or contextual antecedents (for example, social support, career interventions, socio-demographic conditions). Second, within these two overarching domains, more fine-grained themes were inductively identified by comparing how different studies conceptualised and operationalised predictors of career self-efficacy. This iterative process of constant comparison allowed similar constructs to be clustered and recurring mechanisms to be identified across diverse samples and settings.

The approach adopted here follows established practices in systematic review research (Eva et al., 2019; Nguyen et al., 2025) and is designed to ensure that the synthesis is both transparent and replicable. By combining a structured search and screening strategy with systematic coding and thematic analysis, the review provides a coherent overview of what is currently known about the individual and contextual antecedents of students’ career self-efficacy, as well as the key gaps that warrant further investigation.

Table 1: List of journals and number of articles related to career self-efficacy of high school students (1995–2025)

No		No of papers
1	Career Development International	1
2	Journal of Career Assessment	8
3	Journal of Career Development	12
4	Journal of Counselling & Development	2

5	Journal of Counselling & Psychology	1
6	Journal of Hospitality, Leisure, Sport & Tourism Education	2
7	Journal of Science Education and Technology	1
8	Journal of Vocational Behavior	9
9	Leadership Quarterly	1
10	Personality and Individual Differences	3
11	Research in Higher Education	1
12	The Career Development Quarterly	8
Note — compiled by the authors		

Publication outlets

It is interesting to note that most of the papers (77 %) in our review were published in four journals, Journal of Career Assessment, Journal of Career Development, Journal of Vocational Behavior and The Career Development Quarterly (see Table 1). From the analysis of time trends, it is important to note that 66 % of the articles were published after 2009, which leads us to define this issue as a hot topic.

Research methods used in the articles

It can be seen from Table 2, 50 % of empirical studies were conducted by using quantitative research methods, quantitative research methods were preferred by 19 % of articles, while the remaining 27 % utilized a mixed-method approach and 4 % were conceptual studies.

Table 2. Research methods used by career self-efficacy

Research methods		No of papers	% of No
Qualitative	Interviews Observations Policy documents, secondary data Case study, interviews Longitudinal interviews	9	19 %
Quantitative	Survey, questionnaire Secondary data	24	50 %
Mixed	Survey and interviews Documental search from international agencies Survey, secondary data Case study	13	27 %
Conceptual papers		2	4 %
Total		48	100 %
Note — compiled by the authors			

Findings

Existing research provides extensive insights into the factors that shape students' confidence in managing career-related tasks. Across the forty-eight studies reviewed, the antecedents of career self-efficacy cluster into two broad domains: individual characteristics and psychological resources on the one hand, and contextual influences embedded in family, school, and wider social environments on the other. In line with SCCT, these studies show that students' beliefs about their ability to explore options, make decisions, and pursue preferred pathways are co-constructed through the interaction between personal dispositions and the supports and barriers they encounter in their environments.

Individual Antecedents

Empirical research demonstrates that the career self-efficacy of school and university students is strongly shaped by individual characteristics. The first group of studies highlights the role of dispositional tendencies and career-related skills. Work using the Planned Happenstance Career Inventory shows that skills such as curiosity, persistence, flexibility, optimism, and risk-taking are positively associated with career decision self-efficacy, suggesting that students who are more open to unplanned opportunities and able to act on them report greater confidence in their ability to make career decisions (Huang, 2015). Similarly, studies grounded in the Big Five model indicate that conscientiousness, extraversion, and openness to experience are positively linked to career decision self-efficacy, whereas neuroticism tends to undermine students' confidence (Hartman & Betz, 2007). Research on healthy personality profiles further suggests that adolescents and young adults who score higher on indicators of emotional and social adjustment show stronger beliefs in their capacity to handle career tasks (Borgen & Betz, 2008).

A second set of studies focuses on psychological resources that help students regulate emotions and cope with uncertainty. Emotional stability has been identified as a significant negative predictor of career concerns, indicating that students who experience lower levels of anxiety and emotional volatility are less likely to report decision-making difficulties (Bubić & Ivanišević, 2016; Gati et al., 2011). Ability-based emotional intelligence, including the capacity to understand, regulate, and use emotions, is also positively related to career decision self-efficacy (Santos, Wang & Lewis, 2018). Self-esteem emerges as another robust predictor, although some studies report gender differences, with the association between self-esteem and career self-efficacy being particularly pronounced among male students (Lease & Dahlbeck, 2009). Building on SCCT, the research shows that independent self-construal and a sense of personal agency are positively associated with confidence in career decision-making, especially when students have access to adequate occupational information and opportunities for career planning (Gianakos, 2001).

The evidence also sheds light on individual factors that may hinder the development of career self-efficacy. Symptoms of depression, social withdrawal, and low life satisfaction are negatively associated with students' beliefs in their ability to make effective career choices, particularly among those who perceive limited control over their future (Hu, Hood & Creed, 2018). Beyond psychological adjustment, career adaptability has been identified as a key resource: students who report higher levels of concern, control, curiosity, and confidence tend to feel more capable of engaging in career exploration and handling transitions, and interventions designed to enhance career adaptability have been shown to increase career self-efficacy over time (Shin, Steger, & Lee, 2014). Other studies suggest that students' achievement-related beliefs and decision-making styles also matter. High academic achievement and realistic expectations about future performance are linked to stronger career self-efficacy (Choi & Kim, 2013), whereas a maximising orientation — striving to identify the single best option — can lead to rumination and indecision, partly through its effects on outcome expectations (Conklin, Dahling, & Garcia, 2013). Research on younger adolescents shows that early clarity about preferred career fields and realistic perceptions of occupational requirements are associated with higher confidence in navigating educational choices (Jantzer, Stalides & Rottinghaus, 2009).

Finally, several studies examine how person-environment fit shapes students' efficacy beliefs. Findings indicate that congruence between students' interests and their chosen programmes, as well as perceived alignment between personal values and anticipated occupational roles, are positively related to career decision self-efficacy and satisfaction with educational choices (Srsic & Walsh, 2001; Tziner, Oren, & Caduri, 2014). When students feel that their academic and vocational environments reflect their core identities and aspirations, they are more likely to believe that they can successfully pursue and sustain their desired career paths. Taken together, the reviewed studies suggest that individual antecedents of career self-efficacy span stable traits, malleable psychological resources, and dynamic perceptions of fit and capability. This body of evidence underscores that strengthening students' emotional regulation, self-beliefs, and sense of person–environment fit can meaningfully enhance their confidence in navigating career decisions.

Contextual antecedents

In addition to individual traits and psychological resources, the development of students' career self-efficacy is strongly influenced by the environments in which they grow up and study. The reviewed studies consistently show that family, school, peers, religious communities, and broader structural conditions shape how students interpret their abilities, opportunities, and chances of success. In line with SCCT, these contextual influences operate through different forms of support and constraint, such as encouragement, role modelling, access to information, and exposure to barriers.

Social support

A substantial body of work highlights the central role of social support in fostering students' career self-efficacy. Studies with college and university students show that perceived support from family, friends, and significant others is positively associated with confidence in making career decisions and pursuing occupational goals. Students who feel that important people in their lives take their aspirations seriously, offer encouragement, and provide instrumental help (for example, sharing information about courses, jobs, or internships) tend to report higher levels of career decision self-efficacy and lower levels of indecision (Chung-Ju Huan, 2016). Moreover, it was noted that contextual antecedents such as academic support from teachers, usage of college academic resources improve students' career self-efficacy and allow them to search for different career opportunities (Di Fabio & Kenny, 2015).

Parental support emerges as particularly influential during adolescence. Research indicates that adolescents who perceive their parents as warm, involved, and encouraging of exploration are more likely to believe that they can successfully manage educational and career choices (Renn, Steinbauer, Taylor & Detwiler, 2014). Parents who expose their children to different occupations, discuss educational pathways, and model persistence in the face of difficulties help create a climate in which career self-efficacy can develop (Wright et al., 2014). At the same time, overly controlling or critical parental behaviour, or the absence of guidance, can undermine young people's confidence and lead to uncertainty about future plans (Lapan, 2002).

Supportive school environments also play an important role. Perceived teacher support, positive relationships with school counsellors, and constructive interactions with peers are associated with higher career self-efficacy. Teachers who provide feedback on strengths, link classroom learning to real-world work contexts, and encourage students to consider different possibilities help students see themselves as capable of succeeding in further study or work (Lease and Dahlbeck, 2009). Counsellors who actively listen, normalise uncertainty, and provide structured guidance similarly contribute to stronger self-efficacy beliefs (Poux & Fry, 2015). Beyond family and school, some studies emphasise the role of religious and community support (Ginevra, Nota, & Ferrari, 2015). For certain groups of students, feeling supported by faith communities and perceiving a sense of purpose or calling can reinforce confidence in career-related decisions, especially when they face external barriers or discrimination (Duffy & Lent, 2008).

Overall, this evidence suggests that social support functions as a key contextual resource. It not only supplies information but also validates students' aspirations, provides emotional reassurance, and offers concrete opportunities to practice decision-making. When such support is absent or inconsistent, students are more likely to doubt their capabilities and to experience career-related anxiety.

Career Interventions

A second line of research examines how structured career interventions influence students' self-efficacy. The reviewed studies include a range of programmes, such as career decision-making courses, workshops, group counselling, virtual experiences, and intensive camps (Glessner, Rockinson-Szapkiw, & Lopez, 2017). Most interventions adopt pre-test–post-test or quasi-experimental designs and report positive effects on career self-efficacy indicators (Van Raalte et al., 2017).

Career courses that explicitly teach decision-making skills, provide information about educational and occupational options, and offer opportunities for guided reflection consistently lead to increases in career decision self-efficacy and reductions in career indecision (Grier-Reed & Skaar, 2010). For example, students who participate in semester-long courses or short modules that combine lectures, exercises, and individual counselling typically report greater confidence in their ability to identify goals, evaluate alternatives, and implement plans than those in control groups. Interventions that incorporate experiential components — such as role plays, group discussions, or structured exposure to professionals — appear particularly effective, as they provide mastery experiences and vicarious learning opportunities emphasised by SCCT (Grier-Reed, Skaar & Conkel-Ziebell, 2009).

Shorter interventions, including workshops and virtual programmes, can also produce meaningful gains. Studies of online or technology-assisted interventions suggest that interactive exercises, personalised feedback, and scenario-based role modelling can enhance students' confidence, especially when combined with opportunities to reflect on personal strengths and values (Cardoso, Janeiro, & Duarte, 2018). Career camps and intensive programmes, often targeting specific groups of students, show improvements in clarity of aspirations and increased willingness to explore career options (Speight et al., 1995; Reese & Miller, 2006; Fouad, Cotter, & Kantamneni, 2009). However, some studies also note that the durability of these effects over time is not always assessed, and that follow-up measurements are needed to determine whether gains in self-efficacy are sustained (Komarraju, Swanson, & Nadler, 2014; Tansley et al., 2007).

Taken together, the literature indicates that well-designed interventions can strengthen students' career self-efficacy by providing structured environments in which they can practise decision-making, receive feedback, and see others successfully negotiate similar challenges. At the same time, the evidence base remains dominated by short-term evaluations and relatively homogeneous samples, leaving open questions about how to tailor interventions for diverse cultural and socio-economic contexts.

Socio-demographic factors

The final group of contextual antecedents concerns socio-demographic and structural conditions that shape students' career development. Studies focusing on social class, ethnicity, gender, and migration background show that career self-efficacy is not distributed evenly across student populations. Instead, it reflects broader patterns of opportunity and constraint (Jackson, Potere, & Brobst, 2006; Ojeda et al., 2012; Nadermann & Eissenstat, 2018).

Research on ethnicity and acculturation, for instance, finds that students from minority backgrounds may face additional barriers that can weaken their self-efficacy, such as discrimination, limited access to role models in desired occupations, or conflicting expectations between family and host cultures (Bounds, 2017). At the same time, a strong sense of ethnic identity and positive cultural socialisation can act as protective factors, supporting resilience and promoting confidence in career decision-making (Lewis et al., 2017). Studies of acculturation processes suggest that as students become more familiar with the educational and labour market systems of the host country, their career self-efficacy tends to increase, although this relationship is shaped by the quality of support they receive (Nauta & Kahn, 2007; Metz, Fouad, & Ihle-Helledy, 2009).

Socioeconomic status (SES) and parental education also influence students' beliefs about their career capabilities (Harlow & Bowman, 2016). Findings indicate that students from higher-SES families, or whose parents have completed higher education, generally report higher career decision self-efficacy, likely because they have greater access to information, networks, and financial resources (Huang & Hsieh, 2011). In contrast, students from lower-SES backgrounds may perceive more constraints and fewer viable options, which can dampen their confidence even when they have comparable abilities and aspirations (Metheny & McWhirter, 2013). Some studies show that targeted interventions and strong school support can partially compensate for these disadvantages, but they do not fully eliminate structural gaps (Suh & Flores, 2017).

Gender differences are another recurrent theme. In several studies, female students report lower confidence in their ability to pursue certain career paths, particularly in fields that are traditionally male-dominated, despite having similar or higher academic performance. These patterns are linked to gendered expectations, stereotypes about occupations, and differential encouragement from teachers and parents (Scheye & Gilroy, 1994; Metz, Fouad, & Ihle-Helledy, 2009). Conversely, in some contexts male students report lower self-efficacy for pursuing careers that are perceived as less prestigious or less aligned with dominant masculine norms (Albaugh & Nauta, 2005).

In summary, the reviewed studies highlight that contextual antecedents of career self-efficacy extend beyond immediate social support and formal interventions to include wider socio-demographic and structural factors. Family resources, community and cultural contexts, and gendered and ethnic inequalities all shape how students interpret their chances of success and the kinds of careers they consider attainable. Career self-efficacy is therefore best understood as a product of ongoing transactions between individuals and their environments, in which personal dispositions interact with the supports and barriers embedded in family, school, and society.

Discussion

This review set out to synthesise three decades of research on the antecedents of career self-efficacy among school and university students. Drawing on forty-eight empirical studies published between 1995 and 2025, it examined how individual characteristics and contextual conditions shape students' confidence in managing career-related tasks, such as exploring options, making decisions, and pursuing preferred pathways. Guided by SCCT, the analysis shows that career self-efficacy is not a fixed personal trait but a dynamic belief system that develops at the intersection of personal dispositions, psychological resources, and the social and structural environments in which young people are embedded.

The findings highlight two broad domains of antecedents. On the individual side, traits such as curiosity, persistence, openness to experience, and emotional stability are consistently associated with higher levels of career self-efficacy, whereas tendencies towards anxiety, low self-esteem, and pessimistic expectations undermine students' confidence. Psychological resources, including emotional intelligence, self-esteem, and career adaptability, emerge as particularly important because they are malleable and can be strengthened through guidance and intervention. Perceptions of person-environment fit also play a critical role: students who experience congruence between their interests, values, and chosen fields of study report stronger beliefs in their ability to cope with career tasks and transitions. Together, these findings suggest that individual antecedents are best understood as a combination of relatively stable dispositions and more dynamic self-beliefs that can be shaped by experience.

On the contextual side, the review demonstrates that social support, structured career interventions, and wider socio-demographic and structural conditions substantially influence students' career self-efficacy. Supportive relationships with parents, teachers, counsellors, and peers provide encouragement, information, and role models that help students interpret their abilities and opportunities in a more positive way. Well-designed career education and counselling programmes, including courses, workshops, and experiential activities, tend to produce measurable increases in self-efficacy, especially when they offer mastery experiences and opportunities for vicarious learning. At the same time, socio-economic disadvantage, limited parental education, discrimination, and gendered or ethnic stereotypes constrain students' perceived options and can dampen confidence, even when academic ability is high. These patterns underscore that career self-efficacy is unevenly distributed and reflects broader structures of opportunity and inequality.

Taken together, the evidence reviewed here reinforces a central insight of SCCT: students' career beliefs are co-constructed through ongoing transactions between person and context. However, the findings also extend SCCT in several ways. First, they show that contextual supports and barriers are not simply background conditions but can be as proximal and powerful as individual traits in shaping self-efficacy, particularly for students from marginalised groups. Family resources, school climate, and community and cultural supports emerge as active ingredients in the development of career self-efficacy rather than distant influences. Second, the review highlights the importance of psychological resources such as career adaptability and emotional regulation as mediating mechanisms between structural conditions and self-efficacy. These resources help explain why some students maintain confidence despite adversity, while others with similar abilities experience persistent doubt.

The review also refines our understanding of developmental timing. Studies with younger adolescents suggest that early experiences of success or failure, initial clarity of interests, and parental encouragement during key educational transitions lay the groundwork for later self-efficacy beliefs. Research with university students shows that these beliefs continue to evolve as students encounter new demands, such as selecting majors, pursuing internships, or transitioning into the labour market. This developmental perspective suggests that career self-efficacy should be conceptualised as a trajectory rather than a static outcome, with different antecedents becoming more or less salient at different stages.

In terms of the broader literature, this review contributes by bringing together fragmented strands of research on individual and contextual antecedents of career self-efficacy across secondary and higher education. Existing studies often focus either on intra-individual variables, such as personality and self-concept, or on specific forms of support or intervention, such as parental encouragement or career courses. By integrating these strands within a single framework, the review shows that neither perspective is sufficient on its own. High levels of personal agency and adaptability cannot fully compensate for unsupportive schools or severe structural constraints, just as favourable environments may not translate into confident career decision-making if students lack basic psychological resources. A key implication is that theory and practice need to attend more explicitly to the interplay between individual and contextual factors, rather than treating them as separate domains.

Finally, the review identifies several imbalances in the evidence base that shape its conclusions. The field remains dominated by cross-sectional survey studies with relatively homogeneous samples, often drawn from single institutions and middle-class backgrounds. Many interventions are evaluated only in the short term, and there is limited attention to how career self-efficacy develops among students who face intersecting disadvantages related to class, gender, ethnicity, disability, or migration status. Addressing these gaps is essential if future research is to capture the full complexity of how young people construct and sustain beliefs about their career capabilities.

Limitations and directions for future research

Like any systematic review, this study has several limitations that should be acknowledged when interpreting its findings. These limitations also point to promising directions for future research on career self-efficacy among school and university students.

The first limitation concerns the scope of the evidence base. The review focuses on peer-reviewed journal articles published in English between 1995 and 2025. While this ensures a consistent quality threshold, it also introduces potential publication and language bias. Studies published in other languages, as well as dissertations, reports, and conference papers, were not included and may contain additional evidence, particularly from non-Anglophone and low- and middle-income countries. Future reviews could broaden the range of sources and incorporate systematic searches in other languages to capture a more global picture of how career self-efficacy develops in diverse educational systems.

Second, the available research is heavily skewed towards cross-sectional, survey-based designs conducted within single institutions. This limits the ability to draw causal inferences or to understand how career self-efficacy evolves over time. Few studies follow students across key educational transitions, such as moving from lower to upper secondary education, from school to university, or from higher education into the labour market. Longitudinal and multi-wave designs are needed to trace trajectories of career self-efficacy, identify critical periods of change, and examine how early experiences of support or constraint reverberate across later stages. Such designs would also allow researchers to test more complex SCCT-based models that include mediating and moderating processes.

Third, although this review highlights the importance of contextual and structural factors, the empirical literature often treats variables such as socio-economic status, gender, ethnicity, and migration background as control variables rather than as focal constructs. As a result, the field still lacks a nuanced understanding of how intersecting disadvantages shape students' opportunities to develop and sustain strong career self-beliefs. Future research should place these dimensions at the centre of analysis, using intersectional and critical perspectives to explore how social class, gender norms, racialised experiences, and migration histories interact with school and family environments to produce differentiated patterns of career self-efficacy. This would help move the field beyond generic "one-size-fits-all" models.

A related limitation concerns the limited geographical and cultural diversity of the existing studies. Much of the evidence comes from North America, Western Europe, and East Asian contexts, with relatively few studies conducted in the Global South or in countries undergoing rapid social and economic transformation. Yet educational systems, labour markets, and cultural norms around career and family responsibilities differ substantially across regions. Comparative and cross-cultural research is therefore needed to examine whether the antecedents identified in this review operate similarly across contexts, or whether different constellations of individual and contextual factors shape career self-efficacy in different settings. Such work would also help refine SCCT by testing its assumptions in a broader range of cultural and institutional environments.

Finally, the synthesis presented here is shaped by methodological choices in coding and categorising the literature. Although a structured template was used to extract data and to group antecedents into individual and contextual domains, any thematic synthesis inevitably involves a degree of interpretive judgement. Some constructs could reasonably be placed in more than one category, and alternative coding strategies might yield slightly different thematic emphases. Future reviews could build on this work by focusing on more specific subtopics — for example, the role of career adaptability, the impact of parental involvement, or the effectiveness of particular types of intervention — and by triangulating quantitative meta-analyses with qualitative meta-syntheses.

Despite these limitations, the review provides a coherent overview of what is currently known about the individual and contextual antecedents of career self-efficacy among school and university students. Addressing the gaps identified here — through more diverse samples, longitudinal and intersectional designs, rigorous evaluations of interventions, and stronger attention to structural inequalities — would significantly advance theoretical understanding and support the development of more equitable and effective practices in career guidance and education.

Conclusion

This article has synthesised three decades of research on the antecedents of career self-efficacy among school and university students. By systematically reviewing forty-eight empirical studies published between 1995 and 2025, it has shown that students' confidence in managing career-related tasks arises from the interplay between individual traits and psychological resources on the one hand and social, institutional and structural conditions on the other. The evidence confirms a core premise of SCCT: career self-beliefs are not fixed attributes but dynamic constructions that emerge through ongoing person–context transactions.

The review makes three main contributions. First, it consolidates a fragmented literature into an integrated framework of individual and contextual antecedents of career self-efficacy. Individual factors include not only dispositional tendencies such as curiosity, persistence and emotional stability, but also malleable resources such as emotional regulation, self-esteem, career adaptability and perceived person–environment fit. Contextual factors span social support from parents, teachers, counsellors and peers; the design and delivery of career education and counselling interventions; and socio-demographic and structural conditions linked to class, gender, ethnicity and migration background. Viewing these elements together underscores that strong personal agency is difficult to sustain in the absence of supportive environments, and that favourable contexts may not translate into confident decision-making if students lack basic psychological resources.

Second, the review refines SCCT-based understandings of career development for adolescent and emerging adult populations. It highlights that contextual supports and barriers often function as proximal determinants of career self-efficacy, particularly for students facing structural disadvantage, and that psychological resources such as career adaptability and emotional regulation mediate the translation of structural conditions into self-beliefs and behaviour. It also emphasises the developmental nature of career self-efficacy: early experiences of academic success or failure, transitions between educational stages and initial exposure to work and higher education all leave traces in students' subsequent confidence and aspirations.

Third, the review identifies critical gaps that limit current knowledge and provides a focused agenda for future research. Existing studies are dominated by cross-sectional survey designs conducted in a relatively narrow set of countries and often treat socio-economic status, gender, ethnicity and migration background as control variables rather than as central objects of inquiry. Addressing these limitations will require longitudinal, intersectional and cross-cultural designs, as well as more rigorous and longer-term evaluations of career interventions. Such work is essential if research is to capture how career self-efficacy develops among students who navigate complex and unequal educational and labour market systems.

Overall, the review argues that efforts to enhance students' career self-efficacy must operate on multiple levels: by strengthening individual psychological resources, by designing rich and inclusive learning and guidance environments, and by addressing structural barriers that constrain young people's opportunities. For researchers, this implies moving beyond individualised models towards more contextually and critically informed accounts of career development. For practitioners and policy-makers, it points to the importance of aligning career education, guidance services and social policy so that all students — not only those with favourable starting positions — can develop the confidence needed to imagine, plan and pursue meaningful career pathways.

Acknowledgment: This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR24993019).

References

Albaugh, L.M., & Nauta, M.M. (2005). Career decision self-efficacy, career barriers, and college women's experiences of intimate partner violence. *Journal of Career Assessment*, *13*(3), 288–306. https://doi.org/10.1177/1069072705274958

Bandura, A. (1993). Perceived Self-Efficacy in Cognitive Development and Functioning. *Educational Psychologist*. https://doi.org/10.1207/s15326985ep2802_3

Betz, N.E. (2000). Self-Efficacy Theory as a Basis for Career Assessment. *Journal of Career Assessment, 8*(3), 205–222. https://doi.org/10.1177/106907270000800301

Betz, N.E., Borgen, F.H., & Harmon, L.W. (2006). Vocational confidence and personality in the prediction of occupational group membership. *Journal of Career Assessment*, *14*(1), 36–55. https://doi.org/10.1177/1069072705282434

Borgen, F.H., & Betz, N.E. (2008). Career self-efficacy and personality: Linking career confidence and the healthy personality. *Journal of Career Assessment*, *16*(1), 22–43. https://doi.org/10.1177/1069072707305770

Bounds, P.S. (2017). Contextual Factors Related to African American Adolescent Career Development, *65*(June), 131–144. https://doi.org/10.1002/cdq.12087

Bubić, A., & Ivanišević, K. (2016). The Role of Emotional Stability and Competence in Young Adolescents' Career Judgments. *Journal of Career Development*, *43*(6), 498–511. https://doi.org/10.1177/0894845316633779

Chung-Ju Huan, C.-C.C. (2016). College Experiences and Career Barriers among Semi-Professional Student-Athletes: The Influences of Athletic Identity and Career Self-Efficacy. *Career Development International,* 1–37. https://doi.org/http://dx.doi.org/10.1108/MRR-09-2015-0216

Conklin, A.M., Dahling, J.J., & Garcia, P.A. (2013). Linking Affective Commitment, Career Self-Efficacy, and Outcome Expectations: A Test of Social Cognitive Career Theory. *Journal of Career Development*, *40*(1), 68–83. https://doi.org/10.1177/0894845311423534

Di Fabio, A., & Kenny, M.E. (2015). The contributions of Emotional Intelligence and social support for Adaptive Career Progress among Italian Youth. *Journal of Career Development*, *42*(1), 48–59. https://doi.org/10.1177/0894845314533420

Duffy, R.D., & Lent, R.W. (2008). Relation of Religious Support in College Students. *Journal of Career Assessment*, (April), 360–369. https://doi.org/10.1177/1069072708317382

Eva, N., Newman, A., Jiang, Z., & Brouwer, M. (2019). Career optimism: A systematic review and agenda for future research. *Journal of Vocational Behavior*. https://doi.org/10.1016/j.jvb.2019.02.011

Fouad, N., Cotter, E.W., & Kantamneni, N. (2009). The effectiveness of a career decision-making course. *Journal of Career Assessment*, *17*(3), 338–347. https://doi.org/10.1177/1069072708330678

Gati, I., Gadassi, R., Saka, N., Hadadi, Y., Ansenberg, N., Friedmann, R., & Asulin-Peretz, L. (2011). Emotional and personality-related aspects of career decision-making difficulties: Facets of career indecisiveness. *Journal of Career Assessment*, *19*(1), 3–20. https://doi.org/10.1177/1069072710382525

Gianakos, I. (2001). Predictors of Career. *Journal of Career Assessment*, *9*(2), 101–114.

Ginevra, M.C., Nota, L., & Ferrari, L. (2015). Parental Support in Adolescents' Career Development : Parents' and Children's Perceptions. *The Career Development Quarterly*, *63*(March), 2–16. https://doi.org/10.1002/j.2161-0045.2015.00091.x

Glessner, K., Rockinson-Szapkiw, A.J., & Lopez, M.L. (2017). "Yes, I Can": Testing an Intervention to Increase Middle School Students' College and Career Self-Efficacy. *Career Development Quarterly*, *65*(4), 315–325. https://doi.org/10.1002/cdq.12110

Grier-Reed, T.L., & Skaar, N.R. (2010). 23. An Outcome Study of Career Decision Self- Efficacy and Indecision in an Undergraduate Constructivist Career Course.pdf. *The Career Development Quarterly*.

Hartman, R.O., & Betz, N.E. (2007). The five-factor model and career self-efficacy: General and domain-specific relationships. *Journal of Career Assessment*, *15*(2), 145–161. https://doi.org/10.1177/1069072706298011

Huang, J.T. (2015). Hardiness, Perceived Employability, and Career Decision Self-Efficacy Among Taiwanese College Students. *Journal of Career Development*, *42*(4), 311–324. https://doi.org/10.1177/0894845314562960

Jackson, M.A., Potere, J.C., & Brobst, K.A. (2006). Are success learning experiences and self-efficacy beliefs associated with occupational interests and aspirations of at-risk urban youth? *Journal of Career Assessment*, *14*(3), 333–353. https://doi.org/10.1177/1069072706286489

Komarraju, M., Swanson, J., & Nadler, D. (2014). Increased Career Self-Efficacy Predicts College Students' Motivation, and Course and Major Satisfaction. *Journal of Career Assessment*, *22*(3), 420–432. https://doi.org/10.1177/1069072713498484

Koumoundourou, G.A., Kounenou, K., & Siavara, E. (2012). Core Self-Evaluations, Career Decision Self-Efficacy, and Vocational Identity Among Greek Adolescents. *Journal of Career Development*, *39*(3), 269–286. https://doi.org/10.1177/0894845310397361

Lapan, R.T. (2002). Career Self-Efficacy and Perceptions of Parent Support in Adolescent Career Development, *51*(September), 44–55.

Lease, S.H., & Dahlbeck, D.T. (2009). Parental influences, career decision-making attributions, and self-efficacy: Differences for men and women? *Journal of Career Development*. https://doi.org/10.1177/0894845309340794

Lewis, J.A., Raque-Bogdan, T.L., Lee, S., & Rao, M.A. (2018). Examining the Role of Ethnic Identity and Meaning in Life on Career Decision-Making Self-Efficacy. *Journal of Career Development*, *45*(1), 68–82. https://doi.org/10.1177/0894845317696803

Metheny, J., & McWhirter, E.H. (2013). Contributions of Social Status and Family Support to College Students' Career Decision Self-Efficacy and Outcome Expectations. *Journal of Career Assessment*, *21*(3), 378–394. https://doi.org/10.1177/1069072712475164

Metz, A.J., Fouad, N., & Ihle-Helledy, K. (2009). Career aspirations and expectations of college students: Demographic and labor market comparisons. *Journal of Career Assessment*, *17*(2), 155–171. https://doi.org/10.1177/1069072708328862

Moldashev, K., Kozhahmet, S., Yenikeeva, A., & Nurgabdeshov, A. (2019). PREDICTORS OF THE RESEARCH PRODUCTIVITY: ARE TRAININGS FOR SCIENTISTS EFFECTIVE? *Central Asian Economic Review*, (3), 122-132.

Nadermann, K., & Eissenstat, S.H.J. (2018). Career Decision Making for Korean International College Students: Acculturation and Networking. *Career Development Quarterly*, *66*(1), 49–63. https://doi.org/10.1002/cdq.12121

Nauta, M.M., & Kahn, J.H. (2007). Identity status, consistency and differentiation of interests, and career decision self-efficacy. *Journal of Career Assessment*, *15*(1), 55–65. https://doi.org/10.1177/1069072705283786

Nguyen, N.N., Tran, Y., & Nurgabdeshov, A. (2025). International entrepreneurial mobility and innovation: revisiting the phenomenon and future research directions. *Journal of Global Mobility*, *13*(3), 445-467.

Reese, R.J., & Miller, C.D. (2006). Effects of a university career development course on career decision-making self-efficacy. *Journal of Career Assessment*, *14*(2), 252–266. https://doi.org/10.1177/1069072705274985

Sarzhanova, G., & Nurgabdeshov, A. (2025). Mapping psychological well-being in education: A systematic review of key dimensions and an integrative conceptual framework. *Journal of Pedagogical Research*, *9*(3), 327-349.

Scheye, P.A., & Gilroy, F.D. (1994). 31. College Women's Career Self- Efficacy and Educational Environments.pdf. *THE CAREER DEVELOPMENT QUARTERLY*.

Shin, J.Y., Steger, M.F., & Lee, K.H. (2014). Major Incongruence and Career Development Among American and South Korean College Students. *Journal of Career Assessment*, *22*(3), 433–450. https://doi.org/10.1177/1069072713498485

Speight, J.D., Rosenthal, K.S., Jones, B.J., & Gastenveld, P.M. (1995). 29. Medcamp's Effect on Junior High School Students' Medical Career Self- Efficacy.pdf. *THE CAREER DEVELOPMENT QUARTERLY*.

Tansley, D.P., Jome, L.M., Haase, R.F., & Martens, M.P. (2007). The effects of message framing on college students' career decision making. *Journal of Career Assessment, 15*(3), 301–316. https://doi.org/10.1177/1069072707301204

Umirzakov, S., Nurgabdeshov, A., Kozhakhmet, S., & Yenikeeva, A. (2019). The Effect of Human Resource Development on Employee Knowledge Sharing Behavior. *Central Asian Economic Review*, (2), 9-21.

Umirzakov, S.Y., Nurgabdeshov, A.R., Zamanbekov, D.S., & Gassanova, A.A. (2019). Human resource management and employee turnover in kazakhstan's companies. *Vestnik universiteta "Turan" — Bulletin of the "Turan" University*, (3), 191-195 [in Russian].

Wright, S.L., Perrone-Mcgovern, K.M., Boo, J.N., & White, A.V. (2014). Influential factors in academic and career self-efficacy: Attachment, supports, and career barriers. *Journal of Counseling and Development*, *92*(1), 36-46.

https://doi.org/10.31489/2025Ec4/1190

JEL: D83
ORIGINAL RESEARCH
Received: 16.02.2025 | Accepted: 17.06.2025

AI-Powered Libraries: Comparative Insights for an Adaptive Repository

Assem Talasbek[1], Andrey Bogdanchikov[2], Meraryslan Meraliyev[3], Mukhtar Amirkumar[4]*, Meirambek Zhaparov[5]

Abstract

This paper examines how academic staff access and use e-books, and what features they expect from a future digital library platform. Based on survey results, the study reveals that most respondents rely on informal sources, while official databases are used less frequently. At the same time, there is strong demand for key functionalities such as access to a wide collection of e-books, quiz generation, personalized recommendations, and tools for tracking student performance. In addition, the paper provides a comparative review of existing platforms—Perlego, QuizBot, Bookmate, Khan Academy, and Scopus—highlighting both their strengths and limitations, particularly with respect to language support and integration. Taken together, the survey findings and platform analysis underscore the need for a unified adaptive platform that connects publishers, libraries, and users, thereby enhancing accessibility, interoperability, and effectiveness in digital learning.

Keywords: E-books; digital library platforms; adaptive learning; personalization; interoperability; accessibility

Introduction

Library automation has improved in accuracy and efficiency via the application of artificial intelligence techniques. The concept of implementing intelligent library systems to replace conventional ones started to take shape around 1990. Libraries employ intelligent systems that use data analysis to deliver knowledge and information services to patrons. These systems, which serve as an addition to the primary library complex, may make well-informed judgments about where to look for and how to use information resources (Asemi, A et al. 2021).

Teaching and learning methods at all levels of the academic system have changed as a result of the recent, sharp acceleration of the digital revolution in education. At the center of this transition are digital platforms and e-books, which have become crucial instruments for enhancing access to educational materials and fostering more flexible, personalized and interactive learning (Schmidt, J. T., & et al. 2020). The broad trend toward a move to digital techniques intended to enhance the caliber and accessibility of academic information is reflected in their growing adoption by universities, instructors, and students.

E-book platforms have become essential in higher education, where prompt access to relevant learning resources is critical. They provide educators with tools to better promote student engagement, allow institutions to improve resource distribution, and allow learners to connect with information in a more dynamic way (Du Plooy, E. et al., 2024). Platforms that provide creative, flexible, and integrated learning solutions are becoming more and more necessary as the need for intelligent educational technology rises.

The adoption of automated library systems intersects with broader questions of public policy and management. The governance of digital knowledge infrastructures requires balancing efficiency with principles of equity and inclusivity, ensuring that technological change does not deepen information gaps among different groups of users (Wang, 2024). Libraries are increasingly positioned as public institutions that support national and institutional strategies for digital inclusion, making their automation initiatives subject to policy frameworks around access, funding, and accountability (Ciancarini, Giancarlo, & Grimaudo, 2023). Decisions about licensing agreements with commercial e-platforms also reflect management challenges related to transparency, sustainability, and the negotiation of public–private partnerships. These considerations emphasize that intelligent platforms are not only tools for enhancing educational services, but also key components

[1]School of Information Technologies and Applied Mathematics, SDU University, Kazakhstan assem.talasbek@sdu.edu.kz
[2]School of Information Technologies and Applied Mathematics, SDU University, Kazakhstan andrey.bogdanchikov@sdu.edu.kz
[3]School of Information Technologies and Applied Mathematics, SDU University, Kazakhstan meraryslan.meraliyev@sdu.edu.kz
[4]School of Information Technologies and Applied Mathematics, SDU University, Kazakhstan mukhtar.amirkumar@sdu.edu.kz (corresponding author)
[5]Faculty of Information and Computer Technologies, Paragon International University, Phnom Penh, Cambodia mzhaparov@paragoniu.edu.kh

of public information policy and organizational governance. Embedding them within coherent regulatory and managerial structures is therefore critical to ensuring their long-term effectiveness and public value.

This article presents a comprehensive analysis of the technical specifications and functional capabilities of leading e-platforms, including Perlego, QuizBot, Bookmate, Khan Academy, and Scopus. The study aims to identify the key features, strengths, and limitations of each platform in terms of their technological implementation and user experience. Special attention is given to evaluating the applicability of these platforms in the development of an adaptive system for integrating publishing resources with library infrastructures. The analysis highlights the potential of leveraging electronic publishing technologies and artificial intelligence to enhance access to educational and research content. The findings provide a foundation for future development of intelligent library services.

Literature Review

Integration of AI to e-library systems results in various benefits, the most important of which is increased efficiency (Oyetola et al., 2023). However, it does not come without concerns, such as job losses and disruptions of service flow (Ali et al., 2022). Successful integration of AI requires more than just having the right technology. It also involves selecting appropriate practical solutions, addressing algorithmic bias, ensuring ethical use, and establishing clear policies for AI implementation.

In academic and digital library environments, where AI is increasingly seen as a transformational tool for improving service delivery and user engagement, recent research has shed further light on these processes. For example, a comparison of AI strategies in academic libraries in mainland China and the United Kingdom shows glaring disparities in adoption: Chinese universities usually highlight AI in their vision statements, emphasizing the development of new majors and research initiatives in AI technologies, whereas UK university strategies hardly ever mention AI explicitly (Huang et al., 2023). With Chinese libraries more aggressively deploying “smart” or “intelligent” systems, this disparity demonstrates how national goals impact AI integration. In contrast, UK libraries lag in strategic priority, which may restrict their flexibility in digital settings.

Adding to this, empirical research conducted among Polish university librarians using Delphi methods presents AI primarily as a helpful tool rather than a partner or enemy. Experts concur that AI is useful for tasks like indexing, literature reviews, and collection management, but they are hesitant to use it for complex information advising because of concerns about information quality and result verification (Kisilowska-Szurmińska, 2025). In order to reduce biases and promote favorable attitudes toward adoption, the study emphasizes the necessity of developing librarians’ critical thinking and ethical judgment skills as well as their AI literacy. Similar to this, a SWOT analysis of AI applications in Pakistani university libraries finds that while there are opportunities for increased user personalization, the risks of job displacement outweigh the advantages (such as better automation in cataloging and reference services), while also highlighting infrastructure limitations, funding shortages, and ethical concerns like data privacy (Ali et al., 2024). These results are consistent with larger patterns in which societal concerns limit AI’s ability to automate repetitive jobs, highlighting the need for ethical frameworks and employee training to promote fair adoption. Additionally, a study on librarian acceptance of AI in Islamabad’s academic libraries, using the Technology Acceptance Model, found that perceived interactivity and usefulness significantly influence attitudes toward AI, while ease of use has minimal impact, suggesting the need for awareness programs to boost adoption (Hussain, 2025).

By offering tailored recommendations that suit user interests and habits, artificial intelligence (AI) has been crucial in the field of recommendation systems, helping users overcome information overload. On a dataset of 5,000 academic abstracts, a prototype AI-driven framework that combined machine learning (ML) and natural language processing (NLP) for semantic search in digital libraries showed better performance, attaining 87 % precision and 85 % recall as opposed to 68 % and 64 % for conventional keyword-based methods, while increasing user satisfaction by 35 % (Mahadi et al., 2025). While issues like computing costs and data quality dependence present obstacles for smaller institutions, the system’s adaptive ranking and feedback loops demonstrate how hybrid AI models may improve suggestions over time. In addition, sentiment analysis and topic modeling of new developments in library and information science show that people have favorable opinions about AI-powered recommendation systems. Themes highlight personalized book recommendations and RFID-enabled cataloging as ways to improve user engagement and resource efficiency (Rajeevan et al., 2025). As discussed in Library 5.0, where AI allows personalized material suggestions based on user data, such systems not only encourage chance discovery but also facilitate multidisciplinary access, transforming libraries from static archives to dynamic, intelligent ecosystems (Noh, 2023).

Studies supporting transformer-based models like BERT to enhance query comprehension and incorporate knowledge graphs for multilingual support highlight AI's potential for personalization in digital libraries. This enhances recommendation accuracy and inclusivity (Mahadi et al., 2025). Initiatives like European's AI implementations for discoverability serve as examples of how AI recommendation tools are becoming more and more integrated into collection development policies in academic settings, which are moving toward digital-first strategies that use machine learning to improve metadata and provide user-centric recommendations (Hasan & Panda, 2025). However, ethical issues like algorithmic bias in suggestions that can maintain disparities in resource access must be addressed in light of these developments. Additionally, a bibliometric study of 354 papers published between 2010 and 2023 reveals important trends in AI for academic libraries, including chatbots and machine learning, with a spike in production after 2015 and a focus on intelligent systems, exposing ethical and collaborative gaps (Islam et al., 2025).

Though there aren't much research specifically on the subject, AI techniques are becoming increasingly important for assessing user advancement and institutional effectiveness in educational libraries. In line with evaluation frameworks to provide formative feedback and resource optimization, broader analyses show that AI can enable real-time analytics and predictive modeling for borrowing patterns (Scott & Ward, 2025). AI literacy tests, for example, indicate gaps in librarians' readiness in higher education settings; surveys indicate that librarians' self-rated abilities are modest, and training in ethical evaluation of AI outputs is recommended to guarantee objective evaluations (Lo, 2024). When combined with platforms like QuizBot, these techniques allow for compliance with educational standards like Bloom's taxonomy, including progress monitoring and individualized quizzes. However, human control is necessary to reduce biases and hallucinations in created material. To elaborate, the AILIS 1.0 framework assesses AI literacy in library and information science, identifying fundamental aspects such as Functioning, Ethics, and Evaluation. Professionals outperform students in this framework, underscoring the necessity of focused training to address overestimation of competencies (Montesi et al., 2025).

AI-related ethical concerns in libraries have drawn more attention recently, with research identifying prejudice, privacy violations, and transparency as the main challenges. A Delphi survey highlights librarians' concerns about job loss and dehumanization, and it calls for ethical standards that give human supervision in AI implementations first priority (Kisilowska-Szurmińska, 2025). Similar to this, SWOT studies in Pakistani contexts highlight ethical conundrums such as data permission and the effects on society's privacy as threats, calling for laws promoting responsible AI usage (Ali et al., 2020). The necessity for equitable techniques is emphasized by broader evaluations of AI ethics in libraries, with bibliometric analyses revealing a spike in publications from 2020–2025 that concentrate on global consequences, such as psychological hazards and environmental repercussions (Qiu et al., 2025). The integration of AI in Uzbekistan's higher education poses questions about cultural adaptation and legal frameworks, suggesting standards for moral governance to strike a balance between innovation and equality (Khudayberganovna et al., 2025). Together, these observations highlight how crucial it is for libraries to provide explicit ethical guidelines, carry out bias assessments, and make retraining investments in order to capitalize on AI's advantages while preserving user confidence and inclusion.

Moreover, explorations of ChatGPT in library services reveal high awareness and usage for customer service (86.67 %) and research assistance (58.67 %), but privacy (65 %) and ethical concerns (60 %) persist, supporting integration with ethical safeguards (Patra et al., 2025). Similarly, generative AI demonstrates strong potential to enhance access services through personalized learning and efficient circulation management, yet risks such as bias and workforce displacement necessitate comprehensive planning and staff training (Boateng, 2025).

All things considered, research from 2020 to 2025 presents AI as a two-edged sword in e-library systems: a driver of effectiveness, customization, and adaptive services, but also rife with moral, practical, and human-centered issues. Future advancements should put an emphasis on user-centered designs, collaborative frameworks, and multilingual assistance in order to construct integrated platforms that improve equality and accessibility in academic settings. Best practices should be informed by these varied worldwide viewpoints.

Methodology

To better understand how academic teaching staff in Almaty universities use electronic resources and what functions they would like to see in an AI-based e-library system, we carried out an anonymous survey.

The questionnaire was designed in such a way that it did not collect demographic information such as gender, age, university affiliation, or academic rank. This decision ensured maximum anonymity but also limited the possibility of subgroup analysis and assessment of representativeness. The focus of the study was specifically on the practices and expectations of university teaching staff. The survey asked about the most important features of e-book platforms, such as access to content, ease of use, integration with university systems, and interactive learning tools. In total, 114 people took part in the survey, and after data cleaning, 66 complete responses were included in the analysis. 48 responses were excluded because they were incomplete or inconsistent. These responses lacked answers to key sections required for analysis (e.g., methods of access and evaluation of platform functionality). To ensure comparability and reliability of the dataset, only 66 complete and coherent responses were included in the final analysis.

An online questionnaire was used to collect data, designed to study practices of e-book usage and to identify requirements for the functionality of intelligent library platforms. The questionnaire included both closed and open-ended questions, which allowed for gathering both quantitative and qualitative information.

The structure of the survey consisted of several blocks. The first block contained questions about the language of the survey, the intensity of e-resource use compared to printed editions, and the channels of access to digital materials (university subscriptions, open internet resources, commercial platforms, etc.). The second block focused on the evaluation of basic and advanced e-library functions: interface usability, content search, offline access, support for various formats, annotation and bookmarking, as well as integration with learning management systems.

The third block included questions about perceptions of innovative features such as automated quizzes, personalized recommendations, and simulation tools. The final part of the questionnaire consisted of open-ended questions, allowing respondents to specify the most valuable functions, possible challenges, and additional suggestions.

This structure of the instrument made it possible to obtain comprehensive data on current practices and user expectations, providing a foundation for further analysis and the development of recommendations for designing AI-oriented e-library systems.

Today there are many digital platforms that provide access to educational and research content. They differ in their focus, functions, and business models, but all of them try to make learning more convenient and engaging. Below are several examples that are most relevant for our project.

Perlego is often called the “Spotify for textbooks.” It gives users unlimited access to more than one million academic and professional books. The platform includes useful study tools such as highlighting, notes, bookmarks, and automatic citations. A special feature is the AI-powered Smart Search and a Research Assistant that helps students and researchers work more effectively. Perlego works on different devices and supports offline reading.

Access is based on a subscription: $18 per month or $144 per year. For institutions, there are special packages with administrative tools and integration with learning systems. The main limitation is that the platform does not yet support the Kazakh language.

Perlego’s business model is subscription-based. The service offers two plans: monthly for $18 and annual for $144 (equivalent to $12 per month). For institutions, there are special packages with administrative tools and integration with learning systems. This model provides users with full access to an extensive academic library without the need to purchase individual books, increasing accessibility and ease of use. Perlego’s institutional business model is subscription-based and offers two tiers: 1)) up to 100 users—the standard package offers instant registration through the website, access to over a million titles, personalized reading and research tools, and administrative management. 2) For more than 100 users—customized pricing with enhanced support (dedicated manager, training webinars, integration with university systems, and priority support).

Registration is possible via links, email invitations, auto-accounts, and LTI (Learning Tool Interoperability), ensuring seamless integration. This model makes access to academic resources scalable and adaptable to the needs of educational institutions (Perlego, 2024).

Bookmate is a subscription service for e-books and audiobooks. Its catalog includes over 1.8 million titles in 16 languages. One of its main features is social reading: users can follow friends, share quotes, and recommend books. The service also offers personalized book suggestions and offline access on different devices.

Bookmate’s business model is built on a tiered subscription. A free account provides access to a comprehensive library of public domain works, a standard subscription expands the selection of e-books and audiobooks, and a premium plan provides full access to the library, including new bestsellers and exclusive content. This system allows users to tailor their spending to their personal reading habits.

To expand its reach, Bookmate is developing partnerships with mobile operators and equipment manufacturers, such as Kcell (Kazakhstan), Indosat (Indonesia), StarHub (Singapore), Azercell (Azerbaijan), and Tigo (Latin America). These agreements ensure app pre-installation and payment via operator bills. Furthermore, a multilingual catalog and compatibility with various devices make the platform attractive to educational institutions seeking to digitalize their educational materials.

Thus, the combination of a subscription model and strategic partnerships allows Bookmate to position itself as an accessible and flexible tool for a global audience.

Quizbot.ai is designed to help teachers and students create and use quizzes. The system uses artificial intelligence to generate different types of questions (multiple-choice, true/false, fill-in-the-blank, and others). It can work with various sources, such as PDFs, videos, or websites, and export quizzes to LMS platforms like Moodle or Canvas. Quizbot.ai also has plagiarism detection, grammar correction, and tools for tracking student progress.

The platform works on a subscription model, with a free trial and several paid plans. It is flexible and scalable, but like Perlego and Bookmate, it does not yet support Kazakh.

Quizbot.ai's business model is subscription-based. The platform offers a free trial with a limited number of questions and several paid plans that include advanced features, such as plagiarism checking, an AI editor, and tools for identifying AI-generated content. This model allows users to choose the optimal option for both personal and institutional use.

For educational institutions, Quizbot.ai is attractive due to its integration with LMS, multilingual support, and the ability to create differentiated assessments. These features make the platform a useful tool for improving testing systems and enhancing the quality of the educational process.

Khan Academy is a free educational platform that provides courses in many subjects, from math and science to economics and history. Lessons are presented in short video lectures and interactive exercises. The system adapts to the level of each learner, and teachers and parents can monitor progress through dashboards. Gamification elements (badges, points) are also used to keep students motivated. The strengths of Khan Academy are accessibility and adaptiveness. However, advanced university-level content is limited, and live tutoring is not available.

Khan Academy operates on a non-profit, open-access model: all content is provided free of charge, and funding is provided through grants, donations, and partnerships. This model makes education as accessible as possible but limits the development of paid services.

Scopus is one of the largest databases of academic publications. It covers more than 78 million publications and provides detailed author profiles and citation metrics (h-index, CiteScore, etc.). Researchers use it to track trends, analyze literature, and identify reliable journals for publication. Scopus also integrates with many external systems and offers strong visualization tools. The main drawback is the high subscription cost and lack of Kazakh language support.

Scopus operates on a subscription model (B2B): universities, research organizations, and libraries purchase access to the database through a license. Revenue is generated through institutional subscriptions, but the high price limits access for individual researchers and organizations with limited budgets.

Results

As previously mentioned, after cleaning the data and removing incomplete questionnaires, we had **66 respondents** left. In the multiple-choice question, they made a total of **99 selections (clicks)**, which reflects the distribution of access methods to e-books. The majority indicated that they download PDF books from various websites — 64 selections, while fewer reported accessing e-books through the university's subscription to a paid database — 17 selections. Only a small number mentioned buying or renting from publishers' websites — 8 selections, or from platforms such as Amazon and others — 10 selections. This shows that most participants rely on informal sources, with official access channels used much less frequently. The chart presented in Figure 1 clearly illustrates which channels of accessing e-books are the most common and which are used less frequently.

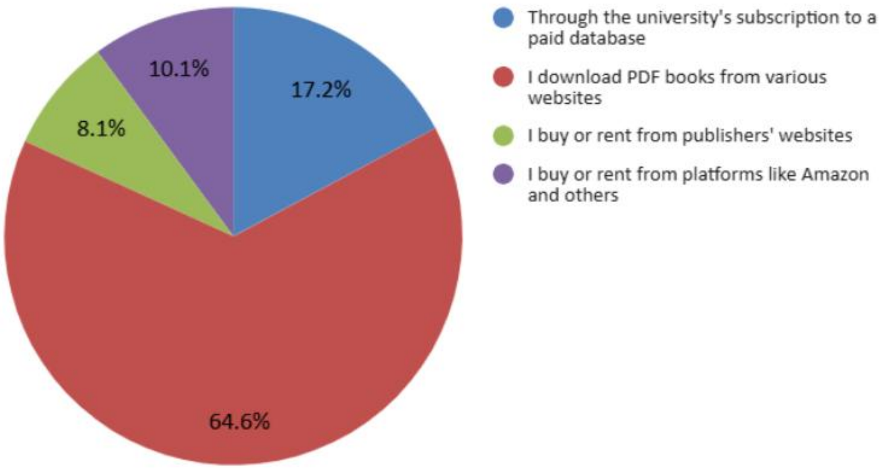

Figure 1. Methods of Accessing E-books

Note — compiled by authors

In Figure 2, the bar chart illustrates the demand for different potential functionalities of the platform, with all ratings indicating consistently high interest, ranging from **8.67 to 9.23** on a ten-point scale. The most significant feature identified by the respondents is **access to a library of electronic books** (9.23), which would provide a wide selection of e-books from various publishers. High demand is also observed for **quiz generation for students** (8.86) and **personalized reading recommendations** (8.86), reflecting the importance of tools that both assess students' understanding and support individualized learning pathways. Slightly lower, though still highly valued, is the functionality of **uploading course learning results and resource recommendations** (8.85), enabling the integration of learning outcomes with relevant e-book suggestions. Respondents also expressed considerable interest in **student performance analysis** (8.74), offering insights into progress and achievement, and in **textbook content suggestions for publishers** (8.67), where course results could inform the creation of more relevant educational materials. Overall, the chart demonstrates a comprehensive demand for such a platform, as respondents expect it to provide not only broad access to resources but also advanced tools for analysis and personalization of the learning process.

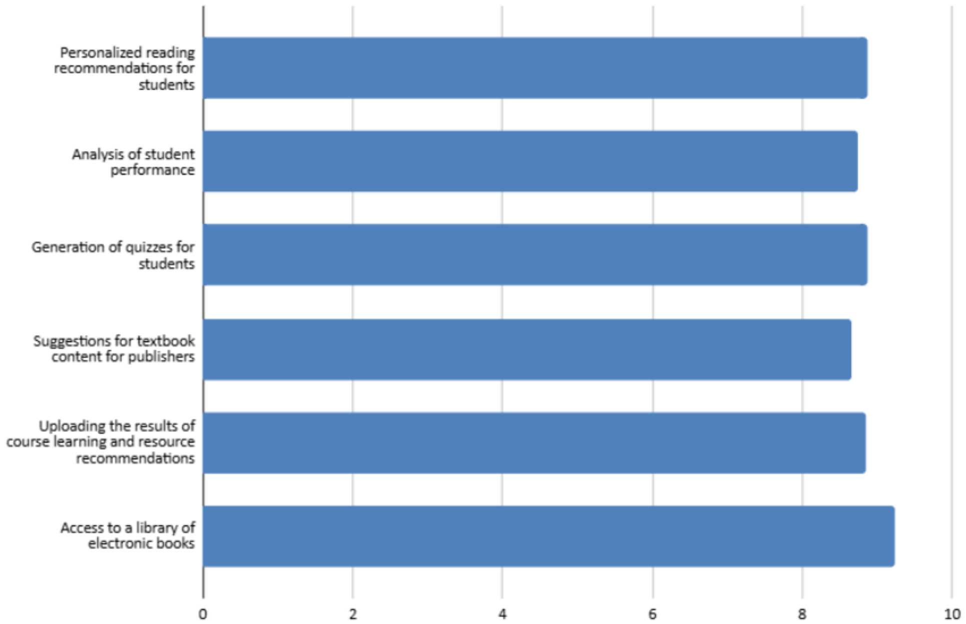

Figure 2. Demand for Platform Functionalities

Note — compiled by authors

Below in Table 1 Comparative summary of existing platforms is presented and highlights the unique strengths of each platform: Perlego emphasizes academic content with AI-powered search and study tools, though its coverage of non-English or region-specific materials may be limited; QuizBot provides adaptive testing and assessment capabilities, but its effectiveness depends on question quality and algorithm accuracy; Khan Academy delivers large-scale, free learning through adaptive pathways, yet it lacks specialized academic resources for advanced research; Scopus offers advanced analytics and citation tracking for research, although access requires costly subscriptions and not all journals are included; and Bookmate stands out with its multilingual catalog and strong social and mobile features, but it is less academically oriented and may not fully support scholarly research needs.

Table 1. Comparative Summary

#	Platform	Learning content	Research Indexing	Personalization	AI tools	Gamification	Integration to other services	Limitations
1	Perlego	Yes (academic books)	No	Yes	Yes	No	Limited	Does not support Kazakh language
2	Quizbot	Yes (quizzes, assessment)	No	Yes	Yes	Partial	No	Does not support Kazakh language
3	Khan Academy	Yes (K-12 prep)	No	Yes	No info	Yes	Yes	- limited options for advanced courses; - support for Kazakh language only subtitles
4	Scopus	No	Yes	No	Partial	No	Yes	Does not support Kazakh language
5	Bookmate	Yes (e-books, audiobooks)	No	Yes	No info	Social	Moderate	Does not support Kazakh language

Note — compiled by authors

In exploring successful digital education tools, three platforms—Perlego, Bookmate, and Quizbot.ai—offer useful models for scalable and user-friendly access to educational content.

Perlego positions itself as an academic-focused e-library. With over one million textbooks available through a monthly or yearly subscription, it caters to both individual learners and institutions. Features like AI-powered search, offline reading, and integrated note-taking make it ideal for academic use. For institutions, Perlego offers scalable subscription models, administrative tools, and LTI-based integration for seamless adoption into learning environments.

Bookmate, by contrast, emphasizes accessibility and engagement. Its library includes over 1.8 million eBooks and audiobooks in 16 languages. The platform supports social reading and personalized recommendations, aiming to make reading a community-driven, mobile-first experience. Bookmate also partners with mobile operators and educational institutions to expand its reach, offering flexible plans to suit different reader profiles.

Quizbot.ai is a different kind of tool—AI-driven and focused on assessment creation. It generates diverse question types from various content formats and aligns them with Bloom's taxonomy. With multilingual support and LMS integration (e.g., Moodle, Canvas), Quizbot is designed for global educational use. Its subscription model ranges from a limited free trial to paid plans with advanced AI features and institutional support options.

Together, these platforms demonstrate how digital education tools can combine content breadth, user-centric design, and institutional scalability. Their diverse approaches to content delivery and monetization offer valuable insights for developing flexible, modern learning ecosystems.

These platforms illustrate a range of approaches to digital content delivery and educational support. While Perlego emphasizes academic resources and institutional scalability, Bookmate combines reading with social engagement and multilingual access, and Quizbot.ai focuses on AI-driven content generation and assessment tools. Each platform's business model reflects a growing demand for flexible, user-centric, and scalable solutions in digital education, offering important insights for the design of future adaptive learning platforms.

Currently, the publishing industry's business model in Kazakhstan is primarily based on selling printed publications, which generate the bulk of revenue. Electronic resources are also available on the market. However, their distribution is primarily limited to the institutional segment: large universities purchase access to digital libraries to support their students and faculty. This approach creates a relatively narrow market for digital solutions and does not reach a broad audience of individual users.

A promising model could include expanding e-book subscriptions to a broader range of universities and developing personalized services for end users. Introducing adaptive features, integration with educational platforms, and improved user-friendliness could generate sustainable interest in electronic publications. Such a transformation would not only stimulate growth in the digital resource segment but also ensure a gradual shift toward a hybrid use of print and electronic formats.

One promising approach is pay-per-use pricing for e-books with revenue sharing. If the project scales up to the CIS, co-financing options are possible. A key prerequisite for the success of this model will be ongoing technical support and an analytics system: data on book usage is necessary for fair profit distribution and assessing content demand.

Recommendations and Development Steps

Author Engagement — Incentivize textbook creation through revenue sharing mechanisms, grant support, and partnerships with universities.

Focus on Digital Format — Develop functionality that increases interest in e-books: interactive elements, quizzes, and analytics for students and teachers.

Technical Support — Built-in content integration and maintenance services to compensate for the lack of IT resources

Analytics System — Transparent accounting of book usage will form the basis for trust and fair revenue distribution.

Regional Development — From the early stages, it is important to consider the requirements and specific features of different CIS countries for future expansion.

Conclusion

The survey results demonstrate that electronic resources already occupy a significant place in the academic practices of university staff. However, most respondents continue to rely on informal access channels, while official institutional subscriptions remain insufficiently used. At the same time, there is a clear and steady demand for advanced functionalities—comprehensive e-libraries, automatic quiz generation, personalized recommendations, and analytical tools—that could support both teaching and research activities. This points to the necessity of developing an integrated digital platform capable of combining wide content access with intelligent educational services.

A comparative overview of platforms such as Perlego, Quizbot.ai, Khan Academy, Bookmate, and Scopus reveals that the market offers a wide range of technological approaches. Each platform has distinct strengths: some emphasize academic rigor and research metrics (Perlego, Scopus), while others focus on user adaptability and accessibility (Quizbot, Bookmate). Despite these differences, the convergence around interoperability, personalization, and intelligent support systems emerges as a general trend. At the same time, existing limitations—lack of support for Kazakh language, partial integration with local educational systems, and incomplete content coverage—remain significant barriers.

In the context of Kazakhstan, the publishing industry still relies mainly on sales of printed publications, which generate the largest share of revenue. Electronic resources are present, but their distribution is largely confined to institutional subscriptions purchased by major universities. This approach restricts the digital market and leaves individual users outside its scope. A more promising direction could be the expansion of subscription access to a wider range of universities, coupled with the development of personalized services for end users. Such steps would not only broaden the audience for digital content but also foster a gradual transition to a hybrid system where print and electronic formats complement each other.

An additional factor for sustainable development is the adoption of new business models. Pay-per-use mechanisms with revenue sharing between publishers and platform providers appear particularly viable. If scaled to the CIS region, such models could be supported by co-financing schemes. However, their success will depend on constant technical support and the establishment of transparent usage analytics, which are essential both for fair distribution of income and for evaluating the demand for content.

Taken together, the findings of this study underscore the urgent need for a unified adaptive platform that would bring together publishers, libraries, and learners. Its effectiveness will depend on AI-powered services, multilingual functionality, and reliable long-term technical maintenance. The implementation of such infrastructure can help overcome current fragmentation, improve accessibility, and contribute to the modernization of Kazakhstan's publishing sector in line with international trends in higher education.

Acknowledgment: This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan under the project titled "Development of an adaptive platform for integrating publishers and libraries using electronic publishing and artificial intelligence technologies" (AP23483919)

References

Ali, M.Y., Naeem, S.B., & Bhatti, R. (2020). Artificial intelligence tools and perspectives of university librarians: An overview. *Business Information Review, 37*(3), 116-124. https://doi.org/10.1177/0266382120952016

Ali, M.Y., Naeem, S.B., Bhatti, R., & Richardson, J. (2024). Artificial intelligence application in university libraries of Pakistan: SWOT analysis and implications. *Global knowledge, memory and communication, 73*(1/2), 219-234. https://doi.org/10.1108/GKMC-12-2021-0203

Asemi, A., Ko, A., & Nowkarizi, M. (2021). Intelligent libraries: a review on expert systems, artificial intelligence, and robot. *Library Hi Tech, 39*(2), 412-434. https://doi.org/10.1108/LHT-02-2020-0038

Boateng, F. (2025). The transformative potential of Generative AI in academic library access services: Opportunities and challenges. *Information Services and Use, 45*(1-2), 140-147. https://doi.org/10.1177/18758789251332800

Ciancarini, P., Giancarlo, R., & Grimaudo, G. (2024). Digital transformation in the public administrations: a guided tour for computer scientists. *IEEE Access,* 12, 22841–22865. https://doi.org/10.1109/access.2024.3363075

Du Plooy, E., Casteleijn, D., & Franzsen, D. (2024). Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement. *Heliyon, 10*(21). https://doi.org/10.1016/j.heliyon.2024.e39630

Hasan, S., & Panda, S. (2025). From Shelves to Servers: Revamping Collection Development Policies for a Digital-First Approach. *Transforming Library Services: Innovation, Access and Research Support,* 3-29. https://doi.org/10.5281/zenodo.15446620

Huang, Y., Cox, A. M., & Cox, J. (2023). Artificial Intelligence in academic library strategy in the United Kingdom and the Mainland of China. *The Journal of Academic Librarianship, 49*(6), 102772. https://doi.org/10.1016/j.acalib.2023.102772

Hussain, A., & Khan, A. (2025). Librarian acceptance of artificial intelligence in academic libraries of Islamabad: An application of the technology acceptance model. *Technical Services Quarterly,* 1-21. https://doi.org/10.1080/07317131.2025.2512280

Islam, M.N., Ahmad, S., Aqil, M., Hu, G., Ashiq, M., Abusharhah, M.M., & Saky, S.A.T.M. (2025). Application of artificial intelligence in academic libraries: a bibliometric analysis and knowledge mapping. *Discover Artificial Intelligence, 5*(1), 59. https://doi.org/10.1007/s44163-025-00295-9

Khudayberganovna, K.D., Asqar o'g, N.M.A., Ziyoviddinovna, M.Z., & Xaydaraliyevna, P.D. (2025). Integration of artificial intelligence technologies into higher education in Uzbekistan: Aspects, challenges and development pathways. *IMRAS, 8*(5), 51-55. https://doi.org/10.5281/zenodo.15449458

Kisilowska-Szurmińska, M. (2025). Artificial intelligence in academic libraries–a tool, a collaborator, an adversary? A Delphi study of university librarians in Poland. *The Journal of Academic Librarianship, 51*(5), 103114. https://doi.org/10.1016/j.acalib.2025.103114

Lo, L.S. (2024). Transforming academic librarianship through AI reskilling: Insights from the GPT-4 exploration program. *The Journal of Academic Librarianship, 50*(3), 102883. https://doi.org/10.1016/j.acalib.2024.102883

Mahadi, A., Pranowo, R.D., Firdaus, M.R., Senobudi, H., & Efgivia, M.G. Semantic Search Optimization in Digital Libraries: Enhancing Precision and User Satisfaction Through NLP-Driven Query Interpretation and Machine Learning Ranking.

Montesi, M., Bornstein, B.Á., Puig, N.B., Ochando, M.B., & Díez, A.S. (2025). AILIS 1.0: A new framework to measure AI literacy in library and information science (LIS). *The Journal of Academic Librarianship, 51*(5), 103118. https://doi.org/10.1016/j.acalib.2025.103118

MS, R., Devi, B., Anoop, V.S., & Mallikarjuna, C. (2025). Mapping of Emerging Technological Trends in Library and Information Science: A Computational Approach Using Sentiment Analysis, Topic Modeling and Network Analysis. *DESIDOC Journal of Library & Information Technology, 45*(4). https://doi.org/10.14429/djlit.21029

Noh, Y. (2023). A study on the discussion on Library 5.0 and the generation of Library 1.0 to Library 5.0. *Journal of Librarianship and Information Science, 55*(4), 889-905. https://doi.org/10.1177/09610006221106183

Oyetola, S.O., Oladokun, B.D., Maxwell, C.E., & Akor, S.O. (2023). Artificial intelligence in the library: Gauging the potential application and implications for contemporary library services in Nigeria. *Data and Metadata, 2*(1), 5. https://doi.org/10.56294/dm202336

Qiu, J., Cheng, L., & Huang, J. (2025). Charting the Landscape of Artificial Intelligence Ethics: A Bibliometric Analysis. *International Journal of Digital Law and Governance*, (0). https://doi.org/10.1515/ijdlg-2025-0007

Patra, N.K., Gaitanou, P., Oladokun, B.D., & Sahu, S. (2025). Exploring the integration of AI in library services: Perspectives and considerations. *Business Information Review, 42*(2), 133-142. https://doi.org/10.1177/02663821251343742

Perlego (2024, November 11). An online subscription for all your academic books. *perlego.com.* Retrieved from https://www.perlego.com

Schmidt, J.T., & Tang, M. (2020). Digitalization in education: challenges, trends and transformative potential. In *Führen und managen in der digitalen transformation: Trends, best practices und herausforderungen* — In *leading and managing in digital transformation: Trends, best practices and challenges* (pp. 287-312). Wiesbaden: Springer Fachmedien Wiesbaden.

Scott, S.M., & Ward, L. (2025). The Use of AI in Academic Libraries. In *Examining AI Disruption in Educational Settings: Challenges and Opportunities* (pp. 59-100). IGI Global Scientific Publishing. http://dx.doi.org/10.4018/979-8-3693-7873-1.ch003

Wang, C., & Si, L. (2024). The Intersection of Public Policy and Public Access: Digital Inclusion, Digital Literacy Education, and Libraries. *Sustainability*, *16*(5), 1878. https://doi.org/10.3390/su16051878

https://doi.org/10.31489/2025Ec4/1200

JEL: C53, I21, J11
ORIGINAL RESEARCH

Received: 19.06.2025 | Accepted: 07.10.2025

Comparative Analysis of Forecasting Models for Student Enrollment in Kazakhstan's General Secondary Education System

Duman Tleuzhan[1], Alexandr Pilipenko[2], Kuandyk Tleuzhanuly[3]*

Abstract
We compared seven forecasting models to predict student enrollment in Kazakhstan's schools using data from 2020-2024. We tested cohort component models, cohort survival models, trend regression with demographic factors, linear trend models, exponential smoothing, multi-factor regression, and weighted moving averages across 20 regions (17 regions and 3 cities) with about 3.9 million students. We measured how accurate each model was using Mean Absolute Percentage Error (MAPE). We trained models on 2020-2023 data and tested them on 2024 numbers. The linear trend model worked best, with 0.70 % MAPE nationally and 0.77 % MAPE across regions. Demographic models didn't work as well—cohort models did poorly at the regional level even though they have good theory behind them. Our forecasts for 2025-2027 show national enrollment growing from 3.9 million to 4.2 million students, but growth varies a lot by region. Big cities like Astana will grow 24.05 % and Almaty 12.81 %, while some regions will barely grow at all. Our results help educational planners pick the right forecasting methods for different areas. This study fills a research gap for post-Soviet countries where detailed forecasting evaluations are hard to find.

Keywords: student enrollment forecasting, educational planning, demographic modeling, time series analysis, Kazakhstan education system, regional analysis.

Introduction

Accurate forecasting of student enrollment is one of the basic challenges facing educational systems worldwide. Educational planners need reliable projections to make sure that infrastructure, teaching staff, and financial resources can meet future demand. This challenge is especially difficult in developing countries where demographic transitions, internal migration, and rapid urbanization create complex planning environments. Kazakhstan's educational system shows these complexities well. The country currently serves about 3.9 million students across 21 major administrative units in a vast and diverse territory. Recent demographic trends have created different pressures: birth rates are declining in some regions while major urban centers see population growth. Meanwhile, large-scale internal migration from rural to urban areas—particularly toward cities like Astana, Almaty, and Shymkent—makes enrollment projections even harder.

The stakes of accurate enrollment forecasting go well beyond simple planning exercises. Enrollment projections directly affect decisions about teacher recruitment and deployment, school construction and renovation programs, budget allocations, and broader educational policy. When forecasts are wrong, the consequences can be serious. Underestimating future enrollment leads to overcrowding and lower educational quality, while overestimating wastes public funds and creates inefficient resource allocation. Despite how important enrollment forecasting is, surprisingly little research has carefully evaluated different forecasting methods using complete data from Kazakhstan's educational system. Most existing studies either focus narrowly on single methods or lack enough historical data to properly test model performance. Moreover, Kazakhstan's unique characteristics—its linguistic diversity, varied settlement patterns, and recent administrative reforms—require specialized analytical approaches that may not transfer directly from other contexts.

This study aims to compare the performance of seven different forecasting models for predicting total student enrollment across Kazakhstan's regional educational system. We work with a complete dataset covering 21 territorial units over five years from 2020 to 2024, including demographic indicators, enrollment statistics, migration patterns, and educational infrastructure characteristics. Our specific goals are fourfold: first, we check whether demographic-based models or statistical time series approaches give more accurate enrollment forecasts; second, we identify which methods work best at different territorial levels from national to regional scales; third, we generate validated forecasts for 2025-2027 to support practical educational

[1]Narxoz University, Almaty, Kazakhstan, duman.tleuzhan@narxoz.kz
[2]Taldau, National Centre for Education Research and Evaluation named after A. Baitursynuly, Astana, Kazakhstan, a.pilipenko@taldau.edu.kz
[3]Narxoz University, Almaty, Kazakhstan, kuandyk.tleuzhanuly@narxoz.kz (corresponding author)

planning; and fourth, we analyze regional variations in enrollment patterns and consider what they mean for resource allocation strategies.

This research contributes through systematic application of multiple forecasting methods to a complete regional dataset, offering evidence-based guidance for educational planners and policymakers. The study addresses a significant research gap by providing real-world validation of model performance within Kazakhstan's distinctive educational and demographic context.

Literature Review

School enrollment forecasting has become essential for educational planning worldwide. It helps policymakers anticipate infrastructure needs, allocate resources efficiently, and ensure access to education. Forecasting methods have evolved with technological advances, but developing countries have shown creativity in adapting these methods to environments where data is limited. Despite growing research in this field, certain regions—particularly post-Soviet countries—remain underrepresented in the literature, pointing toward significant opportunities for investigation.

Time Series Models: Foundation of Enrollment Forecasting

Time series models are the foundational approach to enrollment forecasting, offering solid methods for capturing temporal patterns in educational data. Tang and Yin (2012) established exponential smoothing's relevance in the United States by comparing it with grey prediction models for forecasting education expenditure and enrollment. They found that grey models showed higher accuracy, though exponential smoothing remained valuable as a baseline. Chen (2022) found different results in China, where exponential smoothing actually outperformed grey prediction, ARIMA, and neural network models in forecasting enrollment proportions. These contrasting results show how model performance varies significantly across different national contexts.

The ARIMA methodology has proven versatile in addressing forecasting challenges across both developed and developing nations. Marinoiu (2014) successfully applied the Box-Jenkins methodology to develop an ARIMA model for forecasting gross enrollment ratio in Romanian primary schools, projecting concerning declines. Kornelio et al. (2024) used ARIMA models in Tanzania to project significant enrollment increases in government primary schools. Chen et al. (2021) extended to ARIMAX models incorporating exogenous variables to analyze student-teacher ratios in China's primary education system. Yan (2024) applied Vector Autoregression models to analyze the decline of rural primary schools in China, revealing how urbanization patterns impact educational infrastructure. The comparative effectiveness of these time series approaches shows a crucial finding: no single method dominates across all contexts, making careful consideration of local conditions essential.

Machine Learning and Artificial Intelligence Approaches

The integration of machine learning techniques into enrollment forecasting represents a shift from traditional statistical methods, offering better capabilities for capturing complex, non-linear patterns. James (2021) pioneered the use of Long Short-Term Memory networks at Kansas State University, finding that these deep learning models significantly outperformed both traditional exponential smoothing and standard deep neural networks. Support Vector Machines have emerged as powerful tools when multiple influencing factors must be considered simultaneously. Aksenova et al. (2006) demonstrated this at California State University, Sacramento, where SVM models incorporating demographic variables, income levels, tuition fees, and unemployment rates achieved remarkable accuracy.

Hybrid intelligent systems mark the current frontier in machine learning applications. The Adaptive Neuro-Fuzzy Inference System, as applied by Aji et al. (2023) in Indonesia, combines the learning capabilities of neural networks with the interpretability of fuzzy logic systems. Shafii et al. (2021) applied fuzzy time series models in Malaysia, achieving the highest accuracy for primary school enrollment. The democratization of machine learning tools has also facilitated innovative applications in resource-constrained environments. Sahane et al. (2014) used data mining techniques with accessible tools in India's Aurangabad district, demonstrating that sophisticated analytical capabilities can be deployed even in contexts with limited technical infrastructure.

Spatial and Geographic Information Systems Models

The incorporation of spatial dimensions into enrollment forecasting acknowledges the fundamentally geographic nature of educational access and demand. Geographic Information Systems have changed the field by enabling planners to visualize and analyze the spatial distribution of student populations alongside school infrastructure. Langley's (1997) pioneering work in Leeds, United Kingdom, established the foundation for spatial modeling in educational planning. Haynes (2014) applied spatial Bayesian modeling to enhance small-area enrollment projections in Iowa, improving the accuracy of grade progression rates while accounting for geographic dependencies and migration patterns.

Miller (2008) implemented the Integrated Planning for School and Community model in Wake County, North Carolina, demonstrating how GIS-based approaches can align educational infrastructure development with urban planning. The Greater London Authority's (GLA Intelligence, 2018) school place demand projections illustrated how spatial modeling can inform strategic infrastructure decisions in complex urban environments. Wang et al. (2023) extended spatial methodologies for China's new urban districts by incorporating micro-level factors such as residential location and dwelling type, addressing overcrowding through optimized space utilization.

Demographic and Cohort-Based Approaches

Cohort-survival methods represent the traditional backbone of enrollment forecasting, particularly in contexts with stable demographic patterns. Braden et al. (1972) established the cohort-survival technique as the gold standard for communities with predictable population dynamics. Pajankar and Srivastava's (2019) Reconstructive Cohort Approach for India demonstrates how traditional cohort concepts can be modified for contexts where detailed enrollment data is unavailable, requiring only population figures, repetition rates, and transition rates.

Fabricant and Weinman (1972) applied least squares regression to forecast first-grade enrollment in New York neighborhoods, incorporating variables such as new housing developments, busing policies, and ethnic composition. Grip and Grip's (2019) comparative analysis of confidence intervals and stochastic forecasting using Monte Carlo simulations for New Jersey school districts represents methodological advancement, emphasizing customized approaches based on district size and growth patterns.

System Dynamics and Integrated Modeling Approaches

System dynamics modeling offers a holistic perspective on enrollment forecasting by capturing the complex feedback loops and policy interactions that influence educational participation. Pedamallu et al. (2010) applied system dynamics with cross-impact analysis in developing country contexts, revealing how infrastructural improvements create reinforcing cycles that boost enrollment while reducing dropout rates. The comparative studies by Rynerson and colleagues (2018, 2021, 2022) for various Oregon school districts exemplify integrated approaches, combining demographic analysis, cohort progression, and scenario planning to provide 15-year enrollment forecasts.

National and Comparative Perspectives

The scale of forecasting efforts ranges from local district projections to national policy planning. Hussar and Bailey's (2016) national and state-level enrollment projections for the United States through 2024 illustrate how macro-level forecasting informs federal educational policies and funding allocations. Huynh Van et al. (2019) analyzed Vietnam's six geographical regions and revealed significant disparities in enrollment trends. Chen's (2022) work on China's general and vocational education enrollment ratios highlighted how forecasting must account for regional economic structures and labor market needs.

The evolution of school enrollment forecasting methodologies reflects both technological advancement and adaptive innovation across diverse global contexts. Several key insights emerge from this review. First, no single forecasting method dominates across all contexts; method selection must carefully consider local data availability, demographic patterns, and institutional characteristics. Second, the trend toward hybrid and integrated approaches suggests that future advances may come from creative combinations of existing techniques. Third, the incorporation of spatial dimensions and system dynamics perspectives enriches traditional temporal forecasting.

However, significant gaps remain in the literature. The conspicuous absence of studies from post-Soviet regions and many developing countries suggests that current methodological knowledge may not fully capture the diversity of global educational contexts. Kazakhstan and other Central Asian countries, with their unique demographic transitions and educational system transformations, represent particularly important yet understudied contexts.

Methodology

Data Description

This study utilizes a comprehensive dataset covering Kazakhstan's general secondary education system using the National Education Database (NEDB) for the years 2020-2024. The dataset encompasses 21 territorial units including the national level, 17 regional administrations, and 3 cities of republican significance

(Astana, Almaty, and Shymkent). The primary dataset integrates multiple data sources including demographic indicators, educational enrollment statistics, migration flows, school infrastructure characteristics, and human resource metrics. Demographic variables include birth rates from 2013-2024, enabling cohort tracking with appropriate time lags for school entry age. Educational variables encompass total student enrollment by year, grade-level distributions, and language of instruction categories. Migration data captures both incoming and outgoing student flows between regions, providing insights into population mobility patterns affecting enrollment. Infrastructure variables include school capacity, number of schools, shift arrangements, and facility conditions. Human resource indicators cover teacher numbers, qualifications, and turnover rates. All data sources maintain consistent territorial classifications and temporal coverage, ensuring compatibility across different analytical approaches. Missing values were addressed through interpolation techniques where appropriate, with sensitivity analysis confirming minimal impact on model outcomes.

Forecasting Models

We tested seven different forecasting methods to look at enrollment dynamics from different angles. Each model represents a specific way of predicting enrollment, from demographic-based methods to statistical time series techniques.

Cohort Component Model

The cohort component model is one of the more sophisticated approaches to educational enrollment forecasting. It comes directly from demographic methodology used in population projections. This model works on the idea that future first-grade enrollment can be predicted by tracking specific birth cohorts as they move toward school age. The method starts with historical birth data, usually requiring a six-year lag for standard school entry age. The basic formula is:

$$E_t = B_{t-6} \times S_t \times P_t \times M_t$$

where E_t is enrollment at time t, B_{t-6} is births six years earlier, S_t is the enrollment rate from birth to school age, P_t represents the participation rate in formal education, and M_t captures net migration effects. The model includes survival rates from birth to school age, and participation rates that show what proportion of eligible children actually enter formal education. Migration effects are included through net migration coefficients. For later grades, the cohort progression follows:

$$E_{g,t} = E_{g-1,t-1} \times R_{g,t}$$

where $E_{g,t}$ is enrollment in grade g at time t, and $R_{g,t}$ is the grade progression rate accounting for retention, dropout, and migration. This approach works well in stable demographic settings where birth patterns, migration flows, and educational participation rates follow predictable paths over time.

Cohort Survival Model

The cohort survival model builds on the cohort component framework by adding more sophisticated survival analysis techniques borrowed from actuarial science and epidemiology. Unlike the basic cohort model that uses simple survival rates, this approach recognizes that survival probabilities can vary across different demographic groups, geographic regions, and socioeconomic levels. The model extends the basic formula to include differential survival rates:

$$E_{g,t} = E_{g-1,t-1} \times S_{g,i,t} \times C_{g,i,t}$$

where $S_{g,i,t}$ is differential survival rates for demographic subgroup i in grade g at time t, and $C_{g,i,t}$ denotes continuation probabilities. The model includes differential survival rates that can account for variations in healthcare access, economic conditions, and other factors that influence child survival from birth to school entry age. The cohort survival model adds the concept of educational continuation probabilities, recognizing that not all children who survive to school age will participate in formal education. These continuation probabilities can be modeled as functions of various socioeconomic indicators:

$$C_{g,i,t} = f(X_{i,t})$$

where $X_{i,t}$ is a set of socioeconomic predictors including family characteristics, regional educational infrastructure quality, and economic conditions. This enhanced approach gives more accurate projections where demographic and socioeconomic differences significantly affect educational participation patterns.

Trend Regression with Demographic Factors

Trend regression models with demographic factors combine statistical trend analysis with demographic change theory. These models recognize that enrollment patterns often show both systematic trends over time and responses to underlying demographic drivers. The general form is:

$$E_t = \beta_0 + \beta_1 t + \beta_2 B_{t-k} + \beta_3 N_t + \beta_4 D_t + \epsilon_t$$

where E_t is enrollment at time t, t is the time trend, B_{t-k} is lagged birth rates, N_t captures net migration, D_t represents population density or other demographic variables, and ϵ_t is the error term. The statistical part involves fitting regression equations to historical enrollment data, using time as the main independent variable to capture long-term trends in educational participation. The demographic part adds predictor variables like birth rates, migration flows, age structure changes, and population density variations. More advanced versions may include interaction effects:

$$E_t = \beta_0 + \beta_1 t + \beta_2 B_{t-k} + \beta_3 (t \times B_{t-k}) + \beta_4 N_t + \epsilon_t$$

allowing for non-linear relationships and changing coefficients over time. The methodology often uses multiple regression techniques with various forms including polynomial trends, logarithmic transformations, and piecewise linear functions to capture complex enrollment dynamics. Model specification usually involves testing for autocorrelation using the Durbin-Watson statistic, heteroscedasticity through Breusch-Pagan tests, and structural breaks using Chow tests. These models work well where enrollment patterns are influenced by both long-term demographic transitions and shorter-term policy or economic changes.

Linear Trend Model

The linear trend model, despite its simplicity, is a solid and widely-used forecasting technique that often beats more complex alternatives in practical use. This approach assumes that enrollment changes follow a consistent linear pattern over time, with a constant absolute change per time period. The math involves estimating a simple linear regression equation:

$$E_t = \alpha + \beta t + \epsilon_t$$

where E_t is enrollment at time t, α is the intercept representing baseline enrollment, β trend coefficient, showing average annual change, and ϵ_t is the random error term. Model estimation usually uses ordinary least squares regression, with parameter estimates given by:

$$\hat{\beta} = \frac{\sum_{t=1}^{n}(t - \bar{t})(E_t - \bar{E})}{\sum_{t=1}^{n}(t - \bar{t})^2}$$
$$\hat{\alpha} = \bar{E} - \widehat{\beta \bar{t}}$$

More sophisticated techniques like generalized least squares may be used when autocorrelation is present in the residuals. The forecasting process involves extending the fitted trend line into future periods:

$$\hat{E}_{t+h} = \hat{\alpha} + \hat{\beta}(t + h)$$

where h is the forecast horizon. Prediction intervals are calculated based on the standard error of the forecast:

$$PI_{t+h} = \hat{E}_{t+h} \pm z_{\alpha/2} \times SE(\hat{E}_{t+h})$$

where $SE\left(\hat{E}t+h\right)=s\sqrt{1+\frac{1}{n}+\frac{(t+h-\bar{t})^2}{\sum t=1^n(t-\bar{t})^2}}$ and s is the residual standard error. While critics often dismiss linear models as too simplistic, research in forecasting accuracy often shows that linear trends give surprisingly accurate predictions for many educational systems, especially over short to medium-term horizons.

Exponential Smoothing (Holt's Method)

Exponential smoothing using Holt's method is a sophisticated time series forecasting technique that handles both level and trend components in enrollment data. Unlike simple exponential smoothing that assumes no systematic trend, Holt's method recognizes that many enrollment series show persistent upward or downward movements that must be explicitly modeled. The method uses two smoothing equations: a level equation that updates the estimated current enrollment level and a trend equation that updates the estimated rate of change. The level equation is:

$$L_t=\alpha E_t+(1-\alpha)(L_{t-1}+T_{t-1})$$

and the trend equation is:

$$T_t=\beta(L_t-L_{t-1})+(1-\beta)T_{t-1}$$

where L_t is the level at time t, T_t is the trend at time t, α is the level smoothing parameter $(0<\alpha<1)$, and β is the trend smoothing parameter $(0<\beta<1)$. Parameter estimation often uses optimization techniques to minimize forecast errors over historical data, usually minimizing the sum of squared errors:

$$\min_{\alpha,\beta}\sum_{t=1}^{n}\left(E_t-\widehat{E_t}\right)^2$$

The forecasting process generates predictions by adding trend projections to the current level estimate:

$$\hat{E}_{t+h}=L_t+h\times T_t$$

where h is the forecast horizon. Advanced implementations may include damped trend modifications:

$$\hat{E}_{t+h}=L_t+(\phi+\phi^2+\ldots+\phi^h)T_t$$

where ϕ is the damping parameter $(0<\phi<1)$, which assumes trend effects diminish over longer forecast horizons, addressing the common problem of exponential trend extrapolation producing unrealistic long-term projections.

Multi-Factor Regression Model

Multi-factor regression models are the most comprehensive statistical approach to enrollment forecasting, trying to capture the complex mix of demographic, economic, social, and institutional factors that influence educational participation. The general form is:

$$E_t=\beta_0+\beta_1B_{t-k}+\beta_2P_t+\beta_3I_t+\beta_4C_t+\beta_5T_t+\epsilon_t$$

where E_t is enrollment at time t, B_{t-k} represents lagged birth rates, P_t denotes population density or demographic structure, I_t captures income or socioeconomic indicators, C_t represents school capacity or infrastructure measures, T_t is teacher availability, and ϵ_t is the error term. These models usually include demographic variables like birth rates, age structure, and population density; socioeconomic indicators including income levels, parental education, and employment rates; infrastructure measures like school capacity and teacher availability; and policy variables reflecting changes in educational requirements or funding levels. Model specification requires attention to multicollinearity among predictor variables, checked through variance inflation factors:

$$VIF_j=\frac{1}{1-R_j^2}$$

where R_j^2 is the coefficient of determination from regressing predictor j on all other predictors. Advanced versions may use stepwise regression for variable selection, ridge regression to address multicollinearity with penalty term:

$$\hat{\beta}ridge = arg\,min\,\beta\left[\sum_{t=1}^{n}(E_t - X_t\beta)^2 + \lambda\sum_{j=1}^{p}\beta_j^2\right]$$

or instrumental variables to handle endogeneity problems. The forecasting process requires projecting values for all predictor variables, which can add extra uncertainty.

Weighted Moving Average Model

Weighted moving average models provide a flexible smoothing approach that emphasizes recent observations while including historical patterns to reduce the influence of random fluctuations in enrollment data. Unlike simple moving averages that give equal weights to all observations within the averaging window, weighted moving averages use a declining weight structure that gives more influence to more recent data points. The basic form is:

$$\hat{E}t = \sum i = 0^{k-1}\, w_i E_{t-i}$$

subject to the constraint $\sum_{i=0}^{k-1} w_i = 1$ and typically $w_0 > w_1 >\ldots> w_{k-1}$. The weighting scheme typically follows exponential decay patterns:

$$w_i = \frac{(1-\lambda)\lambda^i}{1-\lambda^k}$$

where λ is a decay parameter ($0 < \lambda < 1$), or geometric progressions, or custom weight distributions designed to match the specific characteristics of the enrollment series. The methodology involves calculating weighted averages of recent enrollment observations and then projecting these smoothed values into future periods. Trend components can be incorporated by calculating the weighted average of first differences:

$$\hat{T}t = \sum i = 0^{k-2}\, w_i(E_{t-i} - E_{t-i-1})$$

and adding these trend estimates to the level projections:

$$\hat{E}_{t+h} = \hat{E}_t + h \times \hat{T}_t$$

The forecasting process often includes automatic weight optimization procedures that minimize historical forecast errors to determine optimal weight parameters, typically by minimizing mean squared error:

$$\min_{w_0,\ldots,w_{k-1}} \sum_{t=k}^{n}\left(E_t - \widehat{E_t}\right)^2$$

Advanced versions may use adaptive weighting schemes that adjust the weight distribution based on how stable or volatile recent enrollment patterns are.

Model Validation and Performance Assessment

We evaluated model performance using out-of-sample validation for accurate assessment. All models were trained only on 2020-2023 data, with 2024 as the testing period. This approach prevents overfitting and gives a realistic assessment of how models perform in actual forecasting situations.

We measured accuracy using Mean Absolute Percentage Error as the main metric. MAPE is easy to interpret because it shows forecast errors as percentages of actual values, making comparison across different areas and enrollment scales straightforward. The MAPE is calculated as:

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n} \left| \frac{E_t - \hat{E}_t}{E_t} \right|$$

where E_t is the actual enrollment at time t, $\hat{E}_t$is the forecasted enrollment, and $n$is the number of observations. We calculated MAPE both at the national level and as average regional performance to check model consistency across different geographic scales. The validation process included sensitivity analysis to test model stability under different parameter settings and data changes. We used cross-validation techniques where possible to make sure performance assessments were solid.

Statistical Implementation

All statistical analyses were conducted using Stata version 18. For the cohort component and cohort survival models, we employed the gen and replace commands to create lagged birth variables and calculate survival rates, with cohort progressions tracked using conditional statements and the bysort prefix for regional groupings. The trend regression with demographic factors utilized the regress command with robust standard errors (vce(robust)) to estimate coefficients, followed by post-estimation diagnostics including the estat vif command to assess multicollinearity, estat hettest for heteroscedasticity testing (Breusch-Pagan test), and estat dwatson for autocorrelation detection (Durbin-Watson statistic). Linear trend models were fitted using regress enrollment year with region-specific estimations executed through statsby loops. Exponential smoothing forecasts implementing Holt's method were generated using the tssmooth exponential command with optimal smoothing parameters identified through the optimize() function, minimizing mean squared error. Multi-factor regression specifications employed stepwise regression procedures for variable selection, with variance inflation factors computed via vif to identify problematic multicollinearity (threshold VIF > 10), and, when necessary, ridge regression was implemented using the ridgereg command to address collinearity issues. Weighted moving average calculations utilized the ma() function within time series operators, with custom weight specifications created through mata programming for non-standard weighting schemes. Model performance evaluation was conducted using out-of-sample validation, where training-test splits were created using if conditions (2020-2023 for training, 2024 for testing), and Mean Absolute Percentage Error was calculated using egen functions combined with summarize commands. Forecasts for 2025-2027 were generated using the predict command with the dynamic() option for multi-step-ahead predictions, and confidence intervals were constructed using the forecast suite of commands with standard error estimates from predict, stdp. All data management, including merging demographic indicators with enrollment statistics, was performed using merge commands with m:1 relationships, and missing values were handled through linear interpolation using the ipolate command where appropriate.

Results

The comparative analysis shows substantial differences in how well various forecasting methods worked across different territorial levels. Table 1 presents MAPE results for all seven models we tested on total student enrollment.

Table 1. Model Performance for Total Student Enrollment Forecasting (MAPE %, 2024)

Model	National Level MAPE	Regional Average MAPE	Overall Rank
Linear Trend Model	0.70 %	0.77 %	1
Exponential Smoothing (Holt's)	1.00 %	1.40 %	2
Multi-factor Regression	1.21 %	4.32 %	3
Trend Regression with Demographics	1.84 %	1.63 %	4
Weighted Moving Average	4.16 %	3.53 %	5
Cohort Survival Model	4.96 %	12.57 %	6
Cohort Component Model	7.92 %	18.28 %	7
Note — compiled by authors based on the sources (estimates based on National Education Database 2020-2024)			

The linear trend model turned out to be remarkably accurate, posting the best scores both nationally (0.70 % MAPE) and across regions (0.77 % MAPE). What makes this particularly useful is how consistently well it performed regardless of whether we looked at the whole country or individual regions. The model's straightforward nature combined with this level of accuracy makes it an obvious choice for actual planning work. Exponential smoothing came in second, doing quite well at the national level (1.00 % MAPE) and maintaining decent regional accuracy (1.40 % MAPE). This method's ability to adapt to recent changes while staying stable explains why it performed reasonably well across the board. What really surprised us was how poorly the demographic models did compare to simpler statistical methods. The cohort component model, which has solid theoretical backing in population dynamics, only managed 7.92 % MAPE nationally and 18.28 % regionally. The cohort survival model wasn't much better at 4.96 % and 12.57 % respectively. This suggests that straightforward demographic relationships aren't capturing what's actually happening in Kazakhstan's education system. The multi-factor regression showed an interesting split—it did well nationally (1.21 % MAPE) but struggled with regional predictions (4.32 % MAPE), which hints at possible overfitting or trouble handling how different regions behave. Since the linear trend model worked so well, we used it to forecast enrollment for 2025-2027. Table 2 shows what we're projecting for different parts of the country.

Table 2. Regional Forecasts Using Linear Trend Model (2025-2027)

Region	Actual 2024	Predicted 2024	MAPE 2024 (%)	Prediction 2025	Prediction 2026	Prediction 2027	Predicted Average Annual Growth Rate 2025-2027 (%)	Predicted Total Growth Rate 2025-2027 (%)
Kazakhstan Total	3,904,496	3,931,649	0.70	4,010,283	4,116,070	4,221,856	2.64	8.13
Abay region	102,494	102,732	0.23	102,876	103,259	103,641	0.37	1.12
Akmola region	142,117	143,257	0.80	144,565	147,013	149,461	1.69	5.17
Aktobe region	185,947	187,020	0.58	192,094	198,241	204,387	3.20	9.92
Almaty region	368,688	371,026	0.63	383,977	399,267	414,556	3.99	12.44
Atyrau region	153,554	153,659	0.07	157,760	161,966	166,171	2.67	8.22
West-Kazakhstan region	126,034	127,571	1.22	129,380	132,725	136,071	2.59	7.96
Zhambyl region	247,648	248,835	0.48	248,972	250,296	251,620	0.53	1.60
Zhetisu region	133,352	133,391	0.03	134,658	135,963	137,269	0.97	2.94
Karaganda region	175,974	177,034	0.60	177,310	178,646	179,982	0.75	2.28
Kostanay region	115,480	116,432	0.82	116,382	117,284	118,185	0.77	2.34
Kyzylorda region	184,917	187,528	1.41	188,260	191,602	194,945	1.78	5.42
Mangystau region	190,643	191,426	0.41	198,277	205,911	213,545	3.85	12.01
Pavlodar region	118,413	119,975	1.32	119,640	120,866	122,093	1.03	3.11
North-Kazakhstan region	74,952	76,229	1.70	74,993	75,034	75,074	0.05	0.16
South-Kazakhstan region	523,618	528,495	0.93	529,168	534,717	540,267	1.05	3.18
Ulytau region	41,218	41,372	0.37	41,923	42,629	43,334	1.68	5.13
East-Kazakhstan region	101,805	102,435	0.62	102,997	104,188	105,380	1.16	3.51
Astana City	286,913	284,716	0.77	309,909	332,906	355,902	7.45	24.05
Almaty City	360,008	366,076	1.69	375,380	390,752	406,123	4.10	12.81
Shymkent City	270,721	272,442	0.64	281,765	292,809	303,853	3.92	12.24
Note — compiled by authors based on the sources (estimates based on National Education Database 2020-2024)								

Nationally, we're looking at steady growth from 3.9 million students in 2024 to 4.2 million by 2027—an 8.13 % jump. This reflects ongoing demographic momentum and continuing educational system expansion. The really striking pattern is how unevenly this growth is distributed. Astana leads the pack with a projected 24.05 % increase over three years. This dramatic growth stems from continued urbanization and people moving to the capital for economic opportunities and government jobs. Almaty City is projected to grow 12.81 %, and Shymkent 12.24 %, showing how these major commercial centers keep attracting families. Among regular regions, Almaty region shows 12.44 % growth, Mangystau hits 12.01 %, and Aktobe reaches 9.92 %. These align pretty well with where economic development is concentrated and where resource industries are pulling in workers and their families. The flip side is what's happening in rural and peripheral areas. North-Kazakhstan region barely budges, with just 0.16 % growth over three years. Abay region shows 1.12 % growth, and Zhambyl comes in at 1.60 %. These modest numbers reflect the rural-to-urban migration wave and falling birth rates in these areas, creating entirely different planning headaches for local authorities.Looking at how accurate our 2024 forecasts were across regions, most came in under 1.5 % MAPE. A few outliers include West-Kazakhstan (1.22 %), Kyzylorda (1.41 %), Pavlodar (1.32 %), and Almaty City (1.69 %), though even these are acceptable for planning purposes. North-Kazakhstan region had the highest error at 1.70 %, possibly because of more erratic demographic shifts or data quality issues there.

Discussion

The fact that simple statistical models beat theoretically sophisticated demographic approaches has significant implications for how educational forecasting should work in Kazakhstan. The linear trend model's accuracy tells us that enrollment patterns over the short term follow fairly predictable growth paths that don't require complex demographic modeling to capture. This goes against what demographers usually recommend for education forecasting.

Why did demographic models perform so poorly? Several Kazakhstan-specific factors probably explain this. First, the link between birth rates and school enrollment becomes complicated when you have the kind of migration Kazakhstan's been experiencing. The country has seen major internal population movements over the past decade—people leaving rural areas for cities, moving from smaller towns to Astana, Almaty, and Shymkent. Traditional cohort models assume populations remain stable geographically, which clearly doesn't match Kazakhstan's reality. Beyond that, educational participation rates and policy changes have likely disrupted old patterns. Recent reforms—changes to school entry age requirements, curriculum overhauls—could have broken historical relationships between birth cohorts and enrollment. Furthermore, five years of data just is not enough for demographic models to establish reliable parameters. These models typically need much longer time series to estimate survival and progression rates properly.

These findings fit with what forecasting research has found more broadly: simple models often beat complex ones in practice, especially when you don't have much historical data. The principle here is that unnecessary complexity introduces new sources of error rather than improving accuracy. Our results support this principle specifically for educational forecasting in developing countries. The enrollment patterns we're projecting show significant regional gaps that need different policy responses. High-growth urban centers like Astana and Almaty face serious infrastructure challenges. That 24 % growth in Astana over three years means they must act quickly on school construction, teacher hiring, and budget allocations. Based on typical class sizes and teacher-student ratios, Astana probably needs 15-20 new schools and around 3,000 new teachers by 2027 just to maintain current standards. Almaty and Shymkent face similar pressures, though not quite as extreme.

Meanwhile, regions with flat enrollment like North-Kazakhstan have different problems concerning efficiency and quality. Maintaining the existing school network becomes harder to justify when student numbers aren't growing. These places might need to think about consolidating schools—fewer institutions with better facilities, more specialized staff, stronger programs—rather than spreading resources thinly across many small schools. That kind of consolidation could actually improve quality while cutting costs.

The urban-rural split in growth patterns mirrors broader trends in Kazakhstan's development. Educational planners must balance infrastructure investment between booming urban areas that need expansion and slower-growth rural areas that need efficiency improvements. This involves difficult trade-offs between equity (keeping rural schools open) and efficiency (concentrating resources where students actually are). Policymakers need better frameworks for making these allocation decisions that consider both immediate enrollment pressures and long-term regional development goals. Regional variations also affect teacher deployment. High-growth regions need many more teachers, creating recruitment and training challenges. Kazakhstan might need incentive programs to attract qualified teachers to move to rapidly growing areas—maybe housing assistance, salary increases, or faster career advancement. At the same time, slow-growth regions might have too many teachers, requiring retraining programs or helping people transition to other education sectors or regions.

Our analysis has several limitations pointing toward future research. First, the short time series (2020-2024) doesn't capture long-term cycles or structural breaks in enrollment patterns. The data includes the COVID period, which may have temporarily distorted enrollment in ways that don't reflect normal trends. Longer time series would allow us to validate models more effectively and understand how enrollment responds to major disruptions. Second, regional aggregation hides important local variations. Regions in Kazakhstan contain diverse territories with different urban-rural mixes, economic bases, and demographic profiles. District-level and school-level analysis would provide much more useful insights for local planning. Future work should examine forecasting at these finer scales, perhaps using spatial models to capture geographic dependencies. Third, we only examined total enrollment without breaking it down by grade. Age-specific forecasts by grade would help with curriculum planning, facility design, and teacher specialization. Different grades might grow at different rates because of varying cohort sizes and changing dropout patterns. Understanding these grade-specific dynamics would sharpen operational planning. Fourth, we did not explicitly model external factors like policy changes, economic shocks, or social shifts. Major policy initiatives or economic disruptions could change enrollment in ways historical trends miss. Future research could incorporate these through scenario analysis or econometric models linking enrollment to external drivers. The system dynamics approaches we discussed in the literature review offer robust frameworks for exploring these complex interactions. Fifth, we did not address quality or outcomes, only quantity. Educational planning ultimately is concerned with quality and results. Future work could link enrollment projections to the resources needed to maintain or improve quality, developing optimization models balancing enrollment accommodation with quality standards.

Conclusion

This study offers the first thorough evaluation of enrollment forecasting models using Kazakhstan's full regional educational dataset. The analysis shows that simple statistical approaches, especially linear trend models, outperform complex demographic models for forecasting total student enrollment in Kazakhstan. This has practical implications for educational planning in Kazakhstan and likely other post-Soviet countries facing similar demographic and institutional situations.

The linear trend model's performance—0.70 % national MAPE and 0.77 % regional MAPE—makes it the obvious choice for operational planning in Kazakhstan. Its simplicity means educational authorities at different levels can use it without the need for sophisticated demographic modeling infrastructure or extensive data collection beyond basic enrollment numbers.

Our 2025-2027 forecasts show continued enrollment growth with significant regional differences reflecting Kazakhstan's ongoing urbanization and economic development. National enrollment should grow from 3.9 million to 4.2 million students, an 8 % jump requiring substantial system-wide capacity expansion. However, this growth is concentrated. Urban centers, especially Astana with 24 % projected growth, face major infrastructure needs demanding substantial investments in school construction, teacher recruitment, and educational resources. Planning authorities in these fast-growing areas need aggressive capacity expansion to prevent overcrowding and maintain quality.

Rural regions with modest growth need different approaches focused on efficiency and quality rather than expansion. These areas might benefit from consolidating schools, concentrating resources in fewer institutions, and implementing new delivery methods like distance learning or shared specialized teachers. Policy frameworks should recognize these different regional needs rather than applying blanket national strategies.

This research contributes to practice by providing evidence-based guidance on model selection and validated projections supporting better resource allocation. The methodological comparison provides educational planners in Kazakhstan and similar contexts practical insights into which forecasting approaches are effective given their data and institutions. Finding that simple models beat complex demographic approaches challenges standard assumptions in educational forecasting and suggests practitioners should not automatically assume theoretical sophistication guarantees forecasting accuracy.

Future research should extend the time series to capture longer-term patterns and validate models across different periods. The current analysis covers a brief period including the unusual COVID disruption; longer datasets would enable stronger conclusions about model performance under various conditions. Increasing geographic resolution to districts and schools would support more detailed planning and allow us to examine local factors influencing enrollment. Adding grade-specific forecasts would improve operational planning for curriculum, facilities, and teacher specialization. Exploring how to incorporate policy variables, economic indicators, and social factors could improve accuracy and enable scenario analysis of how different development paths might affect enrollment.

The methodological framework we've established provides a foundation for improving forecasting practices in Kazakhstan and potentially other Central Asian countries with similar characteristics. The systematic comparison of multiple approaches using comprehensive administrative data offers a model for how educational planning organizations can develop evidence-based forecasting capabilities. Regular updates with new data and periodic reassessment of model performance will ensure planning processes use the most accurate available projections.

The practical implications go beyond academic contribution to directly supporting educational policy and resource planning across Kazakhstan's system. Accurate enrollment forecasts enable proactive rather than reactive planning, allowing authorities to anticipate needs and prepare responses before problems arise. This forward-looking approach supports more efficient resource use, better infrastructure development, and ultimately better educational outcomes through more effective resource management. As Kazakhstan continues to develop and its population changes, we need accurate enrollment forecasting to ensure schools have enough space and all children can receive quality education wherever they live.

Acknowledgment: This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR24993019)

References

Aji, B.W., Septiani, N.Z., Putri, W.M., Irawanto, B., Surarso, B., Farikhin, & Dasril, Y. (2023). Prediction of Indonesia school enrollment rate by using adaptive neuro fuzzy inference system. *Indonesian Journal of Artificial Intelligence and Data Mining*, *6*(1), 40–46.

Aksenova, S.S., Zhang, D., & Lu, M. (2006). Enrollment prediction through data mining. In *Proceedings of the 2006 International Conference on Data Mining* (pp. 510–515). IEEE.

Bernhardt, V.L., Pullum, T.W., & Graham, S.N. (1983, April). *Seattle's small-area approach to forecasting enrollments at the school level*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Braden, B., & Others. (1972). *Enrollment forecasting handbook introducing confidence limit computations for a cohort-survival technique*. New England School Development Council; Educational Facilities Laboratories, Inc.

Chen, C.-C., Chen, Y., & Kang, C.-Y. (2021). Estimating student-teacher ratio with ARIMA for primary education in fluctuating enrollment. *ICIC Express Letters, Part B: Applications*, *12*(6), 515–523.

Chen, Q. (2022). A comparative study on the forecast models of the enrollment proportion of general education and vocational education. *International Education Studies*, *15*(6), 109–126. https://doi.org/10.5539/ies.v15n6p109

Fabricant, R., & Weinman, J. (1972). Forecasting first grade public school enrollment by neighborhood. *Demography*, *9*(4), 625–632.

GLA Intelligence. (2018). *2018 GLA school place demand projections*. Greater London Authority.

Grip, R.S., & Grip, M.L. (2019). Using multiple methods to provide prediction bands of K-12 enrollment projections. *Population Research and Policy Review*, *38*(5), 635–650. https://doi.org/10.1007/s11113-019-09533-2

Haynes, D.A. II. (2014). *Improving enrollment projections through the application of geographic principles: Iowa 1999–2011* (Doctoral dissertation). University of Iowa.

Huynh Van, S., Nguyen Thi, H., Nguyen Vinh, K., Sam Vinh, L., & Giang Thien, V. (2019). Forecasting the results of students attending school in Vietnam by geographical area. *International Journal of Education and Practice*, *7*(3), 274–285. https://doi.org/10.18488/journal.61.2019.73.274.285

Hussar, W.J., & Bailey, T.M. (2016). *Projections of education statistics to 2024* (NCES 2016-013). U.S. Department of Education, National Center for Education Statistics.

James, F. (2021). *Neural network-based time series forecasting of student enrollment with exponential smoothing baseline and statistical analysis of performance* (Master's report). Kansas State University.

Kornelio, S., Balan, R.T., & Deogratias, E. (2024). Forecasting students' enrolment in Tanzania government primary schools from 2021 to 2035 using ARIMA model. *International Journal of Curriculum and Instruction*, *16*(1), 162–174.

Langley, R.J. (1997). *The use and development of geographical information systems (GIS) and spatial modelling for educational planning* (Doctoral dissertation). University of Leeds.

Lavilles, R.Q., & Arcilla, M.J.B. (2012). Enrollment forecasting for school management system. *International Journal of Modeling and Optimization*, *2*(5), 563–566.

Marinoiu, C. (2014). Modeling and forecasting the gross enrollment ratio in Romanian primary school. *Annals of the Constantin Brâncuși University of Târgu Jiu, Economy Series*, *3*, 17–22.

Miller, M.A. (2008). Planning for enrollment growth using land use data to determine future school sites. *Transportation Research Record*, *2074*(1), 12–20. https://doi.org/10.3141/2074-02

Pajankar, V.D., & Srivastava, S. (2019). An approach of estimating school enrolment with reconstructive cohort approach. *Journal of Physics: Conference Series*, *1366*, 012116. https://doi.org/10.1088/1742-6596/1366/1/012116

Pedamallu, C.S., Ozdamar, L., Ganesh, L.S., Weber, G.-W., & Kropat, E. (2010). A system dynamics model for improving primary education enrollment in a developing country. *Organizacija*, *43*(3), 90–100. https://doi.org/10.2478/v10051-010-0010-5

Proehl, R.S. (2000). *Enrollment forecast for Horizon Christian School 2005–2015*. Population Research Center, Portland State University. https://pdxscholar.library.pdx.edu/enrollmentforecasts/71

Rynerson, C., & Ollinger, J. (2018). *Portland Public Schools enrollment forecasts 2018–19 to 2032–33*. Population Research Center, Portland State University. https://pdxscholar.library.pdx.edu/enrollmentforecasts/123

Rynerson, C., & Wei, C. (2021). *Centennial School District enrollment forecasts 2021–22 to 2030–31*. Population Research Center, Portland State University. https://pdxscholar.library.pdx.edu/enrollmentforecasts/148

Rynerson, C., & Wei, C. (2022). *Salem-Keizer Public Schools enrollment forecast 2022–23 to 2041–42*. Population Research Center, Portland State University. https://pdxscholar.library.pdx.edu/enrollmentforecasts/139

Sahane, M., Sirsat, S., Khan, R., & Aglave, B. (2014). Prediction of primary pupil enrollment in government school using data mining forecasting technique. *International Journal of Advanced Research in Computer Science and Software Engineering*, *4*(9), 656–661.

Shafii, N.H., Alias, R., Shamsudin, S.R., & Nasir, D.S.M. (2021). Fuzzy time series for projecting school enrolment in Malaysia. *Journal of Computing Research and Innovation*, *6*(1), 11–21. https://doi.org/10.24191/jcrinn.v6i1.180

Stronge, W.B., & Schultz, R.R. (1981). Models for projecting school enrollment. *Educational Evaluation and Policy Analysis*, *3*(5), 75–81.

Sweeney, S.H., & Middleton, E.J. (2005). Multiregional cohort enrollment projections: Matching methods to enrollment policies. *Population, Space and Place*, *11*(5), 361–379. https://doi.org/10.1002/psp.379

Tang, H.-W.V., & Yin, M.-S. (2012). Forecasting performance of grey prediction for education expenditure and school enrollment. *Economics of Education Review*, *31*(4), 452–462. https://doi.org/10.1016/j.econedurev.2011.12.007

Wang, Q., Wang, Y., & Wang, G. (2023). Research on space potential excavating of existing primary school under the base of schooling demand forecasting model: A case study in new urban districts. *Ain Shams Engineering Journal*, *14*, 101871. https://doi.org/10.1016/j.asej.2022.101871

Yan, B. (2024). Dynamic analysis of influencing factors and forecast of development trend of “disappearance of rural primary schools.” *Journal of Service Science and Management*, *17*(1), 75–88. https://doi.org/10.4236/jssm.2024.171003

Index of articles published in «Bulletin of the Karaganda University. “Buketov Business Review» in 2025