Journal of Education and Human Development
March 2020, Vol. 9, No. 1, pp. 88-102
ISSN: 2334-296X (Print), 2334-2978 (Online)

Published by American Research Institute for Policy Development
DOI: 10.15640/jehd.v9n1a9
URL: https://doi.org/10.15640/jehd.v9n1a9

Updated Content of Education in Kazakhstan: Longitudinal Trajectories of Learning Performance in Mathematics and Science

Dimiter Milkov Dimitrov[1,2] Olga Ivanovna Mozhayeva[3], Aidana Senbekovna Shilibekova[4] Dinara Balakazovna Ziyedenova[5], Zamira Kenshilbayevna Rakhymbayeva[6]

Abstract

Purpose. The purpose of this study is to examine the effect of the Updated Content of Education (UCE) project in Kazakhstan on longitudinal trajectories of learners' performance on diagnostic tests conducted under the piloting of UCE (2015/2019) in the areas of Mathematics and Science.

Research Methods. A longitudinal growth modeling was used to study targeted UCE effects based on a study sample of 2,509 students from 30 pilot schools and 1,082 from control schools.

Findings. The results revealed positive UCE effects on the learners' growth and performance in Mathematics and Science, the presence of two latent classes of growth trajectories in Science, no gender effects, and partial effects of school location (rural/urban) on the learners' performance.

Implications for Research and Practice. It is necessary to continue the work on development of teachers' professional competencies, the development of teaching aids and didactic materials, and criteria-based assessments. Learners from different latent classes of performance in Science need to be identified and provided with support in differential learning strategies.

Keywords: Updated Content of Education, longitudinal performance, growth modeling, latent class analysis.

Introduction

The review of the national educational policy of the Republic of Kazakhstan in secondary education, conducted in 2014 by the Organization for Economic Co-operation and Development (OECD), showed that the secondary school curriculum (a) is extensive and mainly theoretical. and (b) does not contribute to the development of higher-order thinking skills. This is evidenced by the results of Kazakhstani learners in two international studies -- PISA (2012) and TIMSS (2011). The recommendation of OECD experts to the Government of Kazakhstan was to conduct a thorough analysis and revision of the current curriculum for secondary education, reduce its theoretical framework, and ensure the practical application of knowledge. Studies showed that, overall, the school children in Kazakhstan have difficulties in (a) understanding complex written texts, identifying key information, analysis and reasoning for one`s opinion, (b) demonstrating scientific knowledge conducting experiments and providing a rational for their choice, interpreting scientific phenomena, events and processes, applying the theory in real life situations, (c) working with data (tables. diagrams. graphs) and their interpretation, and

[1]Professor Emeritus , George Mason University, USA, and Senior Psychometrician, National Center for Assessment, Saudi Arabia P.O. Box 68566, Riyadh 11537, Saudi Arabia email: ddimitro@gmu.edu
[2]George Mason University, USA
[3] Autonomous Educational Organization "Nazarbayev Intellectual Schools," Kazakhstan AEO "Nazarbayev Intellectual Schools" 21/1 Husein ben Talal street, Nur-Sultan 010000 Republic of Kazakhstan
[4] Autonomous Educational Organization "Nazarbayev Intellectual Schools," Kazakhstan AEO "Nazarbayev Intellectual Schools" 21/1 Husein ben Talal street, Nur-Sultan 010000 Republic of Kazakhstan
[5] Autonomous Educational Organization "Nazarbayev Intellectual Schools," Kazakhstan AEO "Nazarbayev Intellectual Schools" 21/1 Husein ben Talal street, Nur-Sultan 010000 Republic of Kazakhstan
[6] Autonomous Educational Organization "Nazarbayev Intellectual Schools," Kazakhstan AEO "Nazarbayev Intellectual Schools" 21/1 Husein ben Talal street, Nur-Sultan 010000 Republic of Kazakhstan

(d) solving complex problems looking for non-standard methods instead of using traditional algorithms, etc. These findings implied the need to update the system of secondary education in the light of changes taking place in the society worldwide.

Updated Content of Education

To update the system of secondary education in Kazakhstan in light of recommendationsby the OECD, the "BLINDED FOR PEER REVIEW," in collaboration with the Y. Altynsarin National Academy of Education, scientists and teacher-practitioners developed an updated content of education (UCE) which implies the introduction of new State compulsory education standards of the Republic of Kazakhstan, model curricula, subject programs, textbooks andteaching materials, criteria-based assessment system, methods and technologies of teaching, and teacher training courses. The UCE is aimed at the formation of not only academic knowledge, but at the development of functional literacy and a wide range of skills as well.

The revision of the curriculum content in Mathematics and Natural science in primary school is aimed at setting more complex intellectual tasks, such as comparison, analysis, application of knowledge, critical assessment, and argumentation, instead of just memorizing information. The updated content in Mathematics provides for an early acquaintance with mathematical concepts (e.g., ordinary fractions, percentages, etc.), topics on argumentation and interpretation based on resources, combinatorics, and so forth. The updated content in Science adds sections on physics, astronomy and chemistry to the existing materials on biology and geography. In this way the curriculum of Science for primary schools serves as an introductory course for Physics, Chemistry, Biology, and Geography in secondary and high schools, thus ensuring continuity in the school science education.

The transition from the knowledge-centered paradigm to the activity-centered paradigm was carried out through orientation to the depth, not to the volume, of the content of learning in Mathematics and Science. The expected results 'at the exit' are broad-spectrum skills, such as the ability to apply knowledge, think critically, engage in lifelong learning, communicate in diverse communities, possess ICT skills and research skills, work individually and in collaboration with others, and solve tasks and problems responsibly.

Approbation of the UCE

As a part of the transition to the UCE, an approbation of the State Compulsory Standardof Primary Education (SCSPE), subject programs, criteria-based assessment system, methods and technologies of teaching in primary grades of 30 pilot secondary schools of the country wascarried out. The purpose of the monitoring was to assess the quality and effectiveness of the UCE in the context of the real educational process of the pilot schools for its` subsequent implementation in secondary schools. To compare the results, 16 control secondary schools, in which the training was carried out based on the SCSPE of 2012, were involved in the pilot project. The approbation of the UCE was accompanied by four years monitoring (2015-2019), which included diagnostic testing of learners from pilot and control schools in accordance with the SCSPE Approbation Monitoring Methodology (approved by the Ministry of Education and Science of the Republic of Kazakhstan dated July 21, 2015).

Purpose of the Study

The main goal of the present study was to examine UCE effects on longitudinaltrajectories of performance on the diagnostic tests conducted under the piloting of the UCE, as well as to identify possible latent (hidden) classes of such trajectories on diagnostic tests in Mathematics and Science (content on "Natural science & World studies"). The following two research questions (RQs) were addressed:RQ1. What is the immediate pretest-posttest effect of the UCE (pilot vs. control) on thelearners' performance in Mathematics and Science, with the pretest measures at "entrance in grade 1" (September 2015) and posttest measures at "exit in grade 1" (April 2015)?

RQ2. Are there latent classes of learners with different trajectories of performance andwhat is the effect of the UCE (pilot/control), gender, and school location (urban/rural) on the formation of such latent classes (when they exist) and on the growth in Mathematics and Science over the time period of diagnostic testing (2015-2019)?

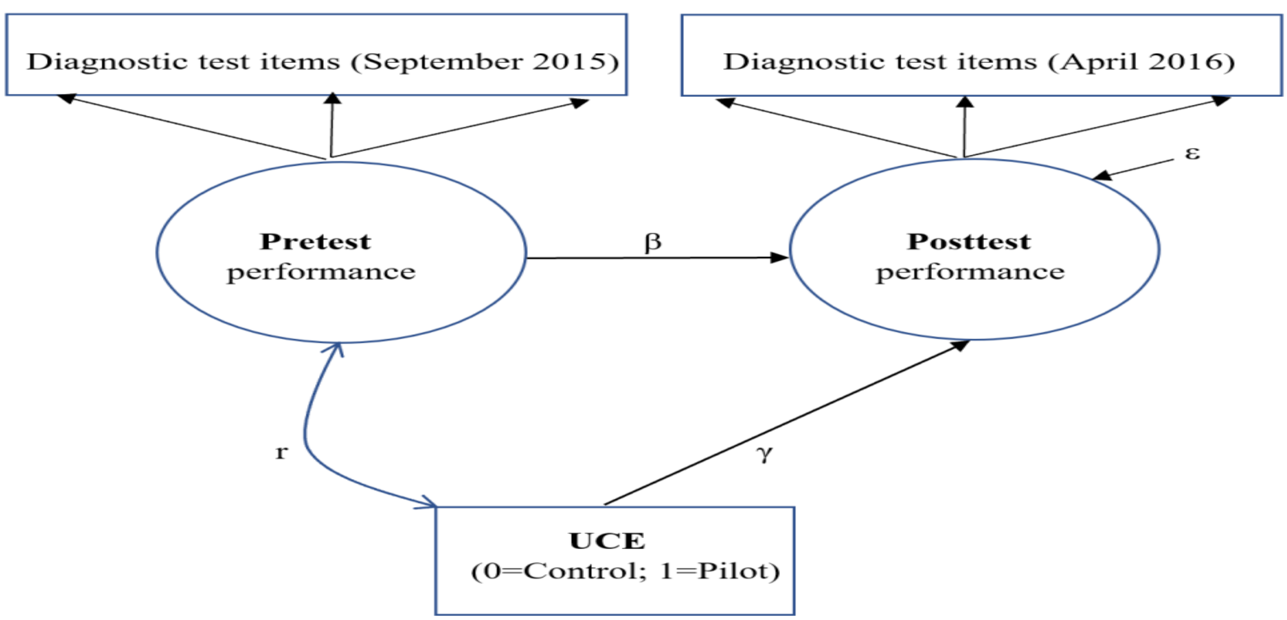
Method

Research Design

The research questions in this study were addressed using a control/experimentallongitudinal growth modeling design for the estimation of UCE effects on the participants' performance in Mathematics and Science across five assessments (diagnostic tests) over a four-year time period (2015-2019). A pretest-posttest design was also used to estimate immediate UCE effect (September 2015 – April 2016). Some details on these two designs is provided next.

Pretest-posttest design. The pretest-posttest design related to addressing RQ1 is developed to compare the UCE (pilot and control) groups on the immediate posttest measures (April 2016) controlling for group differences on the pretest measures (September 2015) in Mathematics and Science. The path diagram of this latent-variable modeling (LVM) design is shown in Figure 1 (e.g., Jöreskog & Sörbom, 1976). In essence, the SEM design in Figure 1 is a LVM analog to the classical analysis of covariance (ANCOVA) approach to analyzing pretest-posttest data (e.g., see Dimitrov, 2013, Chapter 16). The SEM analysis was conducted using the computer program for statistical analysis with latent variables Mplus (Muthén & Muthén, 2009-2018). The analysis under the SEM design in Figure 1 involves two main steps. First, a confirmatory factor analysis (CFA) is used to validate the measurement parts of the model; that is, to validate the factor structure and dependability of the measures associated at the pretest and posttest assessments. Second, the structural part of the SEM model in Figure 1 was analyzed to estimate the UCE effect on the posttest controlling for pretest differences between the groups (pilot vs. control).

Figure 1. Path diagram of LVM design for UCE effect (pilot vs. control) on posttest measures (April 2016) controlling for group differences on the pretest measures (September 2015).



Note. UCE = Updated Content of Education; γ = regression coefficient of the UCE effect on the posttest controlling for pretest differences; β = regression coefficient for the relationship between pretest and posttest measures; r = correlation that provides control of UCE group differences (pilot vs. control) on the pretest measures (e.g.. Jöreskog & Sörbom. 1976).

Longitudinal growth design. A growth mixture modeling (GMM) design was usedto address RQ2 (e.g., Muthén, 2001, 2004). In this case we analyzed longitudinal data across fivediagnostic tests in Mathematics and Science (2015-2019). The GMM analysis allows to evaluate (a) whether the longitudinal trajectories of performance are homogeneous or break into different latent (hidden) classes, and (b) the role of covariates (UCE, gender, and location) in theformation of such classes, as well as the magnitude and significance of the key factors of growth, initial status (intercept) and rate of growth (slope) of growth trajectories by latent classes of learners. A preliminary examination of the learners' longitudinal performance on the diagnostic tests (2005/2015) suggested the appropriateness of using a nonlinear piecewise growth model with linear segments at three stages of longitudinal growth, namely: stage 1 (2015-2016). stage 2 (2016-2018). and stage 3 (2018-2019).

Graphically, the GMM is depicted in Figure 2, where the covariates (UCE, location, gender) are included to facilitate the evaluation of more dynamic process unfolding over time (e.g., Wang & Wang, 2012). The notations of the variables under the GMM in Figure 2 are interpreted as follows:

Y_t= Observed outcome variable at time t of diagnostic testing; (t = 1, 2, 3, 4, 5);
η_0 = initial value (intercept) of the linear trajectory of Yt at stage 1 (2015 – 2016);
η_1 = rate of growth (slope) of Yt at the first stage of growth (2015 – 2016);
η_2 = rate of growth (slope) of Yt at the second stage of growth (2016 – 2018);
η_3 = rate of growth (slope) of Yt at the third stage of growth (2018 – 2019), and
C = categorical latent variable of latent classes.
The analytic form of the conditional GMM is specified for each latent class k as a set ofregression models for the effects of the covariates (UCE, Location, Gender) on the latent factors of growth (η_0, η_1, η_2, η_3) presented as follows:

$Ytk = \eta_{0k} + \eta_{1k}\lambda_{1tk} + \eta_{2k}\lambda_{2tk} + \eta_{3k}\lambda_{3tk} + \varepsilon_{tk}$(1)
$\eta_{0k} = \eta_{00k} + \beta_{01k}(\text{UCE}) + \beta_{02k}(\text{Location}) + \beta_{03k}(\text{Gender}) + \zeta_{0k}$ (2)
$\eta_{1k} = \eta_{10k} + \beta_{11k}(\text{UCE}) + \beta_{12k}(\text{Location}) + \beta_{13k}(\text{Gender}) + \zeta_{1k}$ (3)
$\eta_{2k} = \eta_{20k} + \beta_{21k}(\text{UCE}) + \beta_{22k}(\text{Location}) + \beta_{23k}(\text{Gender}) + \zeta_{2k}$ (4)
$\eta_{3k} = \eta_{30k} + \beta_{31k}(\text{UCE}) + \beta_{32k}(\text{Location}) + \beta_{33k}(\text{Gender}) + \zeta_{3k}$, (5)

where λ_{1tk}, λ_{2tk}, and λ_{3tk} represent time scores for the linear segments at stage 1 (2015-2016), stage 2 (2016-2018), and stage 3 (2018-2019), respectively, of the piecewise growth pattern. The residual term ε_{tk} represents both the random measurement error and time-specific influence of the examinee at time t for latent class k. The intercept coefficient η_{00k} represents the estimated overall mean at the "start" (September 2015), whereas η_{10k}, η_{20k}, and η_{30k} represent the average rate of growth over the time period at stage 1, stage 2, and stage 3, respectively. The slopes (βs) in Equation 2 represent the fixed effects of the covariates (UCE, Location, Gender) on the latent intercept, η_{0k}, whereas those in Equations 3, 4, and 5 are the fixed effects of the covariates on the growth rates η_{1k}, η_{2k}, and η_{3k}, respectively. The error terms ζ_{0k}, ζ_{1k}, $\zeta 2k$, and ζ_{3k} represent the variation in the latent growth factors (intercept in Equation 2 and rates of growth in Equations 3, 4, and 5, respectively).

Research Sample

The study data consist of the learners' test scores obtained from the five diagnostic testsunder the UCE approbation with 30 pilot schools and 16 control schools . The distribution of pilot and control learners across five diagnostic tests with sections in Mathematics and Science is summarized in Table 1. By school location, the 30 pilot schools consist of 18 urban and 12 rural schools, whereas the 16 control schools consist of 8 urban and 8 rural schools. In total, the comparative analysis used the test scores of 3,591 learners, including 1,082 learners of control schools and 2,509 of pilot schools. By gender, the study participants are 1,831 (51.0%) males and 1,760 (49.0%) females. By location, 2,475 (68.9%) are from urban schools and 1,116 (31.1%) from rural schools.

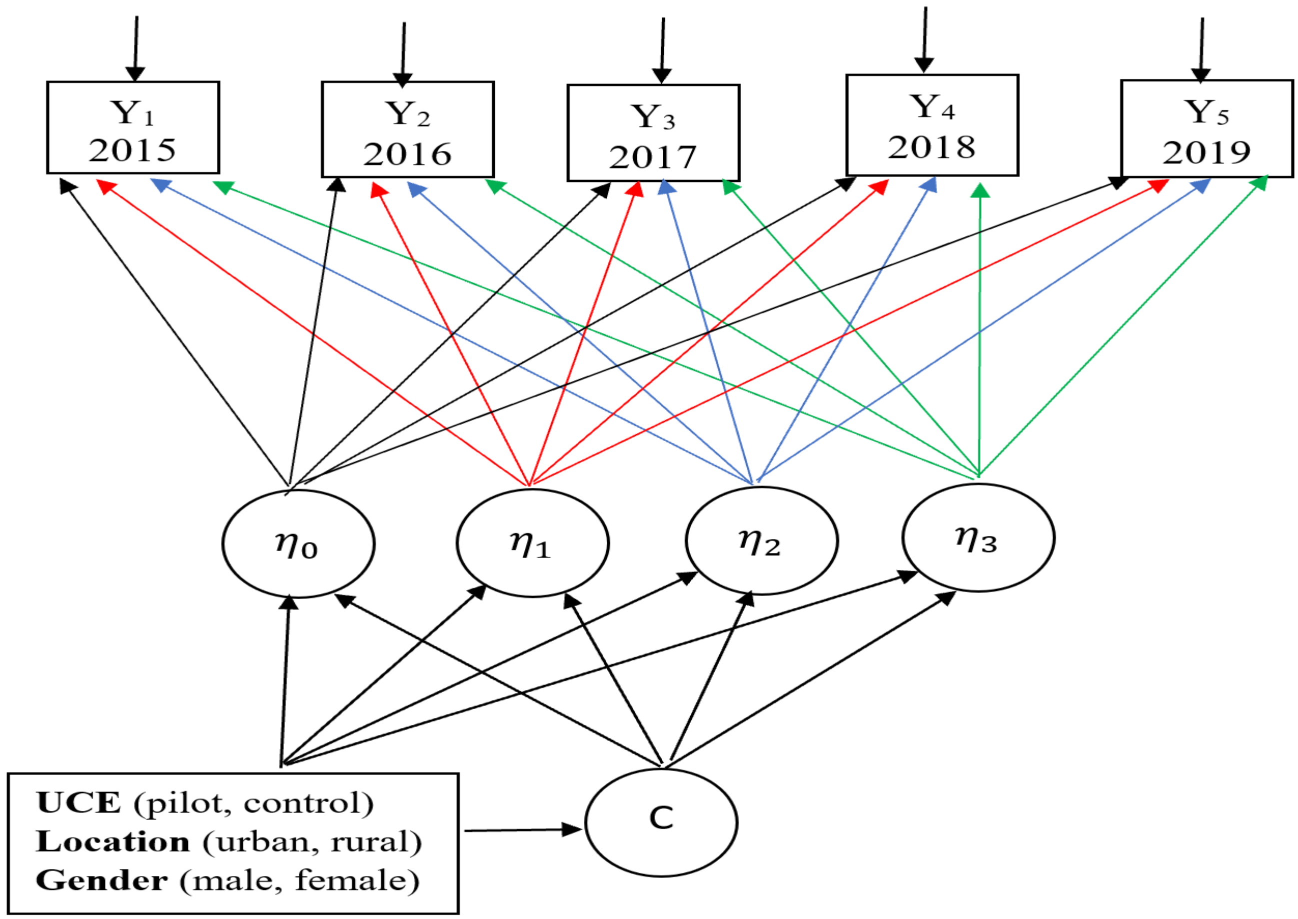


Figure 2. Path diagram of GMM with five longitudinal measures (Y1, Y2, Y3, Y4, Y5) obtained with five diagnostic tests (2015/2019), separately for Mathematics and Science.

Table 1Distribution of Learners Across UCE Conditions

Time of testing	Pilot	Control	Total
September 2015	3,420	1,811	5,231
April 2016	3,347	1,730	5,077
April 2017	3,390	1,633	5,023
April 2918	3,378	1,685	5,063
April 2019	3,301	1,675	4,976

Research Instruments and Procedures

The measures used to address the first research question, RQ1, consist of dichotomously scored items (1 = correct, 0 = incorrect) used at pretesting (September 2015) and immediate posttesting (April 2016) for (a) Mathematics: 14 items at pretest and 23 items at posttest, and (b) Science: 5 items at pretest and 16 items at posttest. The longitudinal measures used to address RQ2 consist of summative scores at each of the five diagnostic tests for each test section separately. Using test equating via common items (anchors) in the framework of item response theory (IRT), the summative scores across the five time points of diagnostic tests are presented on a common scale. Specifically, simultaneous calibration of test booklets with common items was performed using the one-parameter logistic model (OPLM; Verhelst, Glas, & Verstralen, 1995). The OPLM combines useful measurement properties of the Rasch model (Rasch. 1960) with the flexibility of the two-parameter logistic model in IRT where the discrimination power is allowed to vary across items (Birnbaum. 1968). The background variables used in this study are coded as follows: UCE (0 = control, 1 = pilot), gender (0 = male, 1 = female), and school location (0 = rural, 1 = urban).

Data Analysis

The data analysis under the research design in this study was conducted in the framework of structural equation modeling (SEM) using the computer program Mplus (Muthén & Muthén, 2009-2018). Related to RQ1, the data analysis involved testing for data fit of the CFA model in Figure 1 using the following goodness-of-fit indices: chi-square ($\chi2$) goodness-of-fit test, comparative fit index (CFI), standardized mean square residual (SRMR), and root mean square error of approximation (RMSEA) along with its 90 percent confidence interval (90%CI). Based on widely adopted criteria (e.g., Hu & Bentler, 1999), a tenable data fit of a SEM model isdemonstrated when the $\chi2$ statistic is not statistically significant, CFI > 0.90, SRMR < 0.08, and RMSEA < 0.08, with its 90%CI entirely below 0.08 (for excellent fit: CFI > 0.95, SRMR < 0.06, and RMSEA < 0.05, with its 90%CI entirely below 0.05). The $\chi2$ value may not necessarily count in decisions on data fit due to its high sensitivity to sample size and distribution assumptions (e.g., see Dimitrov, 2012).

Related to RQ2, the data analysis involved testing for the number of latent classes of participants, based on their longitudinal performance in Mathematics and Science, using three widely recommended statistics, the Akaike Information Criterion (AIC), the Bayesian Information Criterion adjusted for sample size (aBIC), and the Lo-Mendel-Rubin Adjusted Likelihood Ratio Test (aLRT; e.g., Tofghi & Enders, 2007; Nylund, Asparouhov, & Muthén, 2007). The testing for number of classes starts with a single-class model and gradually increasing the number of latent classes until a decision on the proper number of classes is reached. Under the AIC and aBIC criteria, smaller values indicate better fit. Under that aLRT criterion, the decision on proper number of latent classes is based on the p values associated with this statistic in the comparison of a model with (K – 1) classes versus a model with K classes. If the p value indicates statistical significance for a model with (K – 1) classes (p < 0.05) and the p value for the case of K classes indicates a lack of statistical significance (p> 0.05), the difference in data fit between the two models is negligible, so the more parsimonious model with (K – 1) classes is preferred.

Results

Immediate UCE effects (September 2015 – April 2016)

The estimation of immediate effects of the UCE under the design in Figure 1 involves (a)testing for data fit of the CFA model and (b) SEM-based estimation of the targeted UCE effects.

Based on the goodness-of-fit criteria described in the previous section, the results in Table 2 indicate a tenable model fit for the data in Mathematics and Science. Also, the standardized factor loadings of the items associated with the pretest and posttest measures (not provided here for space consideration) are statistically significant and sizable in magnitude (> 0.30).

Table 2Goodness-of-fit Indices Under the SEM Design for UCE Pretest-Posttest Effects (2015-2016)

Test section	χ^2	*df*	CFI	SRMR	RMSEA	RMSE: 90% CI LL	RMSE: 90% CI UL
Mathematics	3,045.188	660	0.924	0.068	0.032	0.031	0.033
Science	1,308.848	201	0.961	0.076	0.039	0.037	0.041

Note. LL = Lower limit; UL = Upper limit.

Estimates of the coefficients in the SEM model in Figure 1 are provided in Table 3. Forboth Mathematics and Science, the values of the path coefficient γ indicate statistically significant pretest-posttest effect of the UCE. Given the coding of UCE (0 = control, 1 = pilot), the positive sign of γ shows that the pilot learners exceed the control learners on posttest scores controlling for pretest differences between the two groups. The statistically significant coefficient β indicates that higher performance on the pretest (September 2015) is associated with higher performance on the posttest (April 2016) for the learners from both groups. The negative sign of the correlation coefficient *r* indicates that the control learners tend to have higher scores on the pretest than the pilot learners.

The *effect size* (*ES*) of the difference between the pilot and control groups on the posttest,controlling for their pretest differences, is given in Table 3. The *ES* is estimated by standardizing the absolute value of the regression coefficient γ in terms of the latent standard deviation of theresidual variance of the latent variable "posttest" in Mathematics and Science (Hancock, 2004).

Table 3 Unstandardized Coefficients for the UCE Pretest-Posttest Effect (2015-2016)

Test section	γ	β	*r*	VAR(ε)	*ES*
Mathematics	0.503^{**}	0.631^{**}	-0.072^{*}	0.351	0.849
Science	0.516^{**}	0.309^{**}	-0.254	0.194	1.172

$^{*}\ p < 0.05.\ ^{**}\ p < 0.001.\ ^{***}\ p < 0.001.$

Note. $ES = |\gamma| / \sqrt{\mathrm{VAR}(\varepsilon)}$; (Hancock, 2004).

Latent Classes of Longitudinal Growth Trajectories

Related to RQ2, the first step was to test for homogeneity of the population of learners interms of their longitudinal performance on the five diagnostic tests in Mathematics and Science. The question is whether the population is homogeneous or it breaks into latent (hidden) classes of learners with different trajectories of performance across the five assessments (2015-2019). An unconditional GMM was conducted first, without including covariates in the model, to decide how many latent classes of learners to retain using the criteria described in the previous section. The results in Table 4 show that a single latent class is retained for Mathematics and two latent classes for Science.

Thus, the study population of learners is homogeneous in longitudinal performance in Mathematica, whereas they group into two different latent classes of longitudinal performance in Science. As the UCE effect on the longitudinal performance in Mathematics and Science is of main interest, the growth trajectories in Figures 3 and 4, respectively, are provided by UCE (pilot/control) groups. The means and standard deviations of these trajectories acrossthe five diagnostic test scores in Mathematics and Science (2015/2019) are given in Appendix.

Table 4Number of Latent Classes of Learners with Different Growth Trajectories over Five Tests in Mathematics and Science (2015-2019)

Test Section/ Number latent classes	AIC	aBIC	aLRT
Mathematics			
One class*	136,431.323	136488.488	NA
Two classes	134,777.835	134880.130	0.2398 (*p*= 0.240)
Science			
One class	137,034.887	137077.009	NA
Two classes*	136,684.632	136738.789	347.639 ($p < 0.001$)
Three classes	136,432.806	136498.997	252.127 (*p*= 0.486)

Note. NA = Not applicable. * Number of classes retained.

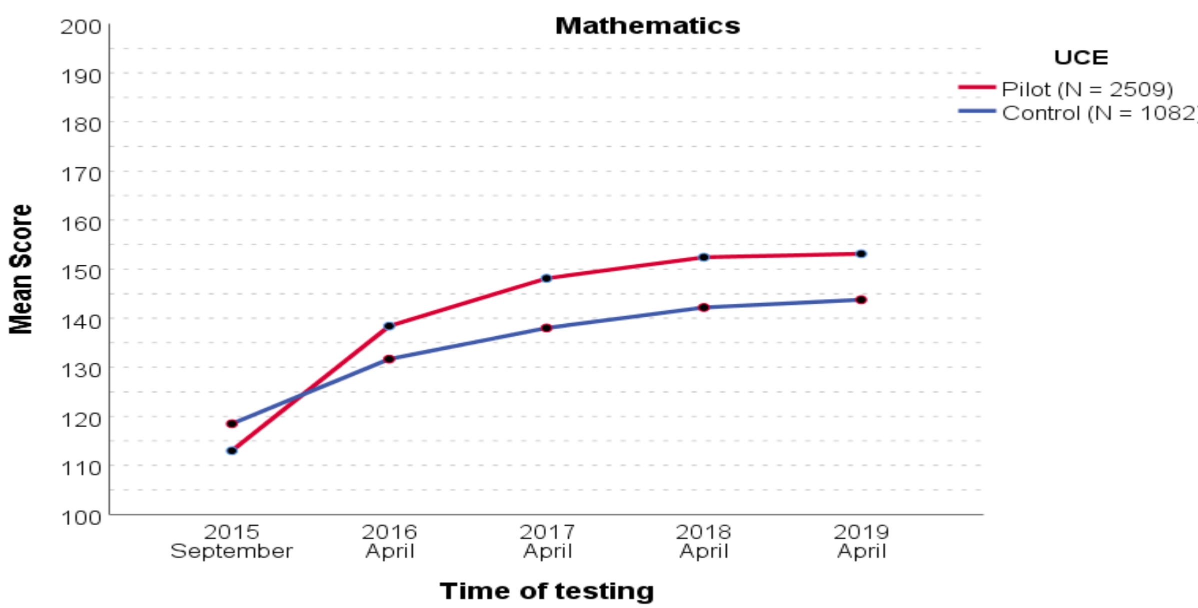


Figure 3. Growth trajectories in Mathematics by UCE condition (pilot/control).

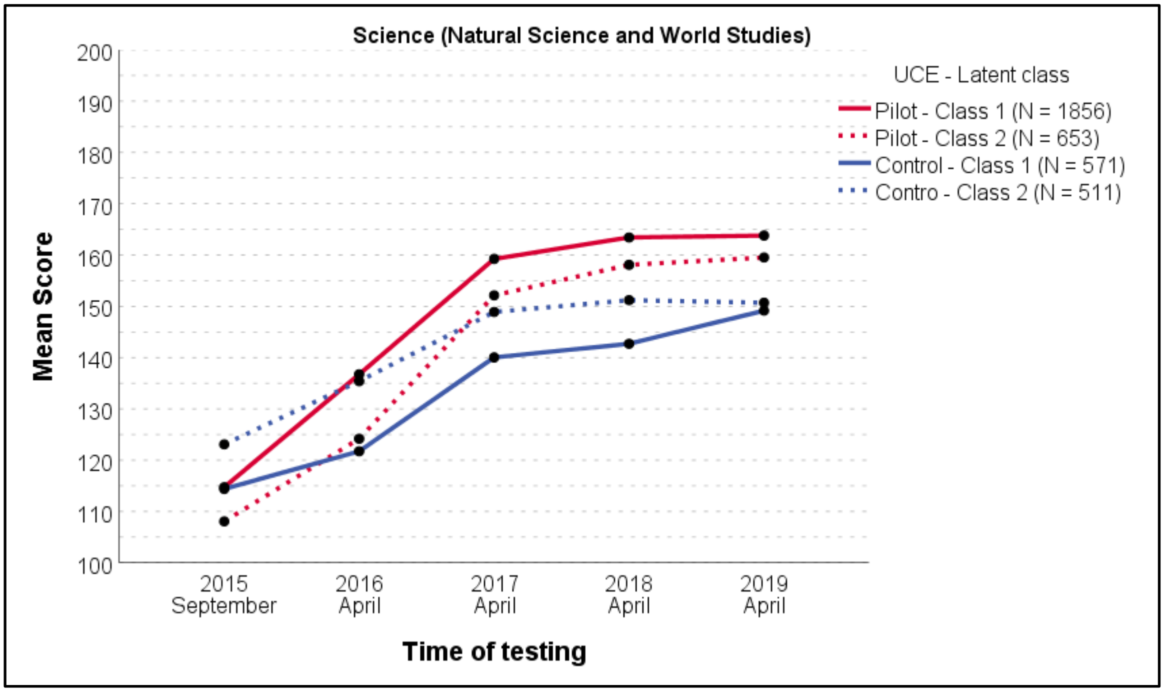


Figure 4. Growth trajectories in Science by latent classes and UCE condition (pilot/control).

Effects of UCE, Location, and Gender on Growth Trajectories

Related to RQ2, the effects of the UCE condition (pilot vs. control), gender, and schoollocation (urban vs. rural) are estimated in the framework of conditional GMM; that is, UCE, gender, and location are included as covariates in the model (Figure 2 and Equations 1-5). As described earlier, the GMM model includes nonlinear piecewise growth patterns with linear segments at three stages of longitudinal growth: stage 1 (2015-2016), stage 2 (2016-2018), and stage 3 (2018-2019). The results for Mathematics and Science are given in Tables 5. The initial GMM analysis showed that gender does not have any significant effect neither on the formation of latent classes, nor on the profiles of longitudinal growth trajectories. Therefore, reported in the following are only results on UCE and location effects.

Mathematics. The statistically significant negative intercept for UCE (-5.356) indicates that the pilot group of learners has lower start at the baseline performance (September 2015) than the control group. The slope of growth is statistically significant at all stages, but with a positive sign at stages 1 and 2 and a negative sign at stage 3. Thus, the pilot learners grow (a) much faster at stage 1 (2015-2016), (b) slightly faster at the middle stage 2 (2016-2018), and (c) at a slightly lower rate at stage 3, compared to the control learners, due to a trend of "ceiling effect" for the pilot group at this stage (April 2018 – April 2019).

Given the coding of location (0 = rural, 1 = urban), the statistically significant positiveintercept for location (2.836) indicates that the urban group of learners has higher start at the baseline performance (September 2015) compared to the rural group. The slope of growth at stage 2 is statistically significant and negative (-1.968) which indicates that the rural group of learners tend to grow faster at this stage, compared to the urban group. The slope values at stages 1 and 3 are not statistically significant.

The interaction term "intercept x slope" (I x S) indicates whether the rate of growth at agiven stage depends on the initial start in performance of the learners. For example, the statistically significant interaction term at stage 1 (-95.516) shows that learners with a lower start (September 2015) tend to grow much faster at stage 1 (September 2015-April 2016) compared to learners with higher start in performance on the Mathematics diagnostic test. At stage 2, the interaction term (-0.781) is not statistically significant, thus indicating that the learners' rate of growth at this stage does not depend on their initial performance. At stage 3, the interaction term is statistically significant and negative (-7.996); that is, the learners with a lower start tend to grow much faster in mathematics performance at stage 3 (2018-2019).

Table 5Growth Factors in Longitudinal Trajectories for Mathematics and Science

Test section/ Latent class (LC)/ Growth factors		Stage 1 2015-2016	Stage 2 2016-2018	Stage 3 2018-2019
MATHEMATICS				
Initial status	UCE	-5.356***		
(*I* = *intercept*)	Location	2.836***		
Rate of growth	UCE	12.738***	1.651***	-1.457**
(*S* = *slope*)	Location	-0.692	-1.968***	-0.867
I x *S*		-95.516***	-0.781	-7.996**
SCIENCE				
Class 1 (LC1): Higher level				
Initial status	UCE	0.147		
(*I* = *intercept*)	Location	2.949***		
Rate of growth	UCE	14.687***	2.854***	-5.919***
(*S* = *slope*)	Location	-2.066***	-1.296***	-2.172**
I x *S*		-95.732***	10.362***	-8.898***
Class 2 (LC2): Lower level				
Initial status	UCE	-14.054***		
(*I* = *intercept*)	Location	1.770*		
Rate of growth	UCE	4.631***	8.223***	1.767
(*S* = *slope*)	Location	0.615	-1.259**	-0.902
I x *S*		-95.732***	10.362***	-8.898***

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Science. The longitudinal growth trajectories in Science break into two latent classes(LCs), denoted here LC1 (higher level) and LC2 (lower level) (see Figure 4). Therefore, the effects of UCE and location, given in Table 5, are discussed in the following for each of these two latent classes.

For LC1, the results related to UCE imply several findings First, the nonsignificant intercept(0.147) shows that the pilot and control groups do not differ on the initial test scores (September 2015). Second, the statistically significant slopes at stages 1 and 2 (14.687 and 2.854) indicatethat the pilot group learners grow faster in performance, especially at the first stage (September 2015-April 2016). At stage 3, however, the control group learners tend to grow faster than the pilot group – see Figure 4 (lower panel) where the control group trajectory slightly increases, whereas the pilot group trajectory is almost "flat" at stage 3 (April 2018-April 2019).Regarding 'location,' the statistically significant intercept (2.949) shows that the urban group of learnersperformed higher than the control group on the start (September 2015). All slopes are statistically significant and negative in sign, thus indicating that the rural group learners grow faster than the urban group in all three stages of longitudinal trajectories.

For LC2,the results related to UCE imply several findings. First, the statistically significant intercept (-14.054) shows that the control group learners performed higher than the pilot group at the start (September 2015). The positive slopes at stages 1 and 2 are statistically significant, thus indicating that the pilot group learners tend to grow faster at these two stages. However, the slope at stage 3 (1.767) is not statistically significant, which indicates that the pilot and control groups do not differ in rate of growth at this stage (April 2018 – April 2019). Regarding 'location,' the intercept is statistically significant (1.770) thus indicating thatthe urban group learners performed slightly higher than the rural group on the initial test (September 2015). The slope at stage 2 (-1.259) is statistically significant which shows that the rural group learners tend to grow faster than the urban group at this stage. At stages 1 and 3, however, the slopes are not statistically significant, Thus, the rural and urban groups do not differ on rate of growth at the first and third stages of longitudinal performance.

For both LC1 and LC2, , the term (I x S) in Table 5 is statistically significant and negative at stages 1 and 3 (-95.732 and -8.898, respectively) and positive at stage 2 (10.362). This indicates that learners with a lower start tendto grow much faster at stages 1 and 3 and, conversely, they grow at a lower rate at stage 2 (April 2016 - April 2018).

Discussion

The overall findings that stem from the results in this study have several main aspects.

First, based on the shape of the growth trajectories, the effects of the UCE (pilot/control), gender, and school location (urban/rural) were investigated at three stages of learners' performance over the time period of five diagnostic tests, (a) stage 1 (2015 – April 2016), (b) stage 2 (April 2016 – April 2018), and (c) stage 3 (April 2018 – April 2019). As *no* gender differences were found in level of performance and rate of growth across the three stages for all sections, summarized in the following are effects of the UCE and school location.

Second, the longitudinal growth trajectories of learners' performance over time points of diagnostic assessments (2015/2019) are homogeneous for Mathematics, but break into two latent classes for Science (Natural science & World studies). Third, the UCE has positive impact on the longitudinal performance of the learners in the sense that the pilot learners tend to grow faster and perform at higher level in both Mathematics and Science. This trend is particularly pronounced for immediate pretest-posttest effects of the UCE, where the pilot group learners outperform the control group on the posttest measures (April 2016) controlling for the pretest differences between the two groups at the pretest (September 2015).

Fourth, regarding the role of school location (rural/urban), the overall effect on rate for growth is in favor of the learners from rural schools. However, some differential results by test sections and stages of growth in performance are worth noting. In Mathematics, there is *no* school location effect on the level of performance and rate of growth with the exception of stage 2 (April 2016 – April 2018), where the learners from rural schools performed at lower level, yetgrew faster, in performance compared to those from urban schools. In Science, (a) the learners from rural schools in the higher-performing latent class grew faster in performance than those from urban schools at all three stages, and (b) the same trend was shown at stage 2, but there were no differences in rate of growth between the learners from urban and rural schools in the lower-performing latent class at the first and third stage of development.

Implications for the Educational Practice and Policies

The practical implications of the study findings are discussed here under the logic thatthey support the approbation of the UCE which, in turn, has direct implications for the educational practice and policies in Kazakhstan. The findings of the UCE approbation, stemming from both quantitative analyses and methodological monitoring were widely discussed among all levels of stakeholders in Kazakhstan which led to practical implementations of the UCE in all secondary schools (more than 7,000 schools) in the country and policies of further developments in education for the next five years (2020-2025). The teachers from UCE pilot schools were provided with professional development courses and targeted methodological support via conferences, seminars, webinars, workshops, reflection on the basis of observations of lessons, and so forth. The systematic, targeted, and individualized methodological support to UCE teachers, along with the updated curriculum and relevant resources, contributed to achieving substantial and sustainable UCE effects on the learners' performance in Mathematics and Science. Furthermore, the UCE pilot schools have become supporting sites for methodological support on implementing the UCE in other schools in the respective regions of the country.

The annual monitoring and diagnostic testing during the UCE led to the followingactivities and changes in the content of the curriculum and assessments:

- Some learning objectives were amended;
- The logical sequence in learning Mathematics and Science was revised;
- The intra-subject integration and profiling, as well as in-depth systematic training on cross-curricula topics in humanitarian subjects, were strengthened;
- The recommendations on formative and summative assessment procedures were modified;
- Video-presentation materials on criteria-based assessment and related support to teacherswere prepared;
- Topics of higher difficulties for learners were identified as follows:

  – in Mathematics: (a) geometric shapes and their classifications, used in the formation of skills and ability to apply knowledge in everyday life situations, and (b) tasks and mathematical model, related to the ability of explaining, solving, and using arithmetic algorithms;

– in Science: (a) plants and animals, aimed at the formation of classification skills,and (b) water, air, and natural resources, aimed at the formation of ability to apply knowledge in performing integrated tasks.

Recommendation for Future Research and Implementation

It is necessary to continue the development of teachers' professional competencies, including the development of teaching aids and didactic materials, video lessons on learning objectives, workshops on primary school subjects, and criteria-based assessments. Based on the research findings on difficulties of learners in particular topics, the practitioners can be provided with the following recommendations:

● Mathematics: There is a need of implementing more practical work into lessons onmodeling of task description in the form of tables/ diagrams/ notes; writing the plan of task solution; dividing complex tasks into simple parts (analysis); gathering simple parts of the task to make a complex one (synthesis).

● Science: There is a need of analyzing information given in the form of pictures, drawings,et., and using integrated topics of cross-curricula units/subjects. Learners from different latent classes of performance need to be identified and provided with support in differential learning strategies.

Conclusion

The approbation of the updated content of secondary education contributed to itsvalidation in the educational practice of Kazakhstan. The results from the four-year diagnostic testing (2015-2019) made it possible to identify difficulties and need for methodological support to teachers and further prospects for the use of curricula in real educational context. Although the study is conducted in the educational context of Kazakhstan, the UCE methodology and research design presented in this paper can be useful to researchers, educators, and policy makers in other countries.

Appendix

Means and Standard Deviations of Longitudinal Growth Trajectories by Conditions UCE (pilot/control), Latent Class (when exist), and School Location (urban/rural)

Table A-1. *Mathematics*

Time/UCE	Location	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*
2015. September							
Pilot	Urban	113.96	11.582	1691	112.99	11.140	2509
	Rural	110.99	9.876	818			
Control	Urban	119.18	13.332	784	118.49	12.709	1082
	Rural	116.68	10.713	298			
2016. April							
Pilot	Urban	138.91	10.021	1691	138.40	10.001	2509
	Rural	137.34	9.883	818			
Control	Urban	133.88	11.544	784	131.67	11.725	1082
	Rural	125.85	10.104	298			
2017. April							
Pilot	Urban	147.02	11.248	1691	148.12	11.705	2509
	Rural	150.41	12.293	818			
Control	Urban	138.13	13.886	784	138.00	12.821	1082
	Rural	137.68	9.476	298			
2018. April							
Pilot	Urban	152.15	10.657	1691	152.42	10.983	2509
	Rural	152.98	11.614	818			
Control	Urban	142.24	12.846	784	142.19	12.056	1082
	Rural	142.06	9.692	298			
2019. April							
Pilot	Urban	152.30	11.971	1691	153.12	11.460	2509
	Rural	154.82	10.121	818			
Control	Urban	142.93	11.451	784	143.76	11.452	1082
	Rural	145.93	11.186	298			

Table A-2. *Science*

Latent class /UCE	Location	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*
September, 2015							
Class 1	Pilot	114.73	10.915	1856			
	Control	114.39	12.258	571	114.65	11.244	2427
Class 2	Pilot	108.06	10.268	653			
	Control	123.07	11.602	511	114.65	13.180	1164
April, 2016							
Class 1	Pilot	136.74	3.450	1856			
	Control	121.73	5.560	571	133.21	7.545	2427
Class 2	Pilot	124.18	4.027	653			
	Control	135.40	4.167	511	129.10	6.909	1164
April, 2017							
Class 1	Pilot	159.23	11.811	1856			
	Control	140.03	10.191	571	154.72	14.052	2427
Class 2	Pilot	152.13	11.567	653			
	Control	148.90	13.217	511	50.71	12.417	1164
April, 2018							
Class 1	Pilot	163.42	11.967	1856			
	Control	142.70	10.542	571	158.55	14.593	2427
Class 2	Pilot	158.09	11.015	653			
	Control	151.19	11.770	511	155.06	11.853	1164
April, 2019							
Class 1	Pilot	163.79	12.229	1856			
	Control	149.15	11.719	571	160.35	13.609	2427
Class 2	Pilot	159.51	13.018	653			
	Control	150.69	9.589	511	155.64	12.431	1164

References

Birnbaum, A. (1968). Some latent trait models (Chapter 17). In F.M. Lord & M.R. Novick.*Statistical theories of mental test scores*. Reading (Mass.): Addison-Wesley.

Dimitrov, D. M. (2012).*Statistical methods for validation of assessment scale data in counseling and related fields*.Alexandria, VA: American Counseling Association.

Dimitrov, D. M. (2009-2013) *Quantitative research in education: Intermediate & advanced methods*. Ocean City, NY: Whittier Publication Inc.

Hancock, G. R. (2004). Experimental, quasi-experimental, and nonexperimental design andanalysis with latent variables. In D. Kaplan (Ed.). *The SAGE handbook of quantitative methodology for the social sciences* (pp. 317-334). Thousand Oaks, CA: SAGE.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structuresanalysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Jöreskog, K. G., & Sörbom, D. (1976). Statistical models and methods for test-retest situations.

In D.N.M. DeGruijter and L.J.Th. Van der Kamp (Eds). *Advances in psychological andeducational measurement*(pp. 135-157). John Wiley & Sons, NY.

Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E.Schumacher (Eds.). *New developments and techniques in structural equation modeling* (pp. 1-33). Mahwah, NJ: Lawrence Erlbaum.

Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniquesfor longitudinal data. In *Handbook of Quantitative methodology for the social sciences*(ed. D. Kaplan). Sage, Newbury Park, CA, pp. 345-368.

Muthén, L. K., & Muthén, B. O. (2009-2018). *Mplus user's guide* (version 8.3). Los Angeles,CA: Author.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes inlatent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 536-569.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen:Danish Institute for Educational Research.

Tofghi, D., & Enders, C. K. (2007). Identifying the correct number of classes in mixture models.In G. R. Hancock & K. M. Samuelsen (Eds.). *Advances in latent variable mixture models* (pp. 317-341). Greenwich, CT: Information Age.

Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. WestSussex, United Kingdom: Wiley.