The Impact of STEM Integration on Secondary School Students' Conceptual Understanding in Physics

Aruzhan Tursumbekova

A thesis submitted to the School of Education and Humanities
in partial fulfillment of the requirements for the degree of

MASTER OF EDUCATIONAL SCIENCES
in 7M01503 - Physics

«SDU University»
Department of Pedagogy of Natural Sciences

Thesis Supervisor:
PhD, Assistant Professor
Nursultan Japashov

©Copyright by Aruzhan Tursumbekova, 2026

«SDU University»
School of Education and Humanities
Department of Natural Sciences Pedagogy

This is to certify that the Master's Thesis of
Aruzhan Tursumbekova
has met the thesis requirements of SDU University.

Approved by

Thesis Supervisor	PhD, Assistant Professor, Nursultan Japashov
M.Ed. Program Coordinator	MSc, Ayatola Gabdulin
Head of Department of Natural Science Pedagogy	PhD, Associate Professor, Zhangyl Abilbek
Dean of School	PhD, Assistant Professor, Dana Moldabayeva

Kaskelen, 2026

«SDU University»
School of Education and Humanities
Department of Natural Sciences Pedagogy

This is to certify that the Master's Thesis of
Aruzhan Tursumbekova
has met the thesis requirements of SDU University.

Approved by

Thesis Supervisor	PhD, Assistant Professor, Nursultan Japashov
M.Ed. Program Coordinator	MSc, Ayatola Gabdulin
Head of Department of Natural Science Pedagogy	PhD, Associate Professor, Zhangyl Abilbek
Dean of School	PhD, Assistant Professor, Dana Moldabayeva

Kaskelen, 2026

List of Tables

Table 3.2 Six-Week STEM-Integrated Newton's Laws Lesson Plan Overview 51
Table 3.3 Assessment Types Used Across the Six-Week STEM Lesson Plan 52
Table 4.1. Demographic Composition of the Experimental and Control Groups 60
Table 4.2 Independent Samples T-Test: Pre-Test and Post-Test FCI Total Score Comparison (Experimental vs. Control) 62
Table 4.3 Homogeneity of Variances Test (Levene's): Pre-Test and Post-Test FCI Total Score 63
Table 4.4 Group Descriptives: Pre-Test and Post-Test FCI Total Score by Group ... 64
Table 4.5 ANCOVA — Post-Test FCI Total Score: Effects of Group, Gender, and Pre-Test Covariate 66
Table 4.6 ANCOVA Assumption Check: Homogeneity of Variances Test (Levene's) Across Group-Gender Cells 68
Table 4.7 FCI Subscale Domain Analysis: Mann-Whitney U Tests at Pre-Test and Post-Test (Experimental vs. Control) 70

List of Figures

Figure 4.1 Mean and Standard Deviation of Pre-Test and Post-Test FCI Total Scores for Experimental and Control Groups 64

Figure 4.2 Estimated Marginal Means of Post-Test FCI Total Score by Gender and Group (ANCOVA) 69

List of Appendices

Appendix A: Three tier FCI 87

Table of Contents

Abstract x

Абстракт xi

Аннотация xii

1. Introduction 12

1.1 Theoretical Background 12

1.1.1 Theoretical Framework 13

1.2 Statement of the Problem 15

1.3 Purpose of the Study 17

1.4 Research Questions and Hypotheses 18

1.5 Significance of the Study 20

1.6 Limitations of the Study 21

1.7 Conclusion 22

2. Literature Review 24

2.1 Conceptual Understanding in Physics Education 25

2.2 Limitations of Traditional Physics Instruction 26

2.3 Studies on the Effectiveness of Integrated STEM Education 28

2.4 STEM-Integrated Instruction in Physics Education 30

2.5 STEM and Conceptual Understanding of Newtonian Mechanics 32

2.6 Teacher Professional Development for STEM Integration in Kazakhstan 35

2.7 Gender Differences in STEM Learning 37

2.8 Research Gap and Rationale for the Study 38

2.9 Conclusion 40

3. Methodology 42

3.1 Research Design 42

3.2 Participants of the Study 42

3.3 Instrument of Data Collection 43

3.3.1 The Force Concept Inventory (FCI) 43

3.3.2 Scoring and Response Categorisation 44

3.3.3 Normalised Gain 45

3.4 Procedure 45

3.4.1 Pre-test Administration 45

3.4.2 Experimental Condition — STEM-Integrated Instruction 45
3.4.3 Control Condition - Traditional Instruction 46
3.4.4 Post-test Administration 46
3.5 Ethical Considerations 47
3.6 Assumptions and Delimitations 47
3.7 Data Analysis Plan 49
3.8 Six-Week STEM-Integrated Lesson Plan 50
Lesson Structure 52
Week 1 — Introduction to Forces and Motion 53
Week 2 — Newton's First Law (Inertia) 54
Week 3 — Newton's Second Law (F = ma) 55
Week 4 — Newton's Third Law (Action-Reaction) 55
Week 5 — Application of Newton's Laws 56
Week 6 — Motion Graphs and Data Analysis 57
Materials and Resources 58
4. DATA ANALYSIS AND RESULTS 60
4.1 Description of Participants 60
4.2 Data Analysis Strategies 61
4.3 Post-Test Comparison of Groups 61
4.4 Main Analysis: Effect of STEM Instruction (ANCOVA) 66
4.5 Gender Differences 68
4.6 Visual Analysis of Results 68
4.7 Mann-Whitney U Subscale Analysis by Domain 69
5. DISCUSSION 73
5.1 Discussion of RQ1: Effect of STEM-Integrated Instruction on Conceptual .. 73
5.2 Discussion of RQ2: Gender Differences 74
5.3 Discussion of RQ3: Domain-Specific Conceptual Gains 74
5.4 Theoretical Implications 75
5.5 Practical Implications for Kazakhstan 76
5.6 Recommendations for Future Research 76
6. LIMITATIONS 78
7. CONCLUSION 80
8. REFERENCES 81

9.APPENDIX 87

List of Abbreviations

Abbreviation	Meaning
ANCOVA	Analysis of Covariance
FCI	Force Concept Inventory
H_0	Null Hypothesis
H_1	Alternative Hypothesis
MoSHE RK	Ministry of Science and Higher Education of the Republic of Kazakhstan
NCPD	National Centre for Professional Development (Orleu)
OECD	Organisation for Economic Co-operation and Development
PD	Professional Development
PISA	Programme for International Student Assessment
RQ	Research Question
SD	Standard Deviation
STEM	Science, Technology, Engineering, and Mathematics
TIMSS	Trends in International Mathematics and Science Study
UNT	Unified National Test (Kazakhstan university entrance examination)

Abstract

This study examines the effect of STEM-integrated instruction on 9th-grade students' conceptual understanding of Newtonian mechanics in Kazakhstan. Using a quasi-experimental pretest-posttest design with 74 students at Al-Farabi Specialized Lyceum, the experimental group (n = 34) received six weeks of STEM-integrated instruction while the control group (n = 40) followed traditional teaching. Conceptual understanding was measured via the Force Concept Inventory (FCI). Results showed that the STEM group significantly outperformed the control group, achieving a medium normalised gain compared to a low gain for controls. A pre-test gender gap closed entirely by post-test, confirming equitable outcomes. The study provides the first controlled evidence from Kazakhstani secondary schools that STEM integration substantially outperforms traditional physics instruction.

Key words: STEM integration, STEM-integrated instruction, conceptual understanding, Newtonian mechanics, physics education, Force Concept Inventory (FCI), quasi-experimental design, normalised gain, secondary school, Kazakhstan, gender equity, traditional instruction, Al-Farabi Specialized Lyceum.

Абстракт

Бұл зерттеу STEM интеграцияланған оқытудың Қазақстандағы 9-сынып оқушыларының Ньютон механикасын тұжырымдамалық түсінуіне әсерін зерттейді. Әл-Фараби атындағы мамандандырылған лицейде 74 оқушы қатысқан квазиэкспериментальды тестілеуден кейінгі дизайнды қолдана отырып, эксперименталды топ (n = 34) алты апта STEM интеграцияланған оқытудан өтті, ал бақылау тобы (n = 40) дәстүрлі оқытуды ұстанды. Тұжырымдамалық түсінік Күш тұжырымдамасын түгендеу (FCI) арқылы өлшенді. Нәтижелер STEM тобы бақылау тобынан айтарлықтай асып түсіп, бақылау тобы үшін төмен өсіммен салыстырғанда орташа қалыпқа келтірілген өсімге қол жеткізгенін көрсетті. Тестілеу алдындағы гендерлік алшақтық тестілеуден кейінгі кезеңмен толығымен жабылды, бұл тең нәтижелерді растады. Зерттеу қазақстандық орта мектептерден STEM интеграциясының дәстүрлі физиканы оқытудан айтарлықтай асып түсетіні туралы алғашқы бақыланатын дәлелдерді ұсынады.

Кілт сөздер: STEM интеграциясы, STEM-интеграцияланған оқыту, тұжырымдамалық түсінік, Ньютон механикасы, физика білімі, күш тұжырымдамасын түгендеу (FCI), квазиэкспериментік жобалау, нормаланған күшейту, орта мектеп, Қазақстан, гендерлік теңдік, дәстүрлі оқыту, Әл-Фараби атындағы мамандандырылған лицей.

Аннотация

В данном исследовании изучается влияние интегрированного STEM-образования на концептуальное понимание ньютоновской механики учащимися 9-го класса в Казахстане. Используя квазиэкспериментальный дизайн с предварительным и последующим тестированием, в исследовании приняли участие 74 ученика специализированного лицея им. Аль-Фараби. Экспериментальная группа (n = 34) в течение шести недель получала интегрированное STEM-образование, в то время как контрольная группа (n = 40) обучалась по традиционной методике. Концептуальное понимание измерялось с помощью Инвентаря концепций силы (FCI). Результаты показали, что группа, получавшая STEM-образование, значительно превзошла контрольную группу, достигнув среднего нормализованного прироста по сравнению с низким приростом у контрольной группы. Гендерный разрыв, существовавший до тестирования, полностью исчез к моменту завершения тестирования, что подтверждает равные результаты. Данное исследование предоставляет первые контролируемые данные из средних школ Казахстана, свидетельствующие о том, что интеграция STEM-образования существенно превосходит традиционное преподавание физики.

Ключевые слова: Интеграция STEM, интегрированное обучение STEM, концептуальное понимание, ньютоновская механика, преподавание физики, тест на определение силы (FCI), квазиэкспериментальное проектирование, нормализованный коэффициент усиления, средняя школа, Казахстан, гендерное равенство, традиционное обучение, специализированный лицей Аль-Фараби.

1. Introduction

1.1 Theoretical Background

Physics education research has found that there is a long-standing gap between student ability and understanding. A student may be able to reproduce a formula, apply that formula to a standard problem, and yet may not be able to explain the physics that is involved in the formula (Hestenes et al., 1992; Redish, 2003). This gap can be attributed to the persistence of scientific misconceptions in students' thinking, and the dominance of procedural fluency over conceptual understanding in science learning and instruction.

In Kazakhstan, these challenges are compounded by systemic factors. Secondary school physics instruction is predominantly teacher-centred, examination-driven, and algorithmically structured — a direct consequence of the Unified National Testing (UNT) system, which rewards procedural recall over conceptual reasoning (Nazarbayev Intellectual Schools, 2019). Research conducted locally confirms the outcome: Japashov et al. (2024) found that only 14.7% of university physics students in Kazakhstan demonstrated genuine scientific knowledge on the Force Concept Inventory (FCI), while 48.15% held firmly entrenched misconceptions. These misconceptions did not decline across academic year levels, indicating that the problem originates in secondary school and is sustained — not corrected — by traditional instruction.

STEM-integrated instruction has emerged globally as a pedagogical alternative designed to address precisely these conditions. By connecting Science, Technology, Engineering, and Mathematics through authentic, problem-centred learning experiences, STEM integration creates the cognitive conditions necessary for genuine conceptual change (Kelley & Knowles, 2016; Thibaut et al., 2018). This dissertation is theoretically grounded in three interconnected frameworks that explain how and why STEM integration can improve conceptual understanding in physics.

1.1.1 Theoretical Framework

The theoretical foundation of this study rests on three pillars: the theory of conceptual change, the integrated STEM framework, and the theory of meaningful learning. Together, these frameworks explain the mechanism by which STEM instruction is expected to produce superior conceptual outcomes compared to traditional teaching.

At the core of such an explanation is the Conceptual Change Theory (Posner et al., 1982): preconceived concepts, including strongly held misconceptions, can only be changed under four conditions, the first of which is that learners must be dissatisfied with their previously held conception because it cannot explain something that they observe. Second, the new conception is intelligible. Third, it is plausible. Fourth, it is productive in the sense of enabling the learner to think of new ideas. Customary, formula-based teaching does not meet any of these criteria, because the students do not find their ideas to be wrong or problematic. In STEM-integrated instruction, by contrast, students are put in a laboratory, in an engineering design task, or asked to collect data, and their intuitions are tested and found to be inadequate. Cognitive dissonance is created in the sense described by Posner et al. as necessary for conceptual change to occur.

Based on the literature from Kelley and Knowles (2016) and a systematic review of 23 secondary studies by Thibaut et al. (2018), the researchers determined five principles of effective STEM education: integration and connections across STEM disciplines, problem-based learning in context, inquiry-based learning, design-based learning, and cooperative learning. All five principles were implemented during the six stages of the lesson cycle (Engagement, Exploration, Explanation, Elaboration, Evaluation, Extension) used in the experimental group in this study. Along with coherence, authenticity, and disciplinary depth, Roehrig et al. (2021) argue that superficial interdisciplinary connections between subjects are not sufficient and real integration is required for meaningful student learning.

The Meaningful Learning Theory (Ausubel, 1968) principle of anchored instruction states that new knowledge is durably learned when it is anchored to existing cognitive structures through meaningful, contextualized engagement. An example is STEM integration, with a balloon-powered vehicle making Newton's Third Law concrete, and acceleration data using a smartphone making Newton's Second Law F = ma a derived result rather than a transmitted formula. These experiences create the assimilative conditions Ausubel deems necessary for retention and transfer.

The study is based on the Constructivist Learning Theory of Lev S. Vygotsky (1978) and Jean Piaget (1964), which states that knowledge is not transferred from one individual to another. In contrast, learners represent, reflect, and construct their knowledge based on their experiences as they act on and interact with the physical and social environment. This is particularly relevant to STEM education, as inquiry and hands-on experiments are how students construct and internalize conceptual knowledge.

Piaget (1964) proposed that cognitive development progresses through a process of adaptation via two complementary processes. One process is assimilation, which involves incorporating new experiences into existing cognitive schemas. The second process is accommodation, a process involving changes in existing cognitive schemas in response to new experiences or the creation of new ones. Active experimentation in the physical environment is central to this process. Most importantly, Piaget argued that children learn when they are active, rather than passive learners. In the learning sciences and STEM fields, the principle has been implemented through laboratory investigations, design challenges and open-ended problems in which students work with materials, tools and phenomena to challenge and define their understanding of the world.

Complementing Piaget's developmental perspective, Vygotsky (1978) emphasized the fundamentally social nature of learning and introduced the concept of the Zone of Proximal Development (ZPD). The ZPD can be considered as the cognitive development area that lies between the level of independent performance and the level

of potential performance, which is achieved through guidance or collaboration with a more skilled peer or adult. In this sense, the ZPD advocates for collaborative dialogue, scaffolding, and co-construction. The social structures of inquiry-based learning environments in STEM are group work, peer-assisted inquiry, experimentation under teacher guidance, that can be viewed as activating the ZPDs of learners who are not yet capable on their own.

Together, the work of Piaget (1964) and Vygotsky (1978) provides the theoretical foundations for understanding the relationship between STEM activities and knowledge building. Piaget's focus on the cognitive reorganization that occurs when an individual learner manipulates hands-on materials is complementary to Vygotsky's focus on the role of collaboration and mediation between the learner and others in the social environment in the learner's cognitive development. Bringing these two perspectives together, high-quality STEM learning experiences engage learners in physical inquiry and provide a social context and structure for collaborative inquiry-oriented meaning-making which enables learners to make sense of their exploration together with others.

Thus, the theoretical background for the design and analysis of this study is Constructivist Learning Theory, as articulated by Vygotsky (1978) and Piaget (1964). This theory supports the view that STEM experiences are epistemological experiences, through which learners construct, negotiate, and validate knowledge through represented, social, and participatory engagements with the world and the material, verbal, and semiotic resources therein. This study extends these theoretical approaches to investigate how organized forms of STEM learning might realize productive learning outcomes, especially the ways inquiry, experimentation, and collaborative dialogue can foster conceptual growth.

1.2 Statement of the Problem

The central problem addressed in this dissertation is the persistent failure of traditional physics instruction in Kazakhstani secondary schools to develop students' conceptual understanding of Newtonian mechanics. This failure is not incidental — it

is structural. The Kazakhstani secondary physics curriculum, as currently implemented, is organised around procedural knowledge delivery: teachers introduce laws through formulas, demonstrate worked examples, and assign standardised practice problems. This approach efficiently produces students who can perform examination tasks but leaves their intuitive, pre-scientific beliefs about the physical world entirely intact.

The depth of this problem has been empirically documented at the local level. Japashov et al. (2024) applied a three-tier FCI to 180 undergraduates across two Kazakhstani universities and found that firmly held misconceptions — incorrect beliefs expressed with high confidence — constituted the most common response pattern, regardless of year level. Students who had spent four years studying university physics still believed, with confidence, that a moving object requires a continuous force to sustain its motion. The secondary school origin of these beliefs is unmistakable: they are not corrected at university because they were never addressed at secondary school.

International large-scale assessments corroborate this picture at the system level. The PISA 2022 assessment — the most comprehensive international benchmark of 15-year-old students' ability to apply knowledge to real-world problems — reveals a striking science proficiency deficit in Kazakhstan. Only 55% of Kazakhstani students attained Level 2 or higher in science, compared to the OECD average of 76% (OECD, 2023). This gap of 21 percentage points is academically significant: PISA Level 2 is not an advanced benchmark — it represents the minimum baseline at which a student can recognise a correct scientific explanation of a familiar phenomenon, apply basic conceptual knowledge to draw a simple conclusion, and identify whether a conclusion follows from data. The inability of 45% of Kazakhstani 15-year-olds to meet even this baseline indicates that conceptual reasoning deficits are not limited to physics specifically but are systemic across the secondary science curriculum.

Including the whole PISA 2022 science achievement distribution: the 2022 proportion of Kazakhstani students scoring levels 5 or 6 is 2% compared to 7% for OECD average (OECD, 2023), suggesting that not only the tail but the entire

distribution is compressed relative to the international standard. Evidence from the local FCI (Japashov et al., 2024) and the foreign PISA data (OECD, 2023) converge that Kazakhstani secondary science education fails to prepare students conceptually but instead makes them procedural experts.

The problem is also an equity issue, as work indicates that male and female students enter secondary physics classrooms with different levels of conceptual understanding that are preserved or exacerbated by customary instruction. In the Kazakhstani context, additional questions are whether STEM-integrated physics instruction might produce different results in male and female students, or reduce the gender gap in physics conceptual understanding. These questions remain empirical questions.

STEM integration has been an educational priority at the national level in Kazakhstan (Ministry of Education and Science of Kazakhstan, 2016), and has been shown to have desirable outcomes in international studies (e.g., Francisca et al., 2021; Asrizal et al., 2023), but has not yet been subject to methodological scrutiny in a secondary school in Kazakhstan. This is problematic because teachers and curriculum experts cannot accept STEM integration without question if evidence is unavailable that it is appropriate in locally representative contexts. This study helps to address that.

1.3 Purpose of the Study

The primary purpose of this study is to investigate whether STEM-integrated instruction produces significantly greater gains in 9th-grade students' conceptual understanding of Newtonian mechanics compared to traditional instruction, and to examine the role of gender in physics conceptual understanding within this instructional context.

The experimental and control groups were formed on the basis of the State-Owned Institution "Specialized Lyceum named after Al-Farabi" of the State Institution "Education Department of Almaty Region", Karasai District, Kazakhstan for the first quarter of the 2025-2026 academic year. The sample consisted of four 9th-grade classes, two of which were taught STEM integrated lessons (n = 34, experimental

group) and two were taught typical lessons (n = 40, control group). A pre-test/post-test quasi-experimental design was used, and a 30-point FCI-based conceptual test was the primary assessment tool.

Along with addressing the main research question, the study proposed to explore the topic-level reasoning (the reasoning in each of the six domains of Newtonian mechanics assessed by the FCI, namely: kinematics; Newton's First Law; Newton's Second Law; Newton's Third Law; superposition of forces; and kinds of force) affected most by STEM instruction. This would help physics teachers plan STEM instruction for topics where it would most benefit students, whilst remaining within the constraints of the national curriculum.

1.4 Research Questions and Hypotheses

The overarching aim of this study is to investigate the impact of STEM-integrated instruction on secondary school students' conceptual understanding of Newtonian mechanics, and to explore the role of gender in physics conceptual understanding within this instructional context. This aim is pursued through the following research questions:

Research Question 1 (RQ1): Does STEM-integrated instruction produce significantly greater gains in 9th-grade students' conceptual understanding of Newton's Laws of Motion compared to traditional instruction at Al-Farabi Specialized Lyceum, Almaty, Kazakhstan?

Research Question 2 (RQ2): Is there a difference in conceptual understanding of physics between male and female students, and does this difference change following STEM-integrated instruction?

Research Question 3 (RQ3): Which specific domains of Newtonian mechanics — kinematics, Newton's First Law, Newton's Second Law, Newton's Third Law, superposition of forces, and kinds of force — show the greatest conceptual gains following STEM-integrated instruction compared to traditional instruction?

RQ1 is the main research question and relates to the main experimental hypothesis; RQ2 is an exploratory research question. Although the experiment does

not call for any specific gender gap, there needs to be exploration of whether there is a gender gap in physics conceptual understanding at the baseline assessment, and whether this gap changes with the use of STEM-integrated curricula. Since prior research on the Kazakhstani secondary school system does not provide guidance in predicting the direction of gender differences, RQ3 identifies which specific topic areas of Newtonian mechanics learning gains are greatest and smallest in response to STEM curricula.

In experimental and quasi-experimental research, hypotheses provide a formal statement of the expected relationship between variables, enabling inferential testing of the research claim. A pair of hypotheses — the null hypothesis (H_0) and the alternative hypothesis (H_1) — is stated for the primary research question (RQ1), which has a clear theoretical basis and a well-established expected direction of effect based on prior empirical literature. RQ2 and RQ3 are treated as exploratory questions and are not accompanied by directional hypotheses.

Null Hypothesis (H_0): There is no statistically significant difference in conceptual understanding gains in Newton's Laws of Motion between students receiving STEM-integrated instruction (experimental group) and students receiving traditional instruction (control group). Any observed difference in post-test scores between the two groups is attributable to chance.

Alternative Hypothesis (H_1): Students receiving STEM-integrated instruction will demonstrate significantly greater gains in conceptual understanding of Newton's Laws of Motion from pre-test to post-test than students receiving traditional instruction, as measured by scores on the FCI-based 30-point conceptual test.

For the alternative hypothesis, H_1, a theoretical basis is provided for a one-tailed hypothesis arguing that STEM-integrated instruction including real-world problem-solving, hands-on inquiry, and engineering design creates the conditions of cognitive conflict and conceptual engagement necessary for the displacement of Newtonian misconceptions and the construction of scientific understanding (Posner et al., 1982). Empirical evidence for H_1 can be found in Francisca et al. (2021), which compared

pre-to-post conceptual gain from STEM-integrated physics instruction through a quasi-experimental design, and Asrizal et al. (2023), which likewise compared pre-to-post conceptual gain in conceptual understanding through STEM-integrated physics materials through t-test and Mann-Whitney U analysis between experimental and control groups. In addition, Abilkhairova et al. (2025) found experimental evidence in a Kazakhstani context in favor of STEM-instructed science students. Research suggests that STEM integration has a medium to large effect on science achievement (Kazu & Şenol, 2022) and, thus, the null hypothesis will be rejected if there is a statistically meaningful difference between groups at the $p < .05$ level.

1.5 Significance of the Study

This study is significant at four interconnected levels: theoretical, empirical, practical, and equity-focused. Each importance level gives its own contextual knowledge gap. The combination of all four importance levels justifies the relevance of this study for the teachers, policymakers, and students of the secondary school system in Kazakhstan.

Theoretically, the study contributes to the conceptual change literature by empirically testing whether STEM-integrated instruction creates the four conditions identified by Posner et al. (1982) as necessary for conceptual change — cognitive dissatisfaction, intelligibility, plausibility, and fruitfulness — in the specific domain of Newtonian mechanics. By analysing not only pre-to-post score changes but also topic-level gains across six Newtonian mechanics domains, the study provides a granular account of where and how STEM instruction disrupts entrenched misconceptions. The study also contributes to the integrated STEM framework literature by examining STEM effectiveness in the Kazakhstani educational context — a nationally specific test in conditions (traditional instruction, high-stakes UNT assessment, documented Newtonian misconceptions) that are representative of many developing educational systems.

Empirically, this study provides evidence that Kazakhstani secondary physics education has not yet produced. No controlled study has previously examined the effect

of STEM-integrated instruction on 9th-grade students' conceptual understanding of Newtonian mechanics in Kazakhstan. Data from PISA 2022 (OECD, 2023) and FCI data (Japashov et al., 2024) show that the conceptual gap is serious and that customary instruction does not address it. The meta-analysis that synthesizes the evidence supporting the STEM integration approach to addressing the conceptual gap (Becker & Park, 2024; Chen et al., 2025) further supports the need for such an approach. This study provides the locally contextualized, methodologically rigorous test of that response.

In practice, the six-stage STEM lesson cycle evaluated in this study can be adopted and implemented within the standard timetable, existing curriculum framework and resource allocation constraints of a Kazakhstani secondary school. It requires no specialist facilities, additional instructional time, or important budget. The teacher's guide provided in the Appendix An of this study can also act as a professional development tool.

From an equity perspective, the gender dimension addresses a documented gap in the Kazakhstani literature. Japashov et al. (2024) found gender differences in physics conceptual understanding among Kazakhstani students. Whether STEM integration can reduce these differences has direct implications for equitable curriculum design.

1.6 Limitations of the Study

Every research design involves trade-offs between what is methodologically ideal and what is practically achievable. The limitations of the present study are acknowledged transparently, as they directly bear on the interpretation and generalisability of the findings.

The main limitation of this study is the quasi-experimental study design, as students could not be randomized to conditions, making it impossible to rule out pre-existing class differences. This was partially controlled by establishing baseline equivalence in the pre-test scores, running an ANCOVA with the pre-test as a covariate, and confirming the results in non-parametric Mann-Whitney U tests. Because of the single-site design, generalization of these results is limited to the school

in which the research was conducted. Therefore, replicating the study in mainstream schools, rural schools, and schools elsewhere in Kazakhstan is necessary before generalizing the results. These limitations do not impinge on the validity of the findings but rather mark them as strong methodologically-sound evidence in a particular setting and suggest areas for future research.

1.7 Conclusion

The theoretical, contextual, and methodological groundwork in this introduction has led to one conclusion that the evidence presented in this chapter reinforces. There is a severe and systematically documented conceptual understanding deficit in Kazakhstani secondary physics education. In Kazakhstan, Japashov et al. (2024) reported that 48.15% of 575 university physics students held resistant Newtonian misconceptions. Internationally, drawing on PISA 2022 data, 45% of 15-year-olds in Kazakhstan failed to reach even the lowest scientific skill level, a long way below the OECD average level, with a 21-percentage-point gap remaining between Kazakhstan and the OECD average (OECD, 2023). These two sets of evidence taken together establish that the problem is real, measurable and urgent.

The theoretical framework, comprised of Conceptual Change Theory (Posner et al., 1982), the Integrated STEM Framework (Kelley & Knowles, 2016; Thibaut et al., 2018; Roehrig et al., 2021), and Meaningful Learning Theory (Ausubel, 1968), supports the pedagogical approach of STEM integration which is substantiated by the meta-analytical findings of this study. Becker and Park (2024) observed a large effect of STEM integration on student learning ($g = 0.661$) across 40 studies involving a total of 15577 learners, which was replicated by the larger synthesis of Chen et al. (2025) involving 124 studies.

The three hypotheses were set up as RQ1 (the main experimental question, formal null and alternative hypotheses), RQ2 (exploratory gender level) and RQ3 (exploratory theme-level). The study was conducted at the State-Owned Institution "Specialized Lyceum named after Al-Farabi" of the State Institution "Education Department of Almaty Region" (Kazakhstan) with 74 students from four 9th grade

classes in the academic year in a quasi-experimental pre-test/post-test design. The importance of the research in theoretical, empirical, practical and equity terms is explained, the limitations noted and the evidence, analysis and implications outlined in the Introduction are developed in further detail in the following eight chapters.

2. Literature Review

This chapter presents a critical review of the theoretical and empirical literature that underpins the present study. The review is organised around six interconnected themes, each of which contributes a distinct layer of justification for the research design, hypotheses, and interpretive framework adopted in this dissertation.

The chapter opens with an examination of conceptual understanding in physics education - what the construct means, how it differs from procedural knowledge, and why it represents the appropriate outcome variable for a study concerned with the depth rather than the breadth of student learning (Section 2.1). Section 2.2 then documents the limitations of traditional physics instruction, establishing why conventional pedagogy systematically fails to produce genuine conceptual understanding and why an alternative instructional approach is warranted in the Kazakhstani secondary school context.

Section 2.3 reviews the empirical evidence on the effectiveness of integrated STEM education broadly, drawing on recent meta-analyses and primary quasi-experimental studies to establish that STEM integration produces significant and meaningful cognitive gains beyond what traditional instruction achieves. Section 2.4 narrows this evidence base to physics education specifically, examining studies that have applied integrated STEM approaches to physics content and documenting the mechanisms through which STEM integration is theorised to produce conceptual gains in the discipline. Section 2.5 further concentrates on the Newtonian mechanics domain - the specific subject matter of the present study - reviewing evidence on student misconceptions about force, motion, and Newton's three laws and on the instructional strategies that have proven effective in addressing those misconceptions within STEM-integrated approaches.

Finally, Section 2.6 situates the present study within the Kazakhstani teacher professional development context, documenting the current gap between national STEM policy ambitions and the practical preparation of physics teachers to implement integrated STEM instruction - a gap that the teacher's guide produced as part of this

dissertation (Appendix A) is designed to begin to address. Together, these sections establish the theoretical grounding, empirical precedent, and contextual rationale for the quasi-experimental investigation conducted in this study.

2.1 Conceptual Understanding in Physics Education

One of the most popular and longstanding issues in physics education research is the difference between procedural knowledge and conceptual understanding. Procedural knowledge is the ability to use algorithms, formulas, and calculation routines. Conceptual understanding can be described as coherent mental models of physical phenomena, or the ability to explain, predict, and reason about the physical world in terms of underlying principles and to transfer that understanding to novel situations (Hestenes et al., 1992; Redish, 2003). A student can possess one but not the other. Though the student may correctly understand the calculation of the acceleration from the equation F = ma, he or she may still incorrectly believe that the velocity of a body must point in the direction of the total force.

The gap has been described in the literature since the 1980s, although systematic study only began in 1985 when Halloun and Hestenes (1985) administered a diagnostic survey of force and motion to incoming university physics students. They found that intuitions about force and motion were poorly correlated with performance on assessments but highly correlated with conceptual learning. However, students with strongly-held misconceptions tended to cling to these misconceived ideas regardless of how well they had done on the exam, indicating that procedural success in physics examinations is not dependent on conceptual understanding.

Decades of research into the classroom have shown that this idea (McDermott 1991 and 2001) is almost universal: if something is moving, a force is pushing it; if the force stops, motion also stops. This idea cannot be presented once and then expected to stick; it needs to be reinforced continually, and is resistant to unlearning. For example, students may write down Newton's First Law of motion in their tests, and then use impetus-theory reasoning to answer a question about the motion of a body that

has stopped moving in circular motion. Instruction that produces procedural competence without conceptual change is characterized by such phenomena.

The most influential theoretical account of the existence of the gap and the conditions under which it can be closed is provided by Posner et al.'s (1982) Conceptual Change Theory, according to which scientifically flawed existing conceptions also function as coherent explanatory systems. The conditions necessary for conceptual change are: 1. dissatisfaction with the existing conception when confronted with a phenomenon it cannot explain, 2. a suitable plausibility structure. a new conception must be intelligible; (3) plausible; and (4) fruitful (i.e. it must open new avenues of understanding.

Subsequent elaborations of the theory (e.g., Vosniadou, 1994) have proposed that students' misconceptions are embedded in framework theories: tacit assumptions that organize and constrain more specific beliefs. This implies that simply addressing students' misconceptions in isolation will often not be sufficient, and that teachers must also ease conceptual change of the underlying assumptions. Thus, inquiry-based, contextually embedded instruction (such as that provided by STEM integration) is far more likely to reach these framework-level beliefs.

2.2 Limitations of Traditional Physics Instruction

In contrast, decades of research comparing customary lecture courses designed to promote conceptual change shows very consistently that customary lecture-based physics instruction is in fact ineffective: Using FCI pre- and post-test data for 6000 students in 62 introductory physics classes, Hake (1998) found that customary lecture-based physics instruction produced relatively low normalized learning gains (average $g \approx 0.23$) while interactive-engagement (IE) instruction produced much higher normalized learning gains (average $g \approx 0.48$). This difference between the institutional contexts further shows that it is instructional approach, rather than student ability or prior knowledge, that drives conceptual learning gains in physics.

The persistence of student misconceptions is a major problem, and Newtonian mechanics is the domain of physics for which misconceptions have been studied the

most. The same misconceptions about force, motion and interaction have been found across cultures, age ranges, and school systems. These misconceptions have been found in Indonesian, Kazakh, American, and two independent samples of European school children (Clement, 1982; Halloun & Hestenes, 1985; Japashov et al., 2024). The prevalence of these misconceptions across cultures suggests that they are not the result of poor education systems, but rather of how humans experience the physical world in everyday situations.

According to Clement (1982), the most common misconception in mechanics is the idea that a moving object has an internal force (that is, impetus) that is consumed in the course of its motion, and this force causes the object to keep moving. But, this idea contradicts Newton's First Law. Halloun and Hestenes (1985) identified many other misconceptions related to the notion of force. Students believe that heavier objects fall faster and that the net force acting on an object will always be in the direction of its motion. They think that force is not acting when an object is stationary. There are similar misconceptions regarding Newton's Third Law as well. In a study of students' reasoning about force interactions, researchers found that when thinking about two objects in contact, students say that the more massive or more powerful object exerts a greater force on the less massive object. This claim contradicts the Third Law (Hestenes et al., 1992).

This phenomenon is further explained by the Conceptual Change Theory. In a lecturing context, the cognitive conflict established by customary teaching is not enough to convince students to abandon deeply-held implicit theories (Posner et al., 1982). While students may be able to recall Newton's First Law for a test, they may not be able to apply impulse-theory reasoning to a new situation (McDermott, 1991).

In Kazakhstan, Japashov et al. (2024) administered the three-tier FCI diagnostic instrument to 180 undergraduates across two universities. Using the CRI format, participants' response patterns were categorized into the following: correct and confident (scientific knowledge), correct but not confident (lucky guess), incorrect and confident (misconception), and incorrect and not confident (lack of knowledge). Most

concerning, 48.15% of all the responses were of the misconception type. In this category, the subjects accepted the wrong principles with high confidence. The results were shown to be constant across the year levels, giving evidence that four years of college-based physics instruction had been ineffective in displacing the misconceptions. Most shared the impetus intuition, which erroneously assumes that a constant force is always needed to maintain motion.

Due to bureaucratic features of the education system such as the Unified National Test (UNT), which is used as an entrance assessment for all universities, teachers are pressured to focus their teaching on recall of factual procedures rather than innovative pedagogical practices (Margot & Kettler, 2019). This is confirmed by international assessments. According to the PISA 2022 (OECD, 2023) survey, only 55% of 15-year olds from Kazakhstan achieved the baseline science skill Level 2 or higher, compared to an average of 76% of 15-year olds across OECD countries. According to the TIMSS 2023 (von Davier et al., 2024) assessment, Kazakhstan's mean score in mathematics in the eighth grade was 454 and in science 443, below the center point of 500. Thus, Kazakhstan's current instructional approaches have not succeeded in producing the conceptual science understanding its own policy goals have required.

2.3 Studies on the Effectiveness of Integrated STEM Education

An increasing number of meta-analyzes and primary studies across multiple educational contexts suggest that integrated STEM education research and practice increases statistically important and practically large quantified gains in students' cognitive outcomes (i.e., conceptual understanding, science achievement) beyond the gains with customary instruction when integrated STEM education is implemented with fidelity to its essential features.

At the meta-analytic level, Zeng and Yao (2025) found, in an analysis of 66 experimental and quasi-experimental studies conducted between 2000 and 2024, that STEM education had the largest cognitive effect in secondary school (Cohen's $d = 0.58$, $I^2 = 62.1\%$). In particular, the STEM approach of active, problem-centered instruction was more effective than nominally STEM-labeled more customary direct

instruction. This finding indeed supports the main argument of the current study, namely that the success of STEM integration cannot be attributed to the label itself, but rather to the conditions of inquiry-based learning.

At the primary study level, Yaki (2022) employed a quasi-experimental pretest-posttest control group design with 112 secondary school biology students to examine the effect of an integrated STEM approach module on critical thinking skills (*European Journal of STEM Education*). The two groups were confirmed as equivalent at baseline (Wilks' $\lambda = 0.93$, $p = .24$). After the intervention, the experimental group showed a large effect size in within-group improvement ($d^2 = 1.56$), while the control group showed a negligible effect ($d^2 = 0.01$). Between-group MANOVA confirmed a significant difference across all five critical thinking subskills ($p < .05$). This study is methodologically close to the present research — quasi-experimental, two intact-class groups, pre-post design, baseline equivalence confirmed — and documents the size of cognitive gains achievable through secondary-level STEM integration.

Uzun and Şen (2023) conducted a quasi-experimental study with 7th-grade students in Turkey (*Journal of Pedagogical Research*), testing the effect of a STEM-based science learning environment on achievement in the electrical energy unit. Using a non-equivalent pretest-posttest control group design, they found a significant difference in science achievement scores favouring the STEM group, while no significant pre-test difference existed between groups — the same methodological assurance applied in the present study. Their finding extends the effectiveness evidence across subject areas and confirms replication in a secondary school context geographically and educationally comparable to Kazakhstan.

Guzey and Li (2023) add a longitudinal perspective to this evidence base (*Journal of Science Education and Technology*). Their study followed 245 middle school students across three academic years, each year involving a different integrated STEM curriculum unit. Mixed-effects modelling revealed that behavioural engagement within STEM activities was a critical predictor of science achievement

gains. Notably, no gender differences in science achievement were found across any of the three STEM units — a finding directly relevant to the gender dimension of the present study — suggesting that well-designed STEM integration can produce equitable outcomes for male and female students.

Taken together, these studies establish that the effectiveness of integrated STEM education is neither confined to any single country nor limited to a particular science subject. The pattern of positive cognitive outcomes is consistent across diverse national contexts, grade levels, and scientific domains, though the effect is reliably largest when instruction fully operationalises the principles of problem-centred, inquiry-based, and design-based learning identified by Thibaut et al. (2018). The present study tests whether this pattern holds in the specific and previously unexamined context of 9th-grade Newtonian mechanics instruction in a Kazakhstan secondary school.

2.4 STEM-Integrated Instruction in Physics Education

The acronym STEM — Science, Technology, Engineering, and Mathematics — has been in educational policy use since the early 2000s, but the conceptualisation of STEM as an integrated instructional approach emerged more gradually. Early uses described efforts to increase participation in STEM fields and improve achievement in individual STEM subjects. The distinctly pedagogical concept of STEM integration — in which two or more STEM disciplines are deliberately connected within a unified learning experience — developed as researchers and practitioners recognised that disciplinary separation was itself a barrier to the holistic, problem-centred learning that science and engineering practitioners actually engage in.

Kelley and Knowles (2016) offer the most widely cited definition: integrated STEM education connects two or more STEM disciplines within a pedagogical framework grounded in authentic, real-world contexts that provide students with opportunities for inquiry, design, and problem-solving. Authenticity is a defining feature: STEM integration requires that the mathematical, technological, or engineering component genuinely contributes to understanding or solving a problem

that science cannot address alone. English (2016) provides a complementary definition emphasising student agency: STEM integration creates environments in which students encounter problems requiring them to draw on multiple disciplines, generate and test hypotheses, design solutions, and interpret data. Honey, Pearson, and Schweingruber (2014), in the National Research Council's comprehensive review, identify a spectrum of integration approaches from loose coordination through full integration, and conclude that degree of integration is positively associated with learning outcomes.

Thibaut et al. (2018) conducted a systematic review of 23 empirical studies on integrated STEM instruction in secondary education and identified five evidence-based principles characterising effective implementation. These principles served as the design criteria for the experimental intervention in the present study. Content integration requires that disciplines are genuinely interconnected, not merely juxtaposed. Problem-centred learning anchors the activity in a real-world problem that motivates the use of STEM knowledge and provides a criterion for evaluating the adequacy of understanding. Inquiry-based learning engages students in active investigation — generating their own data, formulating and testing predictions, and drawing conclusions from evidence — creating the cognitive conflict that is a precondition for conceptual change. Design-based learning engages students in engineering tasks in which they must commit to specific applications of principles and test their designs against physical reality. Cooperative learning structures group work, so that students explain their reasoning to peers and surface implicit assumptions, as well as alternative approaches to physical reasoning.

Theory about why STEM should be used in physics instruction tends to draw its rationale from the convergence of three theoretical perspectives, with the working mechanism lying in Conceptual Change Theory (Posner et al., 1982), which describes cognitive conflict, authentic context, and iterative revision as generally necessary for a valid misconception displacement. Situated Cognition Theory (Brown, Collins & Duguid, 1989) and Ausubel (1968)'s Meaningful Learning Theory explain how, instead of being transmitted via verbal instruction, knowledge is constructed through

participation in authentic practice, and how knowledge is retained by connecting to existing cognitive structures in authentic contexts, while unconnected facts or formulas are forgotten.

Roehrig et al. (2021) brought together these theoretical strands into an overall conceptual framework and states that coherence (disciplines are interdependent), authenticity, and disciplinary depth are essential features of STEM integration. The aspects of authenticity (real-world problem) and disciplinary depth (conceptual rigor rather than accessibility) in this model had implications for the STEM instruction delivered to the experimental group in the current study.

The idea of prioritizing STEM education gained currency in the past decade in Kazakhstan in line with national priorities to diversify an economy that had been dominated by natural resources and establish a knowledge-based working population. The State Program for Education and Science Development 2020-2025 (Ministry of Education and Science of Kazakhstan, 2016) identifies the introduction of integrated inquiry-based and interdisciplinary elements in science and mathematics subjects in the school curriculum as a priority policy. In addition, specialized STEM entities such as the Nazarbayev Intellectual Schools network and STEM centers based in regional universities have been created. STEM education has also been implemented by the Y. Altynsarin National Academy of Education and the NCPD Orleu, which has developed modules on training physics teachers (Abdrakhmanova et al., 2025). Despite these efforts, lack of planning time, a lack of professional development, and the structural pressures of the UNT still limit the effectiveness of these methods (Margot & Kettler, 2019).

2.5 STEM and Conceptual Understanding of Newtonian Mechanics

Meta-analytic syntheses provide the strongest evidence of the effects of STEM integration. Becker and Park (2024) synthesized 40 experimental and quasi-experimental studies with a combined sample of 15,577 students, yielding a strong

overall effect on academic achievement (Hedges g = 0.661, 95% CI [0.548, 0.774]). The moderator analyzes also suggest that the effect was strongest for secondary school students, for science outcomes, and for studies lasting 4-8 weeks; all of which are the case in the current study. The largest meta-analysis by Chen et al. (2025) synthesizes 124 studies published between 2010 and 2022 on STEM integration with K-12 students. The analyzes found a large effect of STEM integration on conceptual knowledge, scientific reasoning, and problem-solving ability; a medium effect on knowledge acquisition; and a medium effect on problem-solving task performance. Kazu and Şenol (2022) similarly documented positive effects of STEM education on academic performance across diverse contexts, with effect sizes ranging from small to large depending on degree of integration, implementation quality, and outcome measure. The consistency of this meta-analytic evidence across different syntheses and analytical frameworks constitutes one of the most reliable findings in contemporary educational research.

Beyond meta-analytic evidence, several primary studies are particularly relevant because of their focus on physics conceptual understanding and quasi-experimental design. Francisca, Paring, Kodama, and Iwayama (2021) conducted the closest published parallel to the present research: their quasi-experimental study of 30 secondary students in Indonesia found a statistically significant improvement in physics conceptual understanding scores before and after STEM instruction, and a significant positive relationship between test scores and agreement with the STEM approach. Asrizal, Festiyed, Ashel, and Amnah (2023) examined the effect of STEM-integrated physics digital teaching materials on secondary students' conceptual understanding in Indonesia, using a quasi-experimental non-equivalent control group design with both independent samples t-test and Mann-Whitney U analysis — the same analytical approach used in the present study — and found STEM-integrated materials significantly outperformed traditional materials on conceptual understanding.

Dominguez, De la Garza, Quezada-Espinoza, and Zavala (2024) examined STEM physics-mathematics integration through modelling and found that the

interdisciplinary approach significantly enriched conceptual understanding, with collaborative modelling activities proving particularly effective. In the Kazakhstani context specifically, Abilkhairova, Ortaeva, and Jumassaeva (2025) reported significantly superior academic outcomes for STEM-taught primary students compared to a traditional control group, providing the first experimental evidence for STEM effectiveness within Kazakhstan. Naushabekov, Maxutov, Mansurova, and Japashov (2025) provided methodological and contextual precedent for FCI-based instruments in Kazakhstani secondary physics education, directly validating the instrument design of the present study.

The most widely used concept inventory in physics education research is the Force Concept Inventory (FCI) developed by Hestenes, Wells, and Swackhamer (1992), which has 30 multiple choice questions, each with a single correct Newtonian answer and four distractor responses based on empirically established student misconceptions. Thus the FCI is both diagnostic and performative. Its content validity is established through expert judgement and curriculum matching (Hestenes and Halloun 1995). Henderson (2002) found that the FCI has adequate reliability (Cronbach's α between 0.78 and 0.86) and strong construct validity in a number of studies. Hake (1998) conducted a survey of the instrument with six thousand students and established the normalized gain index g as the standard comparative metric. Furthermore, Hasan, Bagayoko, and Kelley (1999) developed the Certainty of Response Index (CRI) to complement the FCI. In the CRI, students' ratings are categorized as scientific knowledge, lucky guess, misconception, and lack of knowledge. The CRI provides more diagnostic information than a simple appear correct/incorrect scoring, as it can diagnose the students' misconceptions with a high degree of confidence. The FCI has been used in Kazakhstan in secondary education by Naushabekov et al. (2025) and for the university level by Japashov et al. (2024).

The present study used a 30-question FCI questionnaire modified for the 9th grade Kazakhstani curriculum and the Sure/Not Sure confidence level of the CRI framework at the question level. The default index of the standardized learning gain

was determined using Hake's normalized gain index, which defines the categories of high gain ($g > 0.70$), medium gain ($0.30 \leq g \leq 0.70$), and low gain ($g < 0.30$).

2.6 Teacher Professional Development for STEM Integration in Kazakhstan

Lastly, the quality of STEM integrated instruction in secondary education can also depend on pedagogical content knowledge, interdisciplinary skills and assessment skills of teachers needed for successful adoption and implementation of STEM practices in classroom environments, as well as on the quality of the curriculum and instructional materials. The teacher readiness dimension thus represents one of the biggest challenges, as yet unresolved, on the national STEM agenda.

According to Kazakhstan's Law On Education, teachers are required to undergo formal professional development every five years (Ministry of Education and Science of Kazakhstan, 2016). Historically, however, professional development programs in the country have mainly focused on content updates related to the national examination system and not on pedagogical practices related to STEM integration. The downside of this is that while teachers are prepared to teach in a customary, single-subject, lecture-based classroom, they are ill-prepared to engage students in the engineering design-based inquiry and collaborative learning that are hallmarks of effective integrated STEM teaching and learning (Bakytkazy et al., 2025; Abdrakhmanova et al., 2025).

The strongest evidence of the problem comes from Balta et al. (2024). The researchers conducted a survey with 126 high school physics teachers in Almaty to measure the discontentment of teaching. Discontentment of teaching is defined as the divergence between teachers' pedagogical goals and their actual teaching. Their Bayesian factor analysis suggests the greatest source of discontent for Kazakhstani physics teachers, with a mean discontent score of 1.33 (on a 5 point scale), was in the area of implementing inquiry based teaching and learning. While this value is low in absolute terms, it has an obvious implication: the area of pedagogy that STEM

integration most requires when it comes to teacher efficacy is the one where Kazakhstani physics teachers are least sure of their abilities. Balta et al. argue, explicitly, that inquiry based pedagogies are the preferred approach for physics PD in Kazakhstan, and a matter of national urgency.

This incapacity to enact policy in the classroom is structural. Abdrakhmanova et al. (2025) examined the STEM competency being formed in teacher education programs in Kazakhstan from 2019 to 2023 (Open Education Studies). Although Kazakhstan has established STEM schools, implemented teacher training master's programs under Erasmus+ at three universities, and trained teachers through Nazarbayev University partner Finnish universities, the national framework for STEM teacher preparation at the in-service level has yet to be constructed. Researchers find that in-service and pre-service physics teachers have difficulty with STEM cross-disciplinary integration, inquiry-oriented curriculum development and implementation, as well as practice-oriented modules which would teach them how to implement STEM-based teaching and learning in their physics classrooms. According to Abdrakhmanova et al., although schools are now better equipped with updated laboratories and digital equipment, the pedagogical knowledge necessary to use these in teaching STEM has not kept pace with the advancements.

The teachers' perceptions dimension of this problem was documented in Bakytkazy et al. (2025). In Almaty, 164 physics teachers were surveyed about their pedagogical needs for STEM education. Based on the qualitative data of the interviews and focus group discussions, the authors concluded that although the teachers were aware of STEM as a national education trend, they lacked the pedagogical knowledge, interrelated subject matter knowledge, and teaching materials to implement integrated STEM in the physics curriculum. Teachers identified the lack of STEM professional development, the demands of the UNT, and the pressure to integrate interdisciplinary lesson plans into the Kazakhstani timetable's single-subject format as the most meaningful structural barriers. These barriers to change are not failures of teacher

preparation, but are instead constraints created by the professional environment of Kazakhstani physics teachers.

The national research landscape on this issue is mapped by Zhumabay et al. (2024), who conducted a systematic review of Kazakhstani STEM education research in the *European Journal of STEM Education*. Their review confirms that teacher preparation for integrated STEM is one of the most consistently identified underdeveloped areas in the national evidence base, and that locally grounded, curriculum-embedded instructional resources — designed specifically for the constraints of the Kazakhstani secondary school context — are among the most urgently needed contributions that research can make to STEM implementation in the country.

This body of evidence directly informs the design of the instructional intervention in the present study. Rather than expecting the experimental group teacher to design and adapt STEM lessons independently — a demand that the literature identifies as the primary barrier to adoption — the study provides a complete, ready-to-implement six-stage lesson guide developed in Kazakh and aligned to the 9th-grade Newtonian mechanics curriculum. This design decision is grounded in the evidence that the provision of concrete, curriculum-embedded instructional materials is the most practical scaffold available to physics teachers implementing STEM for the first time in a national context where targeted professional development remains absent (Balta et al., 2024; Bakytkazy et al., 2025; Abdrakhmanova et al., 2025). The teacher's guide produced as part of this dissertation thus contributes not only to the empirical investigation of STEM effectiveness, but directly to the national need for accessible STEM instructional resources for secondary physics teachers in Kazakhstan.

2.7 Gender Differences in STEM Learning

Gender differences in physics education have been documented across multiple dimensions: enrolment and participation rates, academic performance, self-concept and self-efficacy, and conceptual understanding as measured by diagnostic instruments. The present study is concerned specifically with conceptual

understanding, examining whether male and female students differ at baseline and whether STEM instruction changes this pattern.

Gender differences on FCI-based knowledge have been reproducible in Western data sets. Kost, Pollock, and Finkelstein (2009) study the FCI of 2000 introductory physics students and find that women start with lower scores than men and make the same or larger gains but end up with lower scores than men: the gap is partially but not fully closed. Coletta and Phillips (2005) noted that gender differences on the FCI pre-test persisted through instruction in several course models.

However, the relationship between gender and physics conceptual understanding appears to be moderated by instructional approach. Lorenzo, Crouch, and Mazur (2006) found that interactive-engagement instruction not only improved overall FCI gains but significantly reduced the gender gap compared to traditional lecture instruction. This suggests that the gender gap is at least partially a product of pedagogical mismatch: traditional instruction, which favours a transmission model, may be less congruent with the learning patterns of female students. STEM-integrated instruction, emphasising collaborative inquiry, real-world context, and iterative problem-solving, may create a more equitable learning environment.

In Kazakhstan, Japashov et al. (2024) found gender differences in their FCI-based university study, suggesting that gender interacts with conceptual understanding in ways that may be specific to the local educational and cultural context. Whether secondary STEM instruction can reduce or eliminate these differences is an open empirical question that the present study addresses through its exploratory second research question. The equity dimension is significant: evidence that STEM instruction reduces gender inequality in physics conceptual understanding would be directly relevant to United Nations SDG 4 (Quality Education) and SDG 5 (Gender Equality) commitments in Kazakhstan.

2.8 Research Gap and Rationale for the Study

The literature reviewed in this chapter converges on a coherent body of evidence establishing both the nature of the problem and the basis for STEM integration as a

response. The conceptual understanding literature establishes that procedural competence and genuine conceptual understanding are distinct and imperfectly correlated: traditional physics instruction reliably produces the former while consistently failing to produce the latter (Hestenes et al., 1992; Hake, 1998; McDermott, 2001). Conceptual Change Theory provides the process (Posner et al., 1982; Vosniadou, 1994) and the misconceptions literature provides the items on the instructional agenda - impetus theory, force-velocity confusions, Newton's Third Law violations, etc. - which have to be addressed (Clement, 1982; Halloun & Hestenes, 1985; Japashov et al., 2024).

The STEM integration literature shows that teaching that features multiple evidence-based principles, specifically content integration, problem-centered, inquiry, design, and cooperation, leads to considerably better understanding than customary teaching (Thibaut et al., 2018; Kelley & Knowles, 2016). This is supported by meta-analyzes that show large effects on cognitive and academic skills (Becker & Park, 2024; Chen et al., 2025), as well as studies that have shown similar results for physics and mechanics content in particular (Francisca et al., 2021; Asrizal et al., 2023).

The gender literature identifies differences in physics conceptual understanding that are moderated by instructional approach, with interactive-engagement instruction associated with reduced gaps (Lorenzo et al., 2006), though the Kazakhstani secondary school gender dimension remains unexplored.

Despite this body of evidence, four specific research gaps remain. First, no controlled study has examined the effect of STEM integration on 9th-grade physics conceptual understanding in Kazakhstan. The sole Kazakhstani experimental study (Abilkhairova et al., 2025) was limited to primary school students, leaving the secondary level — the level most directly challenged by UNT pressure and most extensively documented for misconceptions — unexamined. Second, no FCI-based study of Newtonian mechanics conceptual understanding has been conducted at the secondary school level in Kazakhstan: Japashov et al. (2024) worked at university level, while Naushabekov et al. (2025) used project-based learning rather than STEM

integration. Third, the gender dimension of STEM integration in Kazakhstani secondary physics education is entirely unexplored. Fourth, the topic-level differential effects of STEM integration across the distinct domains of Newtonian mechanics have not been examined in the Kazakhstani context; existing meta-analyses report only overall effects.

The present study directly addresses all four gaps. By conducting the first controlled quasi-experimental investigation of STEM integration's effect on 9th-grade Newtonian mechanics conceptual understanding in Kazakhstan — using an FCI-based instrument with CRI confidence ratings, across 74 students in four 9th-grade classes, examining gender as an exploratory dimension, and analysing gains across six Newtonian mechanics topic domains (kinematics, Newton's First Law, Newton's Second Law, Newton's Third Law, superposition of forces, and kinds of force) — this research makes a theoretically grounded, empirically rigorous, and contextually specific contribution to the international literature on STEM integration in physics education.

2.9 Conclusion

The literature reviewed in this chapter establishes a coherent and well-supported case for the present study. Taken together, the evidence converges on a consistent conclusion: traditional physics instruction, structured around teacher-centred delivery and algorithmic problem-solving, systematically fails to produce the conceptual restructuring that students need to move beyond persistent Newtonian misconceptions, and this failure is particularly pronounced in examination-oriented systems such as Kazakhstan's, where surface-level performance on calculation tasks can mask deep conceptual gaps (Sections 2.1 and 2.2). At the same time, a growing and methodologically robust body of evidence demonstrates that integrated STEM education — when implemented with fidelity to its defining principles of problem-centred, inquiry-based, design-based, and cooperative learning — produces significant and practically meaningful cognitive gains at the secondary school level, including in physics-specific domains directly relevant to the present study (Sections 2.3 and 2.4).

The evidence on Newtonian mechanics specifically confirms that students' FCI-measurable misconceptions about force and motion are not resolved by traditional instruction but are highly responsive to the contextualised, prediction-observation-explanation cycle that STEM integration provides through authentic engineering design and hands-on investigation (Section 2.6). And the Kazakhstani teacher development context confirms that the absence of ready-to-use, curriculum-aligned STEM instructional resources for physics teachers represents both the primary structural barrier to implementation and the primary practical contribution that research at this scale can offer to the national educational landscape (Section 2.6).

Despite this convergent evidence, a significant research gap remains. No published controlled study has examined the effect of integrated STEM instruction on 9th-grade students' FCI-based conceptual understanding of Newtonian mechanics in the Kazakhstani secondary school context. The international meta-analytic evidence cannot be assumed to generalise uncritically to a system operating under different baseline instructional conditions, different curriculum constraints, and a different assessment culture. The present study directly tests the following in a quasi-experimental pre-post design, with two intact 9th-Grade classes from the State Owned Institution "Specialized Lyceum named after Al-Farabi" in the Karasai District of Almaty's: Is there considerably greater conceptual gain for a six-week STEM-integrated Newton's Laws unit vs. control? Are there important differences between boys and girls? The methodology through which this investigation is conducted is described in Methodology part.

3. Methodology

This chapter describes the methodological framework adopted in the present study. It details the research design, the characteristics of the participants, the data collection instrument, the intervention procedure, the ethical considerations governing the study, its assumptions and delimitations, the proposed data analysis plan, and the full six-week STEM-integrated lesson plan delivered to the experimental group.

3.1 Research Design

This study employed a quasi-experimental non-equivalent pretest-posttest control group design to investigate the effect of STEM-integrated instruction on 9th-grade students' conceptual understanding of Newtonian mechanics. In this design, participants were not randomly assigned to conditions; instead, two existing intact classrooms were assigned to the experimental and control conditions respectively, reflecting the practical constraints of conducting research within a functioning secondary school (Creswell & Creswell, 2018; Shadish, Cook & Campbell, 2002).

The non-equivalent pretest-posttest design is the most widely used quasi-experimental approach for classroom-based educational research where random assignment is ethically or logistically not feasible (Campbell & Stanley, 1963). The use of a pre-test and post-test design for each group, with the same validated pre/post-test (the FCI) in each group, allowed us to assess the difference in gain due to the different treatment, while controlling for the pre-existing differences statistically. We included the pre-test as a covariate (Analysis of Covariance or ANCOVA) because it offered a more sensitive estimate of the effect of the treatment than post-test scores alone.

3.2 Participants of the Study

The study was conducted at the State-Owned Institution "Specialized Lyceum named after Al-Farabi" of the State Institution "Education Department of Almaty Region" in the Karasai District, Kazakhstan, during the fall semester of the 2025–2026

academic year (September–December 2025). The target population was 9th-grade students studying physics within the Kazakhstani national secondary school curriculum.

The sample included 74 students from two intact 9th-grade classes at the same school. One of the classes (n=34) was the experimental group which received the STEM-integrated instruction. The other class (n=40) served as the customary instruction control group. The full sample included 36 male and 38 female students, and both classes were taught by physics teachers from the same school. Thus, school-level variables, such as physical infrastructure, administrative environment, and socio-economic status could be controlled for across the two conditions.

Purposive sampling was used to select the school on the basis of its willingness to participate, its accessibility to the researcher, and the availability of two intact 9th-grade classes at the same academic level (Creswell & Creswell, 2018). The selection of 9th grade was deliberate: Newton's Laws of Motion constitute a core component of the 9th-grade physics curriculum in Kazakhstan (Ministry of Education and Science of Kazakhstan, 2016), making this the appropriate grade level for investigating conceptual understanding of Newtonian mechanics.

Statistical equivalence of the two groups at baseline was confirmed prior to the intervention. An independent samples t-test on pre-test FCI total scores revealed no significant difference between the experimental group ($M = 8.62$, $SD = 5.56$) and the control group ($M = 9.72$, $SD = 5.08$), $t(72) = -0.895$, $p = .374$, Cohen's $d = -0.209$. A Mann-Whitney U test confirmed this finding ($U = 587$, $p = .314$). The two groups were therefore considered statistically equivalent before the intervention commenced.

3.3 Instrument of Data Collection

3.3.1 The Force Concept Inventory (FCI)

The primary data collection instrument for the study was the Force Concept Inventory (FCI) developed by Hestenes, Wells and Swackhamer (1992) and further revised by Halloun et al. (1995). The instrument is a multiple choice test containing

30 questions measuring students' understanding of Newtonian physics concepts. For each question, a participant is presented with a description of a physical situation and then given five possible answers to choose from. One answer is the scientifically correct Newtonian answer, and the other four are common misconceptions identified by physics education research.

The six FCI subscales were: (a) Kinematics, (b) Newton's First Law, (c) Newton's Second Law, (d) Newton's Third Law, (e) the Superposition Principle, and (f) Kinds of Force. The subscale structure of the FCI allowed the present study to investigate not only overall conceptual gains but also the differences between the two conditions within each domain of Newtonian mechanics.

The FCI is the most commonly used tool to assess students' understanding of Newtonian mechanics at high school and first two years of university level (Hake, 1998; Nieminen et al., 2010). In the Kazakhstani context, the FCI was validated by Japashov et al. (2024). Japashov et al. (2024) presented evidence that a three-tier version of the FCI is an appropriate tool to diagnose conceptual understanding of Newtonian mechanics among Kazakhstani university students. A 30-item standard version of the FCI, translated into Kazakh language to be understandable to all participants, was used in the present study.

3.3.2 Scoring and Response Categorisation

Total FCI scores (range: 0-30) were calculated as the number of correctly answered items. Participant responses were also classified using a four-category classification scheme (Japashov et al., 2024) which was adapted to the Kazakhstani context. There were four categories: Scientific knowledge (correct response), Lucky guess (correct response for a difficult item), Misconception (incorrect response that is a documented alternative conception) and Lack of knowledge (incorrect response due to lack of knowledge of the concept). Pre-test categorisation provided a systematic, detailed baseline profile of the student's initial state of understanding before instruction began.

3.3.3 Normalised Gain

To measure the relative magnitudes of conceptual growth under each one of the instructional conditions, Hake's (1998) normalized gain statistic, g, was computed as follows:

g = (post-test mean - pre-test mean) / (maximum score - pre-test mean)

Taking the 6,000 students as a whole, Hake (1998) proposed that the following learning gains can be interpreted as high ($g \geq 0.70$), medium ($0.30 \leq g < 0.70$), and low ($g < 0.30$). The threshold methodology resolves the ceiling effect and is useful for comparing the learning gains of groups of students that have differential precourse knowledge.

3.4 Procedure

3.4.1 Pre-test Administration

At the beginning of the intervention period in September 2025, the FCI was administered in both classes under standard conditions. The students were informed that the test was not graded and that the test was diagnostic as to what concepts students understood in relation to the motion of an object being acted upon by a force. All 74 students who were present during the administration of the pre-test completed the instrument. The pre-test data were coded and secured before the intervention so as not to influence instruction.

3.4.2 Experimental Condition — STEM-Integrated Instruction

The experimental group received the intervention for six weeks and was taught Newton's Laws integrated with STEM content using the six-stage lesson cycle of Engagement, Exploration, Explanation, Elaboration, Evaluation, and Extension based on Thibaut et al. (2018)'s five principles of integrated STEM education: STEM content integration, problem-centered learning, inquiry-based learning, design-based learning, and cooperative learning.

Each weekly session incorporated an engineering context, an experiential laboratory activity, and group work. Examples included an investigative ramp and trolley experiment using a smartphone accelerometer, constructing and testing balloon powered vehicles, and carrying out a series of investigations into multi-force vectors. Lesson activities found in physics education literature, such as the Institute of Physics Teaching Advanced Physics series (Institute of Physics, 2006), were adapted to the Kazakhstani curriculum, available resources, and Kazakh language of instruction, and were mapped to both 9th grade Newtonian mechanics physics content knowledge, and the FCI subscale domains considered in the study.

The six-week lesson plan, aligned to the Kazakh national physics curriculum, was provided to the experimental group's teacher before the intervention. Physics lessons were taught by the class's regular physics teacher for practical considerations of ecological validity. A complete lesson plan is given in section 3.8 of this chapter.

3.4.3 Control Condition - Traditional Instruction

The control was taught the conventional Kazakhstani 9th-grade physics curriculum for six weeks (no modifications); teacher-centered instruction of Newton's three laws of motion, the derivation of formulas, worked examples, and textbook problems, whereby students followed a sequence of teachers' demonstrations and textbooks, lecturing, and questions. No engineering design activities, collaborative inquiry activities, and computer-based digital simulation modeling tools were integrated into the control group lessons. Furthermore, the teacher of the control group was not informed of the content of the experimental lesson guide.

3.4.4 Post-test Administration

At the conclusion of the six-week intervention period in December 2025, the FCI post-test was administered to both groups under the same standardised conditions as the pre-test. All 74 students completed the post-test instrument. Post-test data were collected, coded, and secured immediately following administration for subsequent statistical analysis.

3.5 Ethical Considerations

The present study was conducted in accordance with the ethical principles governing educational research involving human participants. The following procedures were observed throughout the study.

Institutional permission was obtained from the administration of the State-Owned Institution "Specialized Lyceum named after Al-Farabi" prior to commencing data collection. The school principal and relevant academic staff were informed of the study's purpose, design, procedures, and intended use of findings.

Participant welfare was protected throughout. Participation occurred entirely within students' normal classroom instruction time, and no student was subjected to any procedure beyond their regular physics lessons. Designed as a test of learning rather than an achievement test, the FCI sought to eliminate test anxiety and social desirability.

Later responses from participants were assigned numerical codes to protect their confidentiality and personally identifiable information was not maintained in research databases. All data were stored in a secure location accessed only by the researcher and the faculty member supervising the project.

Voluntary participation. Students and parents/guardians were informed that participation in the research component was voluntary and that their refusal to participate would have no negative impact on the student's records or academic standing. All students consented to participate.

Data were used for the purposes of this dissertation only and were not shared with third parties. To protect participant anonymity, results are reported at the group level.

3.6 Assumptions and Delimitations

All research in education rests on premises that are accepted as true or real but that remain unexamined, as well as on assumptions about what can and cannot be

claimed. Here, then, are the premises and assumptions of which I am most conscious, presented honestly and concretely.

The four research assumptions are the following: (1) The Kazakh language version of the FCI instrument can validly measure the conceptual understanding of Newtonian mechanics of 9th-grade students attending Al-Farabi Specialized Lyceum in Almaty. There are other FCI-based studies with Kazakhstani physics students (Japashov et al., 2024). Second, we assumed that the treatment to which both conditions were assigned was implemented for the full 6-week duration of the study. This means that the teacher of the experimental group delivered STEM lesson activities in the prescribed manner and the teacher of the control group delivered instruction about Newton's Laws and motion using conventional non-STEM instruction. Finally, we did not use a fidelity observation protocol in the current study. It is reasonable to assume that the students answered the FCI items honestly, as the FCI was ungraded and the post-test was administered under supervised testing conditions. Fourth, the two intact classes are assumed to have been similar in regard to the unmeasured variables in the FCI pre-test. Factors such as general academic motivation, out-of-class physics experience, and home support for academic work were controlled for by the fact that the two classes were from the same school, grade, and curriculum track, and that the two classes were statistically equivalent on the FCI pre-test ($t(72) = -0.895$, $p = .374$).

The present study was conducted in one school, Al-Farabi Specialized Lyceum, Karasai District, Almaty Region. The total number of students who participated in the present study was 74 in two intact classes. The results here describe what happened in this context in the fall semester of 2025, and it is not possible to generalize the results to other schools, districts, or regions of Kazakhstan without validation by independent research. The intervention only provided the Newton's Laws of Motion unit in the 9th-grade physics course during a six-week period. The only comparison not made is in relation to the effect of STEM integration on other physics content, other science content, or other grade levels. Third, the only form of conceptual understanding measured is the FCI. The study also does not look at other potential consequences of

STEM-based instruction, such as problem solving, scientific reasoning, engineering design, collaboration, and student attitude towards STEM. Fourth, the six weeks of the study was the amount of time allotted by the school where the study was performed for the implementation of the Newton's Laws unit in their STEM curriculum. It is an open question whether a longer-duration STEM intervention would yield larger, smaller, or different effects than those we measured. Fifth, because we applied a quasi-experimental design with intact classes, we cannot rule out the possibility that confounding unmeasured group differences such as teacher effects, classroom characteristics, and within-class student motivation that were not controlled by the baseline equivalence conditions may have contributed to any measured group differences between the two outcome measures.

3.7 Data Analysis Plan

All statistical analyses were conducted using Jamovi (Version 2.7) statistical software (The jamovi project, 2025). The analytical procedures were selected to correspond directly to each of the three research questions.

RQ1 — Does STEM-integrated instruction produce significantly greater gains in 9th-grade students' conceptual understanding of Newton's Laws compared to traditional instruction? Three complementary analyses were conducted. First, to confirm that the groups were statistically equivalent before the intervention, an independent samples t-test and Mann-Whitney U test were conducted on pre-test FCI total scores, with Levene's test used to verify the homogeneity of variances assumption. Second, to compare post-intervention conceptual understanding, an independent samples t-test and Mann-Whitney U test were conducted on post-test FCI total scores, with Cohen's d reported as the parametric effect size and rank biserial correlation reported for the non-parametric test. Third, a two-way ANCOVA was conducted with post-test FCI total score as the dependent variable, group as the primary fixed factor, gender as a secondary fixed factor, and pre-test FCI total score as the covariate, with partial eta squared (η^2) reported as the effect size. Fourth, Hake's (1998) normalised

gain statistic (g) was calculated for each group and interpreted against established thresholds.

RQ2 — Is there a difference in conceptual understanding between male and female students, and does this difference change following STEM-integrated instruction? Independent samples t-test and Mann-Whitney U tests were conducted comparing male and female students on pre-test and post-test FCI total scores across the full sample. Cohen's d was reported to quantify the magnitude of any gender difference at each time point. The Gender × Group interaction term within the ANCOVA described under RQ1 directly tested whether the effect of instructional condition on post-test scores differed between male and female students.

RQ3 — Which specific domains of Newtonian mechanics show the greatest conceptual gains following STEM-integrated instruction? Mann-Whitney U tests were conducted separately for each of the six FCI subscale domains — Kinematics, Newton's First Law, Newton's Second Law, Newton's Third Law, Superposition of Forces, and Kinds of Force — at both pre-test and post-test. Pre-test subscale comparisons confirmed domain-level baseline equivalence between groups. Post-test subscale comparisons identified which specific domains showed statistically significant between-group differences after the intervention. The non-parametric Mann-Whitney U test was selected for subscale analysis given the ordinal nature and smaller range of subscale scores. A significance threshold of $\alpha = .05$ was applied uniformly to all analyses.

3.8 Six-Week STEM-Integrated Lesson Plan

The subsequent section presents the entire six-week STEM-integrated lesson plan for Newton's Laws given to the experimental group during the fall of school year 2025. It was written in the Kazakh language and aligned with the national physics curriculum for the 9th grade. The six stages of the integrated STEM education lesson plan including Engagement, Exploration, Explanation, Elaboration, Evaluation and Extension (Thibaut et al., 2018; Kelley & Knowles, 2016) are closely aligned with each week.

Table 3.2 *Six-Week STEM-Integrated Newton's Laws Lesson Plan Overview*

Week	Main Topic	Key Concepts	STEM Integration	Expected Outcome
1	Introduction to Forces and Motion	Balanced and unbalanced forces	Physics + Mathematics	Students explain the effect of forces and predict motion
2	Newton's First Law (Inertia)	Relationship between force and motion	Engineering + Technology	Students demonstrate inertia through experiment
3	Newton's Second Law (F = ma)	Force, mass, and acceleration	Engineering + Mathematics + Technology	Students verify the force-acceleration relationship through data collection
4	Newton's Third Law (Action-Reaction)	Interaction forces	Engineering Design	Students model the action-reaction principle
5	Application of Newton's Laws	Friction, safety examples	STEM + Design Thinking	Students connect the laws to real-world applications
6	Motion Graphs and Data Analysis	Data visualisation	Technology + Mathematics	Students analyse motion data and compare results

Lesson Structure

Each of the six weekly lessons follows the same six-stage cycle. Stage 1 (Engagement) introduces a real-world problem or question to activate prior knowledge and curiosity. Stage 2 (Exploration) involves student-led experiment and data collection. Stage 3 (Explanation) requires students to analyse their data and draw conclusions before the teacher formalises the physical principle. Stage 4 (Elaboration) presents an engineering design task that requires application of the physical principle. Stage 5 (Evaluation) involves structured discussion, peer review, and reflection on outcomes. Stage 6 (Extension) applies the principle to a new real-world context beyond the classroom activity.

Table 3.3 *Assessment Types Used Across the Six-Week STEM Lesson Plan*

Assessment Type	**Tool**	**Purpose**
Pre-test and post-test	Force Concept Inventory (FCI)	Measure conceptual understanding before and after instruction
Observation checklist	Teacher observation sheet	Monitor STEM activity engagement and participation
Project work	Engineering design rubric	Evaluate quality of engineering solutions and scientific reasoning

Reflection	Open-ended written prompts	Assess students' self-evaluation of their learning experience

Week 1 — Introduction to Forces and Motion

Note. This lesson was adapted and localised for the Kazakhstani educational context from the Institute of Physics Teaching Advanced Physics resource collection, Episode 202: Forces in Equilibrium (Institute of Physics, 2006).

STEM Integration: Physics + Mathematics

Engagement: Students are shown a video of a car suddenly braking and asked: "What force caused the car to stop? What would happen without that force?"

Exploration: Two students hold a rope taut between them. When a mass of 1 kg is hung in the middle of the rope, it is impossible to keep the rope horizontal. This is then followed by a rope taut demonstration (using local materials) that is based on the washing line demonstration from Episode 202. Students also push and pull a variety of objects of different mass across surfaces and measure the force exerted with spring-balance, recording the results in tables and recognizing patterns.

Explanation: The teacher uses student data to introduce balanced forces and unbalanced forces. Students identify balanced or unbalanced forces in their results and use Newton's framework to discuss their observations.

Elaboration: In groups, students create a force diagram representing the forces acting on a bridge supporting a load. The force diagram activities are adapted from Episode 202 TAP 202-2, Forces in Equilibrium (student worksheet), replacing the context of the problems to better suit Kazakhstani students (e.g. the forces on a bridge under a load).

Evaluation: Groups present force diagrams and peer/teacher provide feedback. Students each write down one sentence to explain what happens to motion when forces are unbalanced.

Extension: Students look at a picture of a construction crane, and identify all the forces acting on the load, and predict what would happen if one disappeared**.**

Assessment: Teacher observation checklist, force diagram assessed with rubric.

Materials: 4-5 m long rope, 1 kg hanging mass, 2 spring balances, masses of different weights, both smooth and rough surfaces, ruler, table, data sheet.

Week 2 — Newton's First Law (Inertia)

STEM Integration: Engineering + Technology

Engagement: Students watch a slow-motion video of a passenger lurching forward during sudden braking. Discussion: "Why does the passenger keep moving when the car stops? What force would stop them?"

Exploration: Students place a coin on a card balanced on a glass and flick the card sharply, observing that the coin falls into the glass. They also stack books on a trolley and observe the books sliding when the trolley is suddenly stopped. Smartphone accelerometer apps record the motion profiles.

Explanation: Students use their observations to articulate Newton's First Law in their own words before the teacher formalises it. The concept of inertia is connected to mass — students compare the behaviour of heavy and light objects in their experiments.

Elaboration: Engineering design task: students design a simple seatbelt model using elastic bands and a toy figure on a wheeled platform, testing whether their design prevents the figure from falling forward when the platform is suddenly stopped.

Evaluation: Groups should test if their design successfully restrained their figure, discussing which design features worked and why, and link their overall design back to Newton's First Law.

Extension: Students could research an application of inertia used in engineering designs, such as airbags, crumple zones or anti-lock braking systems. They could write a brief description of the application of Newton's First Law.

Assessment: smartphone data printout; seatbelt design against an engineering design rubric; a written extension task.

Materials: coins, cards, glasses, toy figures, wheeled platform, elastic bands, smartphones with accelerometer apps.

Week 3 — Newton's Second Law (F = ma)

STEM Integration: Engineering + Mathematics + Technology

Engagement: Students see a small car and a large truck, both being pushed by the same force. Discussion: "Which one accelerates faster? Why?"

Exploration: Ramp-and-trolley experiment. Students use a ramp and trolley, masses hanging from the ramp, and the accelerometer available on their smartphone. They vary the force (hanging mass) for a constant trolley mass, then the trolley mass for a constant force, recording and tabulating the resulting acceleration in each case.

Explanation: Students then plot a force-acceleration graph and a mass-acceleration graph with their data and they are guided to derive the formula F = ma rather than being told the formula.

Explanation: Then they use F = ma to calculate the force needed to accelerate the car at that mass by that amount, and compare their answer with the engine output data of real cars described on data cards.

Assessment: Each group presents one of the graphs. They explain whether or not the data supports Newton's Second Law. They explain possible sources of error and how they would improve their experiment.

Extension: Students can apply their understanding of F = ma to calculate what braking force would be required to stop a car traveling at 60 km/h in 5 s, and compare the results for cars of different masses in a road safety context.

Assessment: Completed data table, graphs and a calculation task (F = ma); teacher observation of the exploration phase.

Materials: ramp, trolley, hanging masses, string, pulley, smartphones, graph paper, real vehicle specification data cards.

Week 4 — Newton's Third Law (Action-Reaction)

STEM Integration: Engineering Design

Engagement: Students blow up a balloon and release it without tying it. Discussion: "What pushes the balloon? Where does the force come from? What pushes back?"

Exploration: Students work in groups to construct balloon-powered vehicles using a plastic straw, balloon, tape, cardboard chassis, and plastic wheels. They test their vehicles on a smooth surface and measure the distance travelled with different balloon sizes and inflation levels.

Explanation: Students use their vehicle observations to articulate the action-reaction pair: the balloon pushes air backward; the air pushes the balloon (and vehicle) forward. The teacher formalises Newton's Third Law and guides students to identify action-reaction pairs in everyday contexts — walking, swimming, rocket propulsion.

Elaboration: Students redesign their balloon vehicles to maximise distance travelled, applying their understanding of Newton's Third Law to their engineering decisions. They document their design changes and predict how each change will affect performance before testing.

Evaluation: Vehicle competition: groups measure the distance their final design travels. Students explain their design decisions in terms of Newton's Third Law. The group achieving the greatest distance presents their reasoning to the class.

Extension: Students research the principle of rocket propulsion and draw a labelled diagram showing the action-reaction forces operating during launch. They explain why rockets work in the vacuum of space where there is no air to push against.

Assessment: Balloon vehicle design documentation; engineering design rubric; verbal explanation evaluated during competition debrief.

Materials: Balloons, plastic straws, tape, cardboard, plastic bottle caps (wheels), smooth testing surface, measuring tape.

Week 5 — Application of Newton's Laws

STEM Integration: STEM + Design Thinking

Engagement: A road safety engineer wants to know which playground surface to choose so that the child who falls is not injured. Discussion: "What forces are involved? Which Newton's Laws apply? Which material/designs would you choose?"

Experiments: By pushing or hauling blocks of the same mass over different surfaces (smooth plastic, sandpaper, carpet, wood, etc.) at constant speed and measuring the force used to maintain that speed by means of a spring balance, students can determine the coefficient of friction of each surface.

Explanation: Students connect the friction data with Newton's First Law (constant velocity means all forces are balanced) and Newton's Second Law (friction stops a moving object). The teacher links the net force concept with situations such as safety features.

Elaboration: Students must design a safety device for a playground surface, a vehicle braking system, or a sports helmet. Their designs must clearly explain how all three Newton's Laws are accounted for, include data from their friction experiments if appropriate, and feature a labeled diagram of the device they have designed.

Evaluation: Groups present safety systems, which are assessed in terms of: accurate application of Newton's Laws, practical application of friction data, feasibility of the safety system, and clarity.

Extension: students research a safety-engineering technology (Formula 1 car crash barriers, earthquake-resistant buildings or aircraft landing gear, for example) and explain the Newton's Laws behind it and how they have been applied in real life.

Assessment: Friction data table; safety system design assessed against a four-criteria rubric; extension paragraph.

Materials: Spring balances, identical masses, samples of surface materials, graph paper, design brief cards.

Week 6 — Motion Graphs and Data Analysis

STEM Integration: Technology + Mathematics

Engagement: Show students two motion graphs, one of constant velocity, the other of constant acceleration (without labels). Discuss, "Which graph shows a greater net force acting? How do you know? Which Newton's Law explains each graph?"

Exploration: Students use smartphone accelerometer applications to capture their own motion. They walk at constant speed, accelerate from rest, and come to an unexpected stop, downloading or taking pictures of their acceleration time graphs.

Explanation: Students compare their shared graphs to the three Newton's Laws. The teacher guides students through relating the slope and area on motion graphs to their quantitative physical interpretations in a direct manner.

Elaboration: Data analysis task: students are given a set of motion data from a real-world context (a vehicle braking test) and are asked to determine the average acceleration, use the equation F = ma to determine the net force acting on the object, and determine whether the braking force is able to stop the vehicle within a certain distance.

Evaluation: Students complete an individual written reflection, "Which of Newton's three laws was most clearly demonstrated in your motion data this week? Use your graphs and calculations as evidence." This also acts as a formative assessment of the ideas learned throughout the six weeks.

Extension: A consolidation poster that uses examples from the six weeks of activities, including one application of engineering design to Newton's three laws of motion in the real world, and one graph/diagram for each law to represent it quantitatively.

Assessment: Copies or photographs of smartphone motion graphs with student annotations; data analysis task; individual written reflection; consolidation poster (assessed with a summary rubric).

Materials: Smartphones with accelerometer apps, printed motion data sets, graph paper, and poster making materials.

Materials and Resources

The following resources were used across the six-week lesson plan:

- Institute of Physics Teaching Advanced Physics, Episode 202: Forces in Equilibrium (adapted for Week 1): https://www.stem.org.uk/resources/library/collection/2972/1
- Science Buddies Physics Activities: https://www.sciencebuddies.org/stem-activities/subjects/physics
- LabXChange Physics Simulations: https://www.labxchange.org/subjects#Physics
- STEAM Powered Family Physics Experiments: https://www.steampoweredfamily.com/category/science-experiments/physics/
- Smartphone accelerometer applications (free, cross-platform)
- Standard physics laboratory equipment: spring balances, ramps, trolleys, masses, pulleys
- Engineering materials: balloons, straws, cardboard, tape, plastic caps

4. DATA ANALYSIS AND RESULTS

4.1 Description of Participants

The study sample comprised 74 ninth-grade students from two intact classes at the State-Owned Institution "Specialized Lyceum named after Al-Farabi" in the Karasai District of Almaty Region, Kazakhstan, during the fall semester of the 2025–2026 academic year (September–December 2025). The 34 students in the experimental group had received 6 weeks of instruction based on the lessons of Newton's Laws integrated to STEM, while 40 students in the control group had received instruction based on the standard 9th-grade physics curriculum in Kazakhstan. There were a total of 36 males and 38 females across conditions. Both classes attended the same school and grade level so we could control for school contextual variables. Table 4.1 illustrates the demographics of both groups.

Table 4.1. *Demographic Composition of the Experimental and Control Groups*

Characteristic	Experimental	Control	Total
n	34	40	74
Male	15	21	36
Female	19	19	38
Grade	9	9	9
Instruction type	STEM-integrated	Traditional	—

Note. Both groups attended the same school during the fall semester of the 2025–2026 academic year.

Furthermore, the two groups were equivalent in the number of participants per group and gender distribution (see Table 4.1). This close equivalence of percentage of males to females in sample (49% female in experimental group and 51% in control group) diminishes the likelihood that group differences seen on the post-test may have resulted from unequal gender distributions. The two intact classes from the same school and the same grade level is an example of purposive non-random sampling, appropriate for quasi-experimental educational research (Shadish, Cook & Campbell, 2002). This was done to control for school-

level confounding variables such as teacher quality, infrastructure, and administrative climate in the current research.

4.2 Data Analysis Strategies

All statistical analyses were conducted using Jamovi (Version 2.7; The jamovi project, 2025). A significance threshold of $\alpha = .05$ was applied uniformly to all inferential tests. For each group comparison, both an independent samples Student's t-test and a Mann-Whitney U test were conducted in parallel, ensuring that conclusions were robust to distributional assumptions. Prior to each t-test, Levene's test for homogeneity of variances was conducted; where non-significant, Student's t with pooled variance was used. Cohen's d was reported as the effect size for t-tests — interpreted as small ($d \geq 0.20$), medium ($d \geq 0.50$), or large ($d \geq 0.80$) following Cohen (1988) — and rank biserial correlation for Mann-Whitney U tests. For the two-way ANCOVA, partial eta squared (η^2) was reported as small ($\eta^2 = .01$), medium ($\eta^2 = .06$), or large ($\eta^2 = .14$). The practical magnitude of conceptual growth was assessed using Hake's (1998) normalised gain: $g = (\text{post-test M} - \text{pre-test M}) / (30 - \text{pre-test M})$, classified as high ($g \geq 0.70$), medium ($0.30 \leq g < 0.70$), or low ($g < 0.30$). Domain-specific effects were examined using Mann-Whitney U tests applied separately to each of the six FCI subscale domains at pre-test and post-test.

4.3 Post-Test Comparison of Groups

To address RQ1, the two groups were compared on FCI total scores at both pre-test and post-test using the independent samples t-test and Mann-Whitney U test. The complete Jamovi output — including the inferential test results, assumption checks, and group descriptives for both time points simultaneously — is presented below. Table 4.2 presents the demographic composition of both groups.

Table 4.2 *Independent Samples T-Test: Pre-Test and Post-Test FCI Total Score Comparison (Experimental vs. Control)*

Independent Samples T-Test

		Statistic	df	p		Effect Size
Pre-test total	**Student's t**	-0.895	72.0	0.374	Cohen's d	-0.209
	Mann-Whitney U	587		0.314	Rank biserial correlation	0.137
Post-test total	**Student's t**	3.511	72.0	<.001	Cohen's d	0.819
	Mann-Whitney U	379		0.001	Rank biserial correlation	-0.443

Note. H_a $\mu_1 \neq \mu_2$

At pre-test, Student's t was non-significant ($t(72) = -0.895$, $p = .374$) and the Mann-Whitney U test confirmed this non-parametrically ($U = 587$, $p = .314$). The negligible effect size (Cohen's $d = -0.209$, rank biserial correlation = 0.137) indicates that any mean difference between groups was trivially small in practical terms. The two groups therefore did not differ meaningfully in Newtonian mechanics conceptual understanding before instruction began. At post-test, Student's t was highly significant ($t(72) = 3.511$, $p < .001$) and the Mann-Whitney U test confirmed this non-parametrically ($U = 379$, $p = .001$). The large Cohen's d (0.819) indicates that the mean post-test difference between groups exceeded 0.8 standard deviations — a large effect by Cohen's (1988) conventions and educationally substantial by the standards of the physics education literature (Hake, 1998). The rank biserial correlation ($r = -0.443$) confirms a medium-to-large non-parametric effect, with the negative sign indicating that experimental group (Group 1) scores were systematically higher than control

group (Group 2) scores in the non-parametric ranking. The convergence of both parametric and non-parametric results at both time points demonstrates that these conclusions are not sensitive to distributional assumptions.

Before interpreting the t-test results further, it was necessary to verify that the equal variances assumption was satisfied at both time points. Table 4.3 presents Levene's test for homogeneity of variances for the pre-test and post-test FCI total scores.

Table 4.3 *Homogeneity of Variances Test (Levene's): Pre-Test and Post-Test FCI Total Score*

Homogeneity of Variances Test (Levene's)

	F	df	df2	p
Pre-test total	1.08	1	72	0.303
Post-test total	1.38	1	72	0.244

Note. A low p-value suggests a violation of the assumption of equal variances

At pre-test, Levene's $F(1,72) = 1.08$, $p = .303$. At post-test, Levene's $F(1,72) = 1.38$, $p = .244$. Both tests are non-significant, confirming that the equal variances assumption was satisfied at both time points and that Student's t with pooled variance was appropriate throughout. The lack of importance at post-test, with the large mean difference found in FCI score, is meaningful. The distributions of the post-test FCI scores about the mean were not considerably different and, thus, the high t-statistic for Table 4.2 was not an artifact of a violation of the parametric assumption.

Because inferential statistics are often computed based on group means, and because the Hake's normalized gain shown below involves group means, the means, medians, standard deviations, and standard errors calculated for each group at each of the two times are shown in Table 4.4.

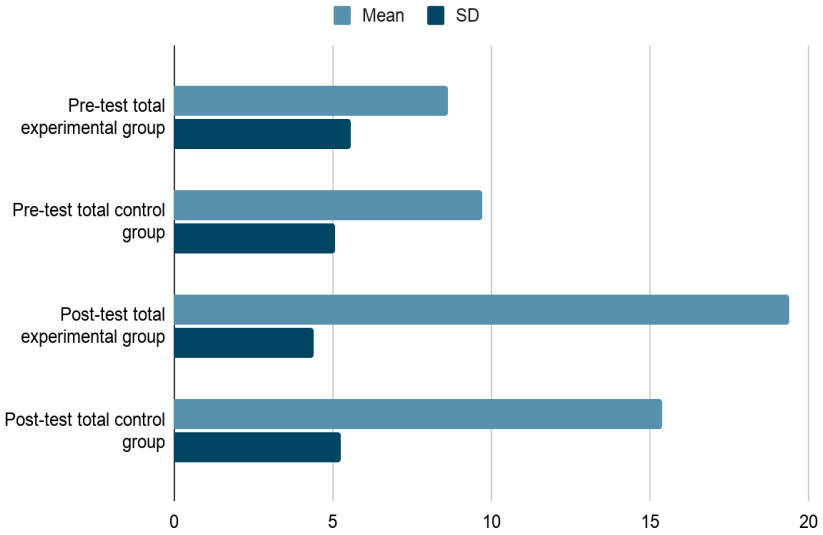
Table 4.4 *Group Descriptives: Pre-Test and Post-Test FCI Total Score by Group*

Group Descriptives

	Group	N	Mean	Median	SD	SE
Pre-test total	1	34	8.62	7.50	5.56	0.953
	2	40	9.72	9.00	5.08	0.803
Post-test total	1	34	19.38	19.50	4.40	0.754
	2	40	15.38	15.00	5.28	0.834

Note. Group 1 = Experimental group (STEM-integrated instruction, n = 34); Group 2 = Control group (Traditional instruction, n = 40). FCI maximum score = 30. SD = Standard Deviation; SE = Standard Error.

Figure 4.1 Mean and Standard Deviation of Pre-Test and Post-Test FCI Total Scores for Experimental and Control Groups



The experimental group (Group 1) had a pre-test mean of M = 8.62 (Median = 7.50, SD = 5.56, SE = 0.953) and the control group (Group 2) had M = 9.72 (Median = 9.00, SD = 5.08, SE = 0.803), only 1.1 points apart on a 30-point scale, both low according to international norms for secondary students before instruction in Newton's Laws (Hestenes et al., 1992). The large standard deviations reflect the within-class heterogeneity that is characteristic of intact secondary school classes. The means and medians were similar across groups; the distributions were sufficiently symmetric to justify the use of parametric analyzes. The experimental groups (M = 19.38, Median = 19.50, SD = 4.40, SE = 0.754) performed statistically considerably better than the control groups (M = 15.38, Median = 15.00, SD = 5.28, SE = 0.834; t (76) = 4.10, P < .001), with an effect size of 4.00 in favor of experimental groups. The standard deviation of the experimental group was reduced from pre-test (M = 19.25, Median = 19.50, SD = 5.56, SE = 0.833), to post-test (M = 19.38, Median = 19.50, SD = 4.40, SE = 0.754), the control group showed no such a similarity, indicating that STEM integrated instruction has reduced the spread of individual scores in the experimental group, thus, increased the concept homogeneity of high and low achievers.

From the values in Table 4.3, Hake's (1998) normalised gain was calculated to quantify the proportional conceptual growth achieved by each group relative to its individual ceiling. The experimental group achieved g = (19.38 − 8.62) / (30 − 8.62) = **0.503** — classified as a **medium gain**. The control group achieved g = (15.38 − 9.72) / (30 − 9.72) = **0.279** — classified as a **low gain**.

Even with these differences in the pretest values controlled for, the experimental group gained close to half the maximum possible gain in Newtonian conceptual understanding, whereas the control group gained less than a third of the maximum possible gain. Tables 4.1, 4.2, and 4.3 provide converging evidence that the six weeks of STEM-integrated instruction in this study produced a statistically meaningful, large, and practically meaningful difference between experimental and control groups on FCI-measured Newtonian mechanics conceptual understanding.

4.4 Main Analysis: Effect of STEM Instruction (ANCOVA)

While Section 4.3 confirms a significant and large post-test group difference, it does not rule out the possibility that this difference reflects residual pre-existing performance variation or differential gender composition. To address both possibilities simultaneously, a two-way ANCOVA was conducted with post-test FCI total score as the dependent variable, group and gender as fixed factors, and pre-test FCI total score as a covariate. Before interpreting the ANCOVA model, the homogeneity of variances assumption across all group-gender cells must be verified. Table 4.5 presents the full ANCOVA results, and Table 4.6 presents the corresponding assumption check, both exactly as produced by Jamovi.

Table 4.4 presents the sum of squares, degrees of freedom, mean square, F-ratio, p-value, and partial eta squared for each source of variance in the ANCOVA model — the pre-test covariate, gender main effect, group main effect, Gender × Group interaction, and residual error — with the dependent variable being post-test FCI total score.

Table 4.5 *ANCOVA — Post-Test FCI Total Score: Effects of Group, Gender, and Pre-Test Covariate*

ANCOVA - Post-test total

	Sum of Squares	df	Mean Square	F	p	η^2
Pre-test total	0.735	1	0.735	0.0295	0.864	0.000
Gender	0.401	1	0.401	0.0161	0.899	0.000
Group	287.718	1	287.718	11.5328	0.001	0.143
Gender ✻ Group	0.765	1	0.765	0.0307	0.861	0.000
Residuals	1721.396	69	24.948			

The pre-test total covariate was non-significant ($F(1,69) = 0.0295$, $p = .864$, $\eta^2 = .000$), indicating that students' initial FCI level did not independently predict post-test scores once group membership was accounted for. This finding suggests that STEM-integrated instruction was similarly effective regardless of students' incoming conceptual level — consistent with Vygotsky's (1978) Zone of Proximal Development, which predicts that scaffolded, collaborative learning can bring students from different initial levels to comparable final understanding. Gender produced a non-significant main effect ($F(1,69) = 0.0161$, $p = .899$, $\eta^2 = .000$), indicating that male and female students did not differ in post-test performance after controlling for pre-test scores and group. The Group factor produced a highly significant main effect with a large effect size ($F(1,69) = 11.5328$, $p = .001$, $\eta^2 = .143$), confirming that the post-test advantage documented in Section 4.3 cannot be explained by pre-existing FCI differences or gender composition — instructional condition was the dominant and independent predictor of post-test conceptual understanding. Group status alone accounts for a large effect in moderate sample- sizes by common conventions (Cohen, 1988). The Gender × Group interaction ($F(1,69) = 0.0307$, $p = .861$, $\eta^2 = .000$) was not meaningful. STEM-integrated instruction was equally helpful for both genders.

The assumption of homogeneity of variances across all four group-gender cells must be established before acceptance of the ANCOVA results. Table 4.5 follows.

Table 4.6 *ANCOVA Assumption Check: Homogeneity of Variances Test (Levene's) Across Group-Gender Cells*

Homogeneity of Variances Test (Levene's)			
F	df1	df2	p
0.680	3	70	0.567

Levene's test was computed across the four group-gender cells. The variance of the post-test FCI scores is not considerably different across the four subgroups (experimental males, experimental females, control males, control females) ($F(3,70) = 0.680$, $p = .567$). Thus, the assumption of homogeneity of variance is met. This satisfies the homogeneity of variances assumption of the ANCOVA model and indicates that the peculiarities of Table 4.4's F-statistics are not a result of heteroscedasticity across the group-gender combinations.

4.5 Gender Differences

In answer to RQ2, independent samples t-tests and Mann-Whitney U tests explored whether there was a difference between female and male total pre-test and post-test FCI scores in the complete sample. These results are presented as part of the ANCOVA results outlined in Section 4.4. On the pre-test measure, females had considerably higher scores than males in the full sample ($t(72) = -2.007$, $p = .049$, Cohen's $d = -0.467$; Mann-Whitney $U = 492$, $p = .038$), indicating a small to medium gender difference at baseline.

By post-test, this gap had entirely disappeared — the difference was negligible and non-significant ($t(72) = -0.122$, $p = .903$, Cohen's $d = -0.028$; Mann-Whitney $U = 669$, $p = .875$).

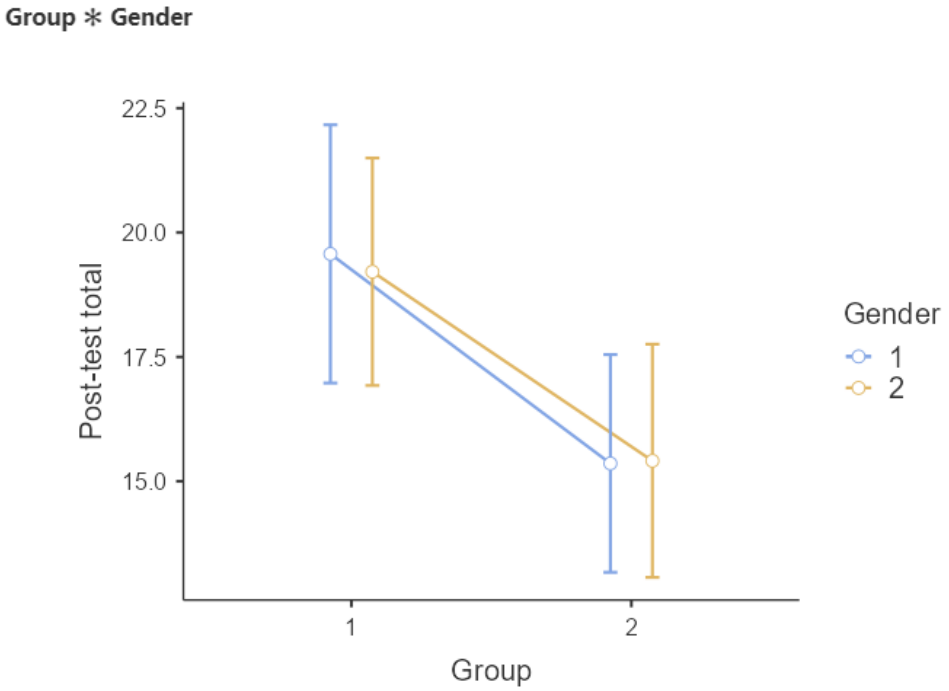
Additionally, the fact that the Gender x Group interaction term in the ANCOVA did not reach importance ($p = .861$, $\eta^2 = .000$) confirms that this convergence in male and female performance was not driven by the instructional group: STEM-integrated instruction was helpful to males and females equally, and the closing of the gender gap from the pre-test is due to conceptual gains overall.

4.6 Visual Analysis of Results

In addition, to visualize the results of the ANCOVA with group and gender as factors, the estimated marginal means plot in figure 4.2 was created directly by the Jamovi software during the ANCOVA with group and gender as factors. This plot

shows post-test FCI adjusted means for both levels of gender on both levels of the instructional type with 95% confidence intervals, controlling for pre-test FCI score.

Figure 4.2 *Estimated Marginal Means of Post-Test FCI Total Score by Gender and Group (ANCOVA)*



Note. Gender 1 = Male (n = 36); Gender 2 = Female (n = 38). Group 1 = Experimental group (STEM-integrated instruction, n = 34); Group 2 = Control group (Traditional instruction, n = 40). The y-axis represents the estimated marginal post-test FCI total score (maximum = 30), adjusted for pre-test FCI score as a covariate. Error bars represent 95% confidence intervals.

4.7 Mann-Whitney U Subscale Analysis by Domain

Having established the overall and ANCOVA-adjusted group effect in Sections 4.3 and 4.4, and having confirmed equitable outcomes across gender in Sections 4.5 and 4.6, the analysis turns to RQ3: which specific Newtonian mechanics domains

showed the greatest conceptual gains following STEM-integrated instruction. To address this, Mann-Whitney U tests were conducted for each of the six FCI subscale domains at both pre-test and post-test. The non-parametric Mann-Whitney U test was selected for subscale analysis given the ordinal nature and restricted score range of individual subscale scores relative to the full 30-item total. Table 4.7 presents all twelve comparisons — six domains at two time points — together in a single Jamovi output section, exactly as produced, covering pre-test baseline equivalence at the domain level and post-test domain-specific instructional effects simultaneously.

Table 4.7 *FCI Subscale Domain Analysis: Mann-Whitney U Tests at Pre-Test and Post-Test (Experimental vs. Control)*

Independent Samples T-Test

		Statistic	p
Kinematics_pre-test	Mann-Whitney U	556	0.170
Kinematics_post-test	Mann-Whitney U	514	0.054
First law_pre-test	Mann-Whitney U	661	0.833
First law_post-test	Mann-Whitney U	503	0.049
Second law_pre-test	Mann-Whitney U	665	0.862
Second law_pro-test	Mann-Whitney U	521	0.069
Third law_pre-test	Mann-Whitney U	585	0.282
Third low_post-test	Mann-Whitney U	358	<.001
Superposition Principle_pre-test	Mann-Whitney U	628	0.548
Superposition Principle_post-test	Mann-Whitney U	592	0.326
Kinds of Force_pre-test	Mann-Whitney U	640	0.663
Kinds of Force_post-test	Mann-Whitney U	392	0.001

Note. $H_a\ \mu_1 \neq \mu_2$

The pre-test subscale results show that all six domains produced non-significant between-group differences before the intervention: Kinematics (U = 556, p = .170), First Law (U = 661, p = .833), Second Law (U = 665, p = .862), Third Law (U = 585, p = .282), Superposition Principle (U = 628, p = .548), and Kinds of Force (U = 640, p = .663). This uniform non-significance extends the total-score baseline equivalence established in Section 4.3 to the domain level, confirming that the two groups entered the intervention at comparable levels of understanding within each specific Newtonian mechanics domain — providing a rigorous basis for interpreting post-test subscale differences as effects of the differential instructional conditions rather than pre-existing domain-specific knowledge gaps.

The post-test subscale results reveal a theoretically coherent and pedagogically interpretable pattern that extends and enriches the total-score findings from Sections 4.3 and 4.4. Three domains produced statistically significant differences favouring the experimental group: First Law (U = 503, p = .049), Third Law (U = 358, p < .001), and Kinds of Force (U = 392, p = .001). Two domains produced marginal trends below the significance threshold: Kinematics (U = 514, p = .054) and Second Law (U = 521, p = .069). One domain showed no meaningful between-group difference: Superposition Principle (U = 592, p = .326). The three significant domains correspond precisely to the FCI areas most directly targeted by the specific engineering design activities in the STEM lesson cycle. Newton's Third Law produced the strongest effect in the entire subscale analysis (p < .001): the balloon-powered vehicle activity in Week 4 created direct, unambiguous physical disconfirmation of the most persistent Third Law misconception — the belief that a heavier object exerts more force than a lighter one in an interaction — one of the most deeply embedded alternative conceptions documented in the literature (Hestenes et al., 1992).The Kinds of Force result (p = .001) reflects the effectiveness of the ramp-and-force-identification tasks in Weeks 1 and 5 in which students were directly measuring, naming, and classifying forces at surfaces, in contrast to the abstract mathematical treatment of force in customary instruction. The result from First Law (p = .049) reinforced the Week 2 inertia

demonstrations, where students physically touched, grasped, and moved common objects to feel them resisting changes to their state of motion. The non-significant Superposition of Forces result ($p = .326$) is theoretically interpretable: vector force decomposition requires the formal integration of conceptual understanding with mathematical vector skills at a level that six weeks of design-based STEM activities cannot fully develop in isolation. The marginal results for Kinematics ($p = .054$) and Second Law ($p = .069$) suggest meaningful trends that may reach significance with a larger sample or a longer intervention duration.

5. DISCUSSION

5.1 Discussion of RQ1: Effect of STEM-Integrated Instruction on Conceptual

The experimental group outperformed the control group on the FCI post-test with a large effect size ($d = 0.819$, $p < .001$) that remained significant after controlling for pre-test scores and gender in the ANCOVA ($\eta^2 = .143$). The experimental group achieved a medium normalised gain ($g = 0.503$) compared to a low gain for the control group ($g = 0.279$).These findings are in agreement with Becker and Park (2024) who found a large meta-analytic effect of STEM integration on secondary science achievement ($g = 0.661$), and with Chen et al. (2025) who found large cognitive effects of full-framework STEM implementation, the version of STEM implemented in the present research. The normalized gain values we found are in agreement with Hake (1998) who reported customary instruction led to an effect size of g in the range of 0.20-0.30, while interactive-engagement led to an effect size of g between 0.40-0.60.

Theoretical explanation of this finding is Posner et al.'s (1982) Conceptual Change Theory: Kazakhstani physics lessons failed to induce cognitive conflict because students could apply formulae without rejecting Aristotelian intuitions about force and motion. Three aspects of the STEM lesson cycle addressed this: the Engagement phase activated prior misconceptions, the Exploration phase produced disconfirming data, and the Elaboration engineering design tasks made conceptual understanding a necessity (not a choice) (Kelley and Knowles, 2016). All five principles proposed by Thibaut et al. (2018) were operationalised in the lesson cycle, with Zeng and Yao (2025) identifying it as the main moderator of STEM effectiveness magnitude.

FCI gain was proven meaningful in Indonesian secondary school students with STEM intervention, as shown by Francisca et al. (2021), and with convergent parametric and non-parametric analyzes as shown by Asrizal et al. (2023), approaches that also contribute to the findings of the present study.

5.2 Discussion of RQ2: Gender Differences

At the time of the pre-test, girls outperformed boys ($d = 0.467$, $p = .049$), the reverse of the usual gender outcome in Citro et al. (2022) and other western FCI studies. The opposite gender outcome, however, was consistent with a meaningful female pre-test advantage in the study by Japashov et al. (2024) using the Kazakhstani FCI, and the findings on academic dispositions of boys and girls in the highly routinized and examination-driven culture of the post-Soviet secondary school system.

The gender effect had disappeared at post-test ($d = 0.028$, $p = .903$) and the ANCOVA did not find an important Gender x Group interaction ($p = .861$, $\eta^2 = .000$), suggesting that integrated STEM interventions work equally well for both genders. This was also confirmed in the Guzey and Li (2023) longitudinal study, which showed no gender difference three years after integrated STEM units were delivered. The collaborative, inquiry-based, and consequential nature of the engineering design tasks meant that the prior informal mechanics experience advantage male learners usually receive in customary delivery of mechanics instruction (McCullough, 2004; Beroíza-Valenzuela & Salas-Guzmán, 2024) was overcome. The results are compelling classroom-level evidence in countries such as Kazakhstan, where national STEM education policy is motivated by gender equity goals (Zhumabay et al., 2024; Abdrakhmanova et al., 2025), that gender equity can be achieved through STEM integration without differential treatment of gender in the curriculum.

5.3 Discussion of RQ3: Domain-Specific Conceptual Gains

For the FCI's subscale domains, the experimental group had an important difference on the post-test compared to the control group for Newton's Third Law ($p < .001$), Kinds of Force ($p = .001$), and Newton's First Law ($p = .049$). No meaningful differences were found for Superposition of Forces ($p = .326$), Kinematics ($p = .054$, marginal), and Newton's Second Law ($p = .069$, marginal).

This all makes sense as well. The three major topics line up with the three engineering design tasks carefully targeting those topics: construction of balloon

rockets to show Third Law action-reaction pairs, ramp and force identification tasks for Kinds of Force, and inertia demonstrations to establish the validity of the First Law. For all three cases the activity allowed a prediction-observation-explanation cycle in the mode of conceptual accommodation (Posner et al., 1982). Newton's Third Law showed the greatest effect because the misconception that an object of greater mass has to have a greater force was the most resistant to change. The balloon vehicle activity provided immediate, unambiguous, physical disconfirmation of this belief.

Albeit non-important, the Superposition of Forces result is theoretically interpretable: vector force decomposition requires the formal mathematical integration of conceptual understanding that six weeks of design-based engineering activities cannot provide. This mathematical scaffolding cannot be offloaded onto students through design-based engineering tasks, as qualitative engineering practices cannot replace mathematics concepts (Alarabi et al., 2022). This marginal Second Law result suggests a disconnect between the relatively straightforward numbers students were calculating and the more involved, multi-force nature of the FCI items in the domain.

5.4 Theoretical Implications

This domain specific pattern of results adds to the Thibaut et al. (2018) model by providing suggestive evidence that design-based learning is more effective for qualitative and represented properties but less effective for mathematically abstract properties requiring formal vector reasoning. The fact that the ANCOVA did not yield a meaningful effect for the pre-test covariate ($p = .864$) could also imply that STEM integration decreases the extent to which differences in ability predict differences in physics learning outcomes, as suggested by Vygotsky (1978)'s Zone of Proximal Development. Since both conclusions are correct, the most fruitful approach would be for STEM activities to develop the cognitive conflict and, subsequently, authentic application experiences that are integral to both Posner et al.'s (1982) Conceptual Change Theory and Ausubel's (1968) Meaningful Learning Theory to ease conceptual change.

5.5 Practical Implications for Kazakhstan

At the classroom level, the teacher's guide (Appendix A) provides a replicable, curriculum-aligned STEM instructional model that does not require specialist training — the barrier Balta et al. (2024) and Bakytkazy et al. (2025) identify as the primary obstacle to STEM adoption among Kazakhstani physics teachers. The large effect achieved by a first-time STEM implementer confirms that structured instructional materials can compensate for the absence of prior STEM pedagogy training.

At the policy level, Kazakhstan's PISA 2022 science proficiency rate (55% at Level 2+) remains 21 percentage points below the OECD average (OECD, 2023), reflecting a systemic deficit in conceptual science understanding. This study provides direct evidence that curriculum-aligned STEM integration produces approximately double the conceptual gains of traditional instruction within the same six-week timeframe. The teacher's guide produced here represents the kind of curriculum-embedded, practice-oriented professional support that Wen et al. (2023) identify as most effective for enabling first-time STEM implementation, and that Zhumabay et al. (2024) identify as the most urgent practical contribution research can make to Kazakhstan's STEM education landscape.

5.6 Recommendations for Future Research

Multi-school, multi-teacher designs with larger samples are needed to confirm the generalisability of these findings across diverse Kazakhstani school types and regions. Longitudinal designs examining whether gains are retained and whether cumulative STEM exposure produces larger effects over multiple academic years would address the duration limitation. Mixed-methods designs incorporating formal fidelity measures, classroom observation, and student reflection data would clarify the specific lesson design features responsible for conceptual change. Future STEM lesson cycles for Newton's Laws should incorporate more explicit vector mathematics scaffolding for the Superposition of Forces domain, and research comparing lesson

designs with and without digital simulation components would clarify the contribution of virtual laboratory tools to FCI gains in mathematically demanding domains.

6. LIMITATIONS

There are five limitations to the present study that the authors would like to bring to the reader's attention.

Design limitation. The quasi-experimental non-equivalent pretest-posttest design with two intact classes does not permit random assignment to conditions within or between classes. Despite statistically similar pre-test scores across all FCI subscales, time-varying unobserved confounding variables (such as teacher effectiveness, classroom goal structures, and changes in peer group dynamics) cannot be discounted as alternative explanations for statistically important differences in post-test scores. The presence of quasi-experimental controls increases internal validity but does not ensure it.

Sample limitations. This research took place at a single school, the Specialized Lyceum named after Al-Farabi in the Karasai District of the Almaty Region, over a single semester. While the findings describe results in one institutional context, they should not be extrapolated to urban public schools, rural schools, schools in other areas in Kazakhstan, or other types of schools without further independent research.

Implementation fidelity limitation. No observation protocol or fidelity checklist could be implemented during the 6-week intervention. Because the teacher was implementing STEM-integrated instruction for the first time, and because of the differences between the designed and implemented lesson activities, these effect sizes may understate the impact a fully trained, experienced STEM educator can have delivering this same lesson plan.

Measurement limitation. The FCI was the sole dependent variable. While the most extensively validated instrument for Newtonian mechanics conceptual understanding, it captures one dimension of learning. Other outcomes theorised to benefit from STEM integration — engineering design competence, scientific reasoning, collaborative skills, STEM career interest, and attitudinal change — were not measured and may represent additional unmeasured benefits of the experimental condition that are invisible in the present data.

Duration limitation. The six-week intervention captures only the immediate effect of a single curriculum unit. The study cannot determine whether the conceptual gains are retained over time, whether they transfer to related physics content, or whether repeated STEM exposure across multiple units and grade levels would produce cumulative effects beyond those documented here. The non-significant Superposition of Forces result in particular suggests that some FCI domains may require a longer instructional timeframe than six weeks to respond to STEM-integrated approaches.

7. CONCLUSION

This study investigated the effect of six-week STEM-integrated instruction on 9th-grade students' conceptual understanding of Newton's Laws of Motion in a Kazakhstani secondary school, using a quasi-experimental design with 74 students and the Force Concept Inventory as the measurement tool.

The results demonstrated that STEM-integrated instruction produced significantly greater conceptual gains than traditional instruction, with a large effect size ($d = 0.819$, $p < .001$) and a medium normalized gain ($g = 0.503$) for the experimental group. A pre-existing gender gap favouring female students fully closed by post-test, confirming that the approach supports equitable learning outcomes for both genders. Subscale analysis showed the strongest gains in Newton's First and Third Laws and Kinds of Force, while Superposition of Forces required additional mathematical scaffolding beyond what the six-week program provided.

These findings provide the first controlled quasi-experimental evidence from Kazakhstan that curriculum-aligned STEM integration is significantly more effective than traditional physics instruction, while also supporting national gender equity goals in STEM education. Future research should replicate these results across multiple schools using longitudinal mixed-methods designs to further validate and deepen these findings.

8. REFERENCES

Abdrakhmanova, K., Kadirbayeva, R., Kudaibergenova, K., Zharmukhanbetov, S., & Nurmukhanbetova, G. (2025). Formation of STEM competencies of future teachers: Kazakhstani experience. *Open Education Studies, 7*(1), 20240058. https://doi.org/10.1515/edu-2024-0058

Abilkhairova, Z., Ortaeva, G., & Jumassaeva, A. (2025). STEM education impact on the development of primary school students' meta-subject skills: An experimental study in Kazakhstan. *Frontiers in Education, 10*, 1669765. https://doi.org/10.3389/feduc.2025.1669765

Alarabi, K., Tairab, H., Wardat, Y., Belbase, S., & Alabidi, S. (2022). Enhancing the learning of Newton's Second Law of Motion using computer simulations. *Journal of Baltic Science Education, 21*(6), 946–966.

Asrizal, A., Festiyed, F., Ashel, H., & Amnah, R. (2023). STEM-integrated physics digital teaching material to develop conceptual understanding and new literacy of students. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7), em2289. https://doi.org/10.29333/ejmste/13275

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. Holt, Rinehart and Winston.

Bakytkazy, T., Kuralay, N., Nursultan, J., & Nurman, Z. (2025). Examining the pedagogical abilities and needs of Kazak physics teachers to implement STEM education. In M. E. Auer & T. Rüütmann (Eds.), *Futureproofing engineering education for global responsibility. ICL 2024. Lecture Notes in Networks and Systems* (vol. 1261). Springer. https://doi.org/10.1007/978-3-031-85649-5_25

Balta, N., Japashov, N., Bakytkazy, T., Abdikadyr, B., & Nurgaliyeva, K. (2024). High school physics teachers' pedagogical discontentment: The effect of curriculum improvement and professional development programs. *Heliyon, 10*(2), e24274. https://doi.org/10.1016/j.heliyon.2024.e24274

Becker, K., & Park, K. (2024). A meta-analysis of STEM integration on student academic achievement. *Research in Science Education*. https://doi.org/10.1007/s11165-024-10216-y

Beroíza-Valenzuela, F., & Salas-Guzmán, N. (2024). STEM and gender gap: A systematic review in WoS, Scopus, and ERIC databases (2012–2022). *Frontiers in Education, 9*. https://doi.org/10.3389/feduc.2024.1378640

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Rand McNally.

Chen, B., Chen, J., Wang, M., Tsai, C.-C., & Kirschner, P. A. (2025). The effects of integrated STEM education on K-12 students' achievements: A meta-analysis. *Review of Educational Research.* https://doi.org/10.3102/00346543251318297

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50*(1), 66–71.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Coletta, V. P., & Phillips, J. A. (2005). Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. *American Journal of Physics, 73*(12), 1172–1182.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.

Dominguez, A., De la Garza, J., Quezada-Espinoza, M., & Zavala, G. (2024). Integration of physics and mathematics in STEM education: Use of modeling. *Education Sciences*, *14*(1), 20. https://doi.org/10.3390/educsci14010020

English, L. D. (2016). STEM education K–12: Perspectives on integration. *International Journal of STEM Education*, *3*(1), 3.

Francisca, D., Paring, I. R. B., Kodama, K., & Iwayama, T. (2021). STEM integrated approach in improving students' physics conceptual understanding. *Bulletin of Aichi University of Education, 70* (Natural Sciences), 33–41.

Guzey, S. S., & Li, W. (2023). Engagement and science achievement in the context of integrated STEM education: A longitudinal study. *Journal of Science Education and Technology*, *32*(2), 168–180. https://doi.org/10.1007/s10956-022-10023-y

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74.

Halloun, I. A., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics, 53*(11), 1043–1055.

Halloun, I., Hake, R. R., Mosca, E. P., & Hestenes, D. (1995). *Force Concept Inventory* (Revised). Unpublished instrument.

Hasan, S., Bagayoko, D., & Kelley, E. L. (1999). Misconceptions and the certainty of response index (CRI). *Physics Education, 34*(5), 294–299.

Henderson, C. (2002). Common concerns about the Force Concept Inventory. *The Physics Teacher, 40*(9), 542–547.

Hestenes, D., & Halloun, I. (1995). Interpreting the Force Concept Inventory. *The Physics Teacher, 33*(8), 502–506.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher, 30*(3), 141–158.

Honey, M., Pearson, G., & Schweingruber, H. (Eds.). (2014). *STEM integration in K–12 education: Status, prospects, and an agenda for research.* National Academies Press.

Institute of Physics. (2006). *Episode 202: Forces in equilibrium. Teaching Advanced Physics.* STEM Learning. https://www.stem.org.uk/resources/library/collection/2972/1

Japashov, N., Abdikadyr, B., Balta, N., Maxutov, S., Postiglione, A., & Tzafilkou, K. (2024). Analysing the structure of Kazakhstan university undergraduate students' knowledge about the force concept: Findings from a three-tier FCI survey. *Physics Education, 59*(2), 025003. https://doi.org/10.1088/1361-6552/ad1656

Kazu, İ. Y., & Şenol, C. (2022). The effect of STEM education on academic performance: A meta-analysis study. *Turkish Online Journal of Educational Technology, 21*(2), 123–135.

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*(1), 11. https://doi.org/10.1186/s40594-016-0046-z

Kost, L. E., Pollock, S. J., & Finkelstein, N. D. (2009). Characterizing the gender gap in introductory physics. *Physical Review Special Topics — Physics Education Research, 5*(1), 010101.

LabXChange. (n.d.). *Physics simulations.* https://www.labxchange.org/subjects#Physics

Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics, 74*(2), 118–122.

Margot, K. C., & Kettler, T. (2019). Teachers' perception of STEM integration and education: A systematic literature review. *International Journal of STEM Education, 6*(1), 2.

McCullough, L. (2004). Gender, context, and physics assessment. *Journal of International Women's Studies, 5*(4), 20–30.

McDermott, L. C. (1991). Millikan Lecture 1990: What we teach and what is learned — closing the gap. *American Journal of Physics, 59*(4), 301–315.

McDermott, L. C. (2001). Oersted Medal Lecture 2001: Physics education research — the key to student learning. *American Journal of Physics, 69*(11), 1127–1137.

Ministry of Education and Science of Kazakhstan. (2016). *State program for education and science development of the Republic of Kazakhstan for 2016–2019.* Government of Kazakhstan.

Naushabekov, Z., Maxutov, S., Mansurova, A., & Japashov, N. (2025). Effect of project-based learning on middle school students' conceptual understanding of mechanics: A quasi-experimental study. European Journal of Science and Mathematics Education, 13(3), 225–237. https://doi.org/10.30935/scimath/16826

Nazarbayev Intellectual Schools. (2019). *Updated content curriculum: Natural sciences.* Autonomous Education Organization.

Nieminen, P., Savinainen, A., & Viiri, J. (2010). Force Concept Inventory-based multiple-choice test for investigating students' representational consistency. *Physical Review Special Topics — Physics Education Research, 6*(2), 020109.

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education.* OECD Publishing. https://doi.org/10.1787/53f23881-en

Piaget, J. (1964). Cognitive development in children: Development and learning. *Journal of Research in Science Teaching, 2*(3), 176–186. https://doi.org/10.1002/tea.3660020306

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211–227.

Redish, E. F. (2003). *Teaching physics with the physics suite.* Wiley.

Roehrig, G. H., Dare, E. A., Ring-Whalen, E., & Wieselmann, J. R. (2021). Understanding coherent and comprehensive models of integrated STEM education. *Disciplinary and Interdisciplinary Science Education Research, 3*(1), 11.

Science Buddies. (n.d.). *Physics activities.* https://www.sciencebuddies.org/stem-activities/subjects/physics

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Houghton Mifflin.

STEAM Powered Family. (n.d.). *Physics experiments.* https://www.steampoweredfamily.com/category/science-experiments/physics/

The jamovi project. (2025). *Jamovi* (Version 2.7) [Computer software]. https://www.jamovi.org

Thibaut, L., Ceuppens, S., De Loof, H., De Meester, J., Goovaerts, L., Struyf, A., & Depaepe, F. (2018). Integrated STEM education: A systematic review of instructional practices in secondary education. *European Journal of STEM Education, 3*(1), 2.

Uzun, S., & Şen, N. (2023). The effects of a STEM-based intervention on middle school students' science achievement and motivation. *Journal of Pedagogical Research, 7*(1), 228–242. https://doi.org/10.33902/JPR.202315486

von Davier, M., Kennedy, A., Reynolds, K., Fishbein, B., Khorramdel, L., Aldrich, C., Bookbinder, A., Bezirhan, U., & Yin, L. (2024). *TIMSS 2023 international results in mathematics and science.* Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.timss.rs6460

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4*(1), 45–69.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Wen, X., Walters, S. M., & Tao, Y. (2023). In-service STEM teachers professional development programmes: A systematic literature review 2018–2022. *Computers & Education, 199,* 104790. https://doi.org/10.1016/j.compedu.2023.104790

Yaki, A. A. (2022). Fostering critical thinking skills using integrated STEM approach among secondary school biology students. *European Journal of STEM Education, 7*(1), 06. https://doi.org/10.20897/ejsteme/12481

Zeng, S., & Yao, J. (2025). Systematic review and meta-analysis of the impact of STEM education on students' learning outcomes. *Frontiers in Psychology*, *16*, 1579474. https://doi.org/10.3389/fpsyg.2025.1579474

Zhumabay, N., Varis, S., Abylkassymova, A., Balta, N., Bakytkazy, T., & Bowen, G. M. (2024). Mapping the Kazakhstani STEM education landscape: A review of national research. *European Journal of STEM Education*, *9*(1), 16. https://doi.org/10.20897/ejsteme/15576

9. APPENDIX

APPENDIX A: Three tier FCI

Ethical Approval Note. Participation is voluntary and anonymous.No identifiable information is collected. All data are used solely for research purposes.

Part I: Demographics

Сіз оқитын мекеме мектеп

Қай сыныпта _____

Жынысыңызды көрсетіңіз ер/әйел_____

Болашақ мамандығыңызды көрсетіңіз_____

Өткен жылғы физикадан алған жылдық бағаңыз қанша_____

Part II: FCI for students

1. Өлшемдері бірдей екі металл шардың біреуі екіншісінен екі есе ауыр. Екі шар бір мезетте бір қабатты үйдің төбесінен құлады. Шарлардың жерге құлауына қажетті уақыт:

A) Ауыр шар жеңіл шардан екі есе аз уақытта түседі.

B) Жеңіл шар ауыр шардан екі есе аз уақытта түседі.

C) Екеуі бірдей уақытта түседі.

D) Ауыр шар жеңіл шардан бірінші түседі, бірақ екі есе болуы міндетті емес.

E) Жеңіл шар ауыр шардан бірінші түседі, бірақ екі есе болуы міндетті емес.

Жоғарыда берген жауабыңызға (екеуінің бірін таңдаңыз):

A) Сенімдімін B) Сенімді емеспін

1 ші сұраққа берген жауабыңыздың физикалық тұрғыдағы түсіндірмесін төменге жазыңыз.

_____

_____

2. Алдыңғы тапсырмадағы шарларды горизонталь үстелден бірдей жылдамдықпен домалатып жіберді. Бұл жағдайда:

A) Екі шар да еденге үстелден шамамен бірдей қашықтықта түседі

B) Ауыр шар жеңіл шарға қарағанда үстелге екі есе жақын түседі

C) Жеңіл шар ауыр шарға қарағанда үстелге екі есе жақын түседі

D) Ауыр шар жеңіл шарға қарағанда үстелге жақын түседі, бірақ екі есе болуы міндетті емес

E) Жеңіл шар ауыр шарға қарағанда үстелге жақын түседі, бірақ екі есе болуы

міндетті емес

3. Тас жерге бір қабатты үйдің шатырынан құлайды:

A) Құлау басталған соң біраз уақыт ішінде максимал жылдамдыққа жетіп, содан соң бірқалыпты жылдамдықпен құлайды.

B) Үнемі үдемелі қозғалады, себебі жерге жақындаған сайын гравитациялық тартылыс артады.

C) Тұрақты тартылыс күшінің әсерінен үдемелі қозғалады.

D) Барлық денелер Жер бетіндегі табиғи тыныш күйіне ұмтылып құлайды.

E) Тасты төмен қарай тартатын ауырлық күші мен оны ауа тарапынан төмен қарай итеретін күштің қосынды әсерінен құлайды..

4. Үлкен жүк көлігі жеңіл автокөлікпен бетпе-бет соқтығысқан. Соқтығыс кезінде:

A) Жүк көлігінің жеңіл автокөлікке әсер ететін күші автокөліктің жүк көлігіне әсер ететін күшінен көп.

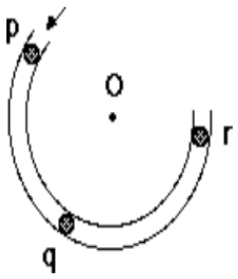
B) Автокөліктің жүк көлігіне әсер ететін күші жүк көлігінің автокөлікке әсер ететін күшінен көп.

C) Жүк көлігі де жеңіл автокөлік те бір-біріне күш түсірмейді, жеңіл автокөлік жүк көлігінің жолында болғандықтан қирайды.

D) Жүк көлігі автокөлікке күш түсіреді, бірақ автокөлік жүк көлігіне күш түсірмейді.

E) Жүк көлігі жеңіл автокөлікке қандай күш түсірсе, автокөлік те жүк көлігіне сондай күшпен әсер етеді.

КЕЛЕСІ ЕКІ СҰРАҚҚА ТӨМЕНДЕ КӨРСЕТІЛГЕН СУРЕТ ПЕН БЕРІЛГЕН ШАРТТЫ ҚОЛДАНЫП ЖАУАП БЕРІҢІЗ (5 және 6-сұрақ)



Берілген суретте центрі «О» нүктесінде болатын шеңбер сегменті пішініндегі үйкеліссіз арна (жол) көрсетілген. Арна үйкеліссіз көлденең үстелге бекітілген. Бақылаушы үстелге жоғарыдан қарап тұр. Ауаның әсер ету күші ескермеуге болатындай аз. Кішкентай шар үлкен жылдамдықпен арнаға "p" нүктесінен кіріп, "r" нүктесінен шығады.

Төмендегі күштерді қарастырыңыз:

1.Төменге бағытталған ауырлық күші.

2. Арна тарапынан әсер ететін, «q» нүктесінен «О» нүктесіне бағытталған күш.

3. Қозғалыс бағыты бойынша күш.

4. «О» нүктесінен «q» нүктесіне бағытталған күш

5. Арнаның ішіндегі "q" нүктесінде осы күштердің қайсысы шарға әрекет етеді?

A) тек 1.

B) 1 және 2.

C) 1 және 3.

D) 1,2 және 3.

E) 1,3 және 4.

6. Арнаның "r" нүктесінен шыққан шардың стол бетімен үйкеліссіз қозғалысы шамамен қай бағытта өрбиді?

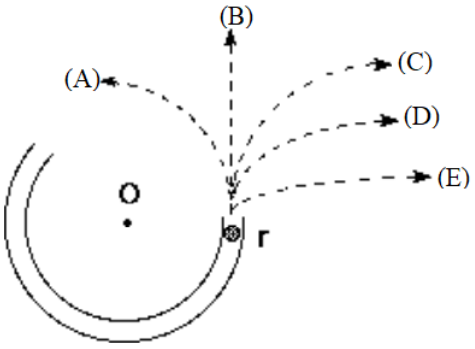
A) A

B) B

C) C

D) D

E) E



7. Бала жіпке байланған болат шарды суретте көрсетілгендей горизонталь жазықтық бойымен айналдырады. Кенет P нүктесінде жіп шарға бекітілген нүктесінен үзіледі. Егер бақылаушы суреттегідей жоғарыдан қарап тұрса, жіп үзілгеннен кейін шардың ұшу бағытын шамамен анықтаңыз

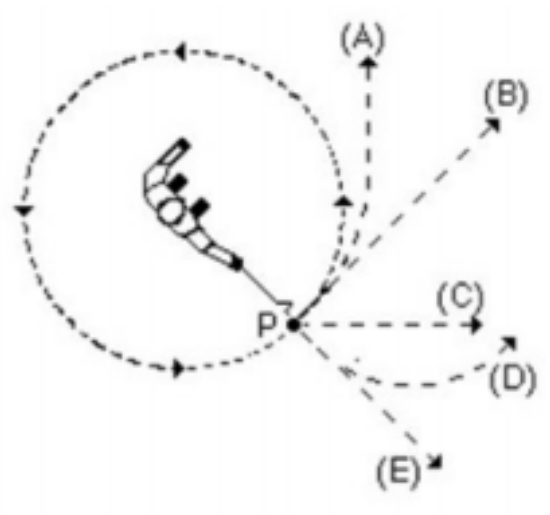
A) A

B) B

C) C

D) D

E) E



КЕЛЕСІ ТӨРТ СҰРАҚҚА ТӨМЕНДЕ КӨРСЕТІЛГЕН СУРЕТ ПЕН БЕРІЛГЕН ШАРТТЫ ҚОЛДАНЫП ЖАУАП БЕРІҢІЗ (8 және 11-сұрақ)

Суретте мұз бетімен a нүктесінен b нүктесіне қарай үйкеліссіз бірқалыпты v_0 жылдамдықпен қозғалып бара жатқан шайба көрсетілген (бақылаушы үстінен қарап тұр). Ауаның кедергісі ескермеуге болатындай өте аз. b нүктесіне жеткенде суреттегі бағыт бойымен шайбаны жылдам тебеді. Егер шайба b нүктесінде тыныштықта тұрған болса, онда шайбаны тепкіден кейін ол v_k жылдамдықпен тепкен бағытта қозалатын еді.



8. Тепкеннен кейін шайба шамамен суреттерде көрсетілген бағыттардың қайсысымен қозғалатындығын анықтаңыз

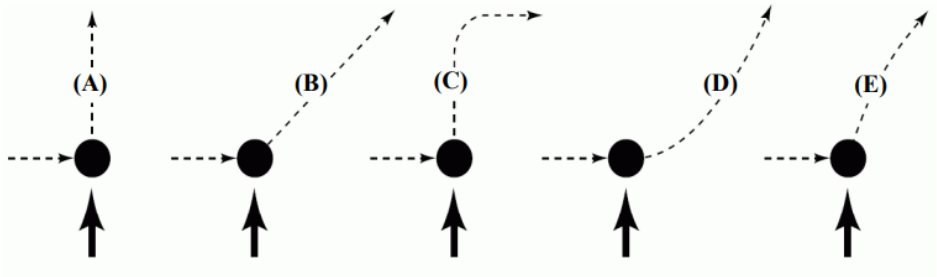
A) A

B) B

C) C

D) D

E) E



9. Алдыңғы тапсырмадағы шайбаның тепкіден кейінгі жылдамдығы:

A) Тепкіге дейінгі қозғалған v_0 жылдамдық модуліне тең.

B) Тебу нәтижесінде шайба модулі бойынша v_k жылдамдықпен қозғалады, және бастапқы v_0 жылдамдыққа тәуелсіз.

C) Модулі бойынша v_0 және v_k жылдамдықтарының арифметикалық қосындысына тең

D) Модулі бойынша v_0 және v_k жылдамдықтарының әрқайсысынан кіші

E) Модулі бойынша v_0 және v_k жылдамдықтарының әрқайсысынан үлкен, бірақ екі жылдамдықтың арифметикалық қосындысынан аз.

10. 8 тапсырмадағы шайбаның тепкіден кейінгі мұз бетіндегі үйкеліссіз қозғалысының жылдамдығы:

A) Тұрақты.

B) Үнемі артады.

C) Үнемі кемиді.

D) Басында артады, содан соң кемиді.

E) Басында тұрақты, содан соң кемиді.

11. 8 сұрақтағы үйкеліссіз жолда тепкіден кейін шайбаға әсер ететін күш(тер):

A) төмен қарай бағытталған ауырлық күші.

B) төмен қарай бағытталған ауырлық күші және қозғалыс бағытында әсер ететін горизонталь күш.

C) төмен қарай бағытталған ауырлық күші, жазықтық тарапынан жоғары қарай әсер ететін күш және қозғалыс бағытында әсер ететін горизонталь күш.

D) төмен қарай бағытталған ауырлық күші, жазықтық тарапынан жоғары қарай әсер ететін күш

E) Шайбаға ешқандай күш әсер етпейді.

12. Суретте көрсетілгендей жартас төбесінен зеңбіректен оқ атылды. Атылған оқтың жолын жуықтап көрсетіңіз

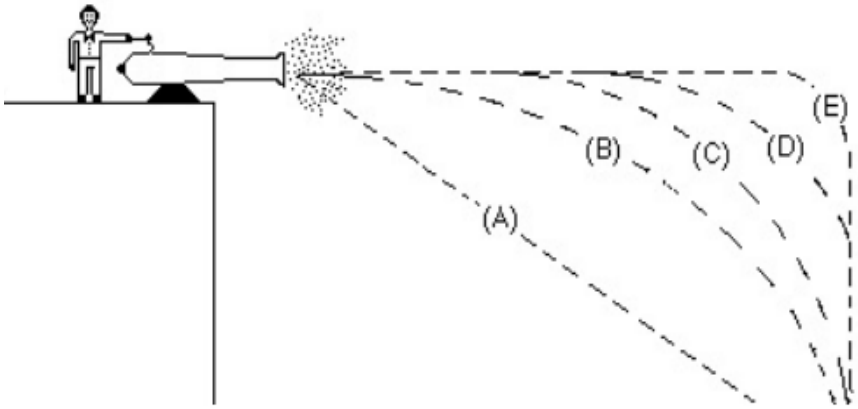
A) A

B) B

C) C

D) D

E) E



13. Бала болат шарды тік жоғары лақтырды. Шардың баланың қолынан шыққаннан бастап жерге түсерге дейінгі қозғалысын қарастырыңыз. Ауаның кедергі күшін ескермеуге болады. Осы жағдайда шарға келесідей күштер әсер етеді:

A) төмен қарай бағытталған ауырлық күші және жоғары қарай бағытталған үнемі кемитін күш.

B) Баланың қолынан шыққаннан бастап максималды биіктікке жеткенге дейін жоғары қарай бағытталған үнемі кемитін күш; жерге жақындаған сайын үнемі артып отыратын төмен қарай бағытталған күш, төмен қарай бағытталған ауырлық күші.

C) төмен қарай бағытталған шамамен тұрақты ауырлық күші, максимал биіктікке жеткенге дейін жоғары қарай бағытталған үнемі кемитін күш; төмен қарай бағытта тек қана ауырлық күші.

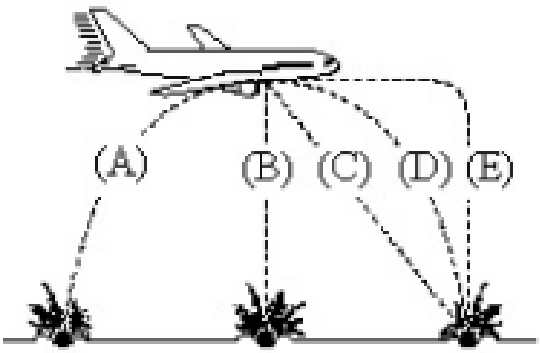
D) төмен қарай бағытталған шамамен тұрақты ауырлық күші.

E) Шарға ешқандай күш әсер етпейді, шар жерге табиғи орнықтылығын сақтау үшін құлайды.

14. Горизонталь бағытта ұшып бара жатқан ұшақтан боулинг шары (үлкен әрі ауыр шар) байқаусызда құлап кетті.

Ұшаққа суреттегідей қарап тұрған жер бетіндегі бақылаушыға шар қандай траекториямен құлайтындығын анықтаңыз.

A) A
B) B
C) C
D) D
E) E



КЕЛЕСІ ЕКІ СҰРАҚҚА ТӨМЕНДЕ КӨРСЕТІЛГЕН СУРЕТ ПЕН БЕРІЛГЕН ШАРТТЫ ҚОЛДАНЫП ЖАУАП БЕРІҢІЗ (15 және 16-сұрақ)

Жолда үлкен жүк көлігі бұзылып қалды, оны суретте көрсетілгендей шағын көлік итеріп бара жатыр.



15. Жүк көлігін итеретін көлік үдеу ала бастағанда:

A) Автокөліктің жүк көлігіне түсіретін күші, жүк көлігінің автокөлікке түсіретін күшіне тең.

B) Автокөліктің жүк көлігіне түсіретін күші, жүк көлігінің автокөлікке түсіретін күшінен кіші.

C) Автокөліктің жүк көлігіне түсіретін күші, жүк көлігінің автокөлікке түсіретін күшінен үлкен.

D) Жүк көлігін итеруші автокөліктің моторы жұмыс істеп тұр, сондықтан автокөлік жүк көлігіне күш түсіреді. Ал жүк көлігінің моторы жұмыс істемейді, сондықтан жүк көлігі автокөлікке күш түсіре алмайды. Жүк көлігі жеңіл автокөліктің жолында тұрғандықтан ғана алға қарай қозғалып келеді.

E) Автомобиль де, жүк көлігі де бір-біріне күш түсірмейді. Жүк көлігі жеңіл автокөліктің жолында тұрғандықтан ғана алға қарай қозғалып келеді.

16. Жүк көлігін итеруші автомобиль белгілі бір тұрақты жылдамдыққа жеткеннен кейін:

A) Автокөліктің жүк көлігіне түсіретін күші, жүк көлігінің автокөлікке түсіретін күшіне тең.

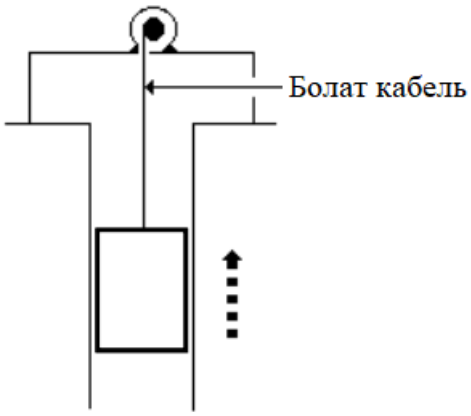
B) Автокөліктің жүк көлігіне түсіретін күші, жүк көлігінің автокөлікке түсіретін күшінен кіші.

C) Автокөліктің жүк көлігіне түсіретін күші, жүк көлігінің автокөлікке түсіретін күшінен үлкен.

D) Жүк көлігін итеруші автокөліктің моторы жұмыс істеп тұр, сондықтан автокөлік жүк көлігіне көлік күш түсіреді. Ал жүк көлігінің моторы жұмыс істемейді, сондықтан жүк көлігі автокөлікке күш түсіре алмайды. Жүк көлігі жеңіл автокөліктің жолында тұрғандықтан ғана алға қарай қозғалып келеді.

E) Автомобиль де, жүк көлігі де бір-біріне күш түсірмейді. Жүк көлігі жеңіл автокөліктің жолында тұрғандықтан ғана алға қарай қозғалып келеді.

17. Болат кабель суретте көрсетілгендей шахтадағы лифтті тұрақты жылдамдықпен көтереді. Барлық үйкеліске қатысты әсерлер елеусіз аз. Бұл жағдайда лифтке әсер ететін күштер:



A) Кабельдің жоғары қарай тарту күші, төмен қарай бағытталған ауырлық күшінен үлкен.

B) Кабельдің жоғары қарай тарту күші, төмен қарай бағытталған ауырлық күшіне тең.

C) Кабельдің жоғары қарай тарту күші, төмен қарай бағытталған ауырлық күшінен аз.

D) Кабельдің жоғары қарай тарту күші төмен қарай бағытталған ауырлық күші мен ауа тарапынан төмен қарай әсер ететін күштің қосындысынан үлкен.

E) Жоғарыдағы жауаптардың ешқайсысы дұрыс емес. (Лифт кабельден түсірілетін күштің әсерінен емес, ол тек кабелдің қысқаруына байланысты жоғары көтеріледі).

18. Суретте баланың «А» нүктесінен жоғары нүктеден бастап арқанмен тербеліп тұрғаны көрсетілген.



Төмендегі күштерді қарастырыңыз:

1.Төмен қарай бағытталған ауырлық күші.

2. Арқанның «А»-дан «О»-ға қарай бағытталған тарту күші.

3. Баланың қозғалыс бағытындағы күші.

4. «О»-дан «А»-ға қарай бағытталған күш.

Бала «А» нүктесіне жеткенде оған қандай күштер әсер етеді?

A) Тек 1.

B) 1 және 2.

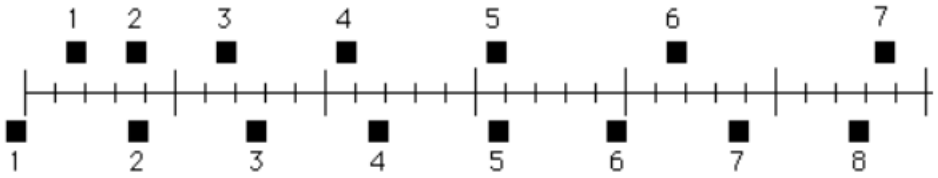
C) 1 және 3.

D) 1,2 және 3.

E) 1,3 және 4.

19. Екі нысанның орны әр 0,2 секунд сайын нөмірленген шаршылармен белгіленген.

Нысандар оңға қарай қозғалып барады.



Бұл нысандардың қозғалыс жылдамдықтары бірдей ме?

A) Жоқ.

B) 2-мезетте бірдей.

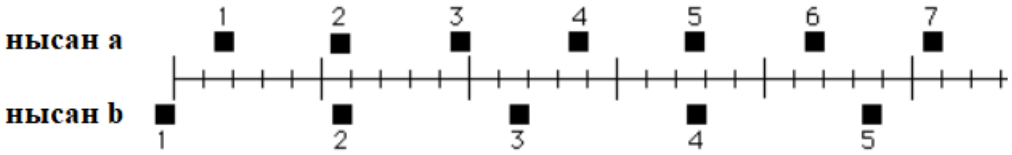
C) 5- мезетте бірдей.

D) 2 және 5- мезеттерде бірдей.

E) 3 пен 4- мезеттердің арасында бірдей.

20. Екі нысанның орны әр 0,2 секунд сайын нөмірленген шаршылармен белгіленген.

Нысандар оңға қарай қозғалып барады.



Бұл нысандардың үдеулері арақатынасы төмендегідей:

A) «a»-ның үдеуі «b» -ның үдеуінен үлкен.

B) «a» -ның үдеуі «b» -ның үдеуімен бірдей. Екі үдеу де нөлден үлкен.

C) «a» -ның үдеуі «b» -ның үдеуінен кіші.

D) «a» -ның үдеуі «b» -ның үдеуімен бірдей. Екі үдеу де нөлге тең.

E) сұраққа жауап беру үшін ақпарат жеткіліксіз.

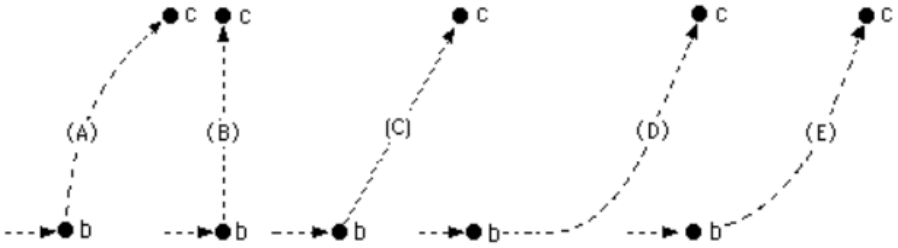
КЕЛЕСІ ТӨРТ СҰРАҚҚА ТӨМЕНДЕ КӨРСЕТІЛГЕН СУРЕТ ПЕН БЕРІЛГЕН ШАРТТЫ ҚОЛДАНЫП ЖАУАП БЕРІҢІЗ (21 мен 24-сұрақтар арасы)

Ғарыш кеңістігінде зымыран суретте көрсетілгендей «a» нүктесінен «b» нүктесіне қарай қозғалады. Зымыранға ешқандай сыртқы күштер әсер етпейді. «b» нүктесінде зымыранның қозғалтқышы іске қосылады. Қозғалтқыш тудыратын тарту күші «ab» сызығына тік бұрыш жасай бағытталған. Тарту күші зымыран «c» нүктесіне жеткенге дейін сақталады.



21. Қай сызық зымыранның «b» нүктесінен «c» нүктесіне дейінгі жолын көрсетеді?

A) A
B) B
C) C
D) D
E) E



22. Зымыран «b» нүктесінен «c» нүктесіне қозғалғанда оның жылдамдығы:

A) тұрақты.

B) үнемі артады.

C) үнемі кемиді.

D) біраз уақытқа дейін артады, содан кейін тұрақты болады.

E) біраз уақытқа дейін тұрақты болады, содан кейін кемиді.

23. «c» нүктесінде қозғалтқыш сөніп, күш бірден нөлге дейін кемиді. Зымыранның «c» нүктесінен бастап жүрген жолын қай сызық көрсетеді?

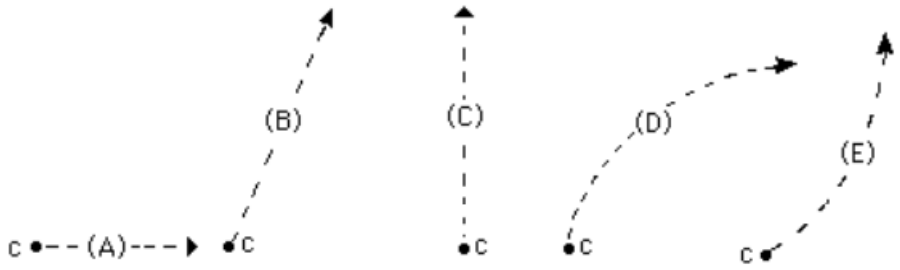
A) A

B) B

C) C

D) D

E) E



24. «c» нүктесінен бастап зымыранның жылдамдығы:

A) тұрақты.

B) үнемі артады.

C) үнемі кемиді.

D) біраз уақытқа дейін артады, содан кейін тұрақты болады.

E) біраз уақытқа дейін тұрақты болады, содан кейін кемиді.

25. Әйел үлкен қорапты (жәшік) горизонталь бағытта тұрақты күшпен итереді. Нәтижесінде қорап горизонталь бағытта еден бетімен тұрақты «v_0» жылдамдықпен қозғалады.

Әйел горизонталь бағытта түсірген тұрақты күші:

A) қораптың салмағына тең.

B) қораптың салмағынан үлкен.

C) қораптың қозғалысына қарсы әсер етуші күштің шамасына тең.

D) қораптың қозғалысына қарсы әсер етуші күштен үлкен.

E) қораптың қозғалысына қарсы әсер етуші күш пен қораптың салмағынан үлкен.

26. Егер алдыңғы сұрақтағы қорапқа әйел горизонталь бағытта екі есе үлкен күшпен әсер етсе, онда қорап:

A) «v_0» жылдамдықтан екі есе көп тұрақты жылдамдықпен қозғалады.

B) «v_0»-дан үлкен, алайда екі есе үлкен болуы міндетті емес, тұрақты жылдамдықпен қозғалады.

C) біраз уақыт «v_0» жылдамдықтан үлкен тұрақты жылдамдықпен қозғалады,

содан кейін жылдамдығы артады.

D) біраз уақыт жылдамдығы артады, содан кейін тұрақты жылдамдықпен қозғалады.

E) үнемі жылдамдығы артатын қозғалыспен қозғалады.

27. Егер 25-сұрақтағы қорапқа әйел горизонталь бағытта әсер ететін күшін кенеттен тоқтатса, онда қорап:

A) бірден тоқтайды.

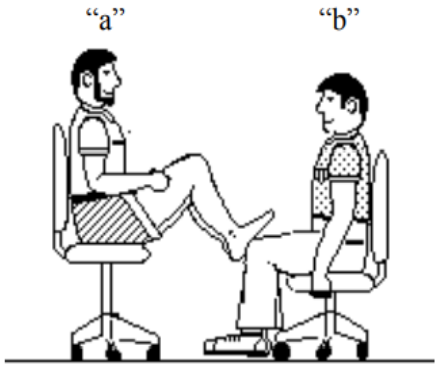
B) тұрақты жылдамдықпен біраз уақыт қозғалады, содан кейін толықтай тоқтағанға дейін жылдамдығы кемиді.

C) бірден баяулай бастайды және толық тоқтағанға дейін жылдамдығы кемиді.

D) тұрақты жылдамдықпен қозғала береді.

E) жылдамдығы біраз уақытқа артады, содан кейін толық тоқтағанша кемиді.

28. Оң жақтағы суреттегі «a» студентінің массасы 95 кг, ал «b» студентінің массасы 77 кг. Олар бірдей дөңгелегі бар орындықта бетпе-бет отыр. «a» студенті аяқтарын суретте көрсетілгендей «b» студентінің тізелеріне қояды да, содан кейін аяғын серпи итереді. Соның салдарынан екі орындық артқа қарай қозғалады.



Студенттер бір-біріне жанасып тұрып серпи итеру кезінде:

A) Олардың ешқайсысы бір-біріне күш түсірмейді.

B) «a» студенті «b» студентіне күш түсіреді, бірақ «b» студенті «a» студентіне күш түсірмейді.

C) әр студент екіншісіне күш түсіреді, бірақ «b» көбірек күш түсіреді.

D) әр студент екіншісіне күш түсіреді, бірақ «a» көбірек күш түсіреді.

E) екі студент бір-біріне модулі бірдей күш түсіреді.

29. Еденде бар бос орындық тұр. Келесі күштерді қарастырыңыз:

1.Төмен қарай бағытталған ауырлық күші.

2. Еден тарапынан әсер ететін, жоғары бағытталған күш.

3. Ауа тарапынан әсер ететін, төмен қарай бағытталған күш.

Орындыққа қандай күштер әсер етеді?

A) Тек 1.

B) 1 және 2.

C) 2 және 3.

D) 1,2 және 3.

E) Ешқандай күш әсер етпейді, себебі орындық қозғалыста емес.

30. Дауылды желге қарамастан теннисші допты ракеткамен тордан асып түсетіндей етіп ұрды.

Келесі күштерді қарастырыңыз:

1.Төмен қарай бағытталған ауырлық күші.

2. Соққының әсер ету күші.

3. Ауа тарапынан әсер ететін күш.

Соққыдан кейін, яғни ракеткадан ажыраған соң жерге түскенше ұшу барысында допқа қандай күштер әсер етеді?

A) Тек 1.

B) 1 және 2.

C) 1 және 3.

D) 2 және 3.

E) 1,2 және 3.