

COGNITIVE AND STRUCTURAL DIMENSIONS OF ECONOMIC REASONING IN STANDARDIZED MATHEMATICS EXAMINATIONS: A COMPARATIVE ANALYSIS OF SAT AND NATIONAL CERTIFICATION SYSTEMS

Anvar Nuriddinov

University of World Economy and Diplomacy
Tashkent, Uzbekistan

ORCID: <https://orcid.org/0009-0009-2970-8153>

Email: anvar.nuriddinov2008@gmail.com

<https://doi.org/10.5281/zenodo.20187083>

Abstract: This study examines how economic reasoning is structurally embedded in two high-stakes mathematics examination systems: the SAT and a national certification examination. Although economic contexts frequently appear in standardized mathematics assessments, limited research has comparatively analyzed how such tasks differ in structural modeling depth and optimization orientation. Drawing on research in mathematical reasoning and cognitive demand, this study introduces the Economic Reasoning Assessment Framework (ERAF) to classify economic-context items according to four levels of structural complexity.

A qualitative document analysis was conducted on 20 examination items (12 SAT, 8 national certification). The results indicate a clear structural contrast. In the SAT dataset, 75% of items were concentrated in Levels 1 and 2, emphasizing applied proportional reasoning and constraint-based modeling, while only 8% reached Level 4 optimization structures. In contrast, 50% of national certification items were classified as Level 4, reflecting explicit profit maximization and formal analytical reasoning.

These findings suggest that although both systems incorporate economic contexts, they differ in the depth and structural orientation of economic reasoning they operationalize. The ERAF framework provides a structured lens for comparative analysis of task design in standardized mathematics assessments.

Keywords: Economic reasoning; Cognitive demand; Mathematical modeling; Standardized assessment; Optimization; Task structure

АННОТАЦИЯ: Данное исследование рассматривает, каким образом экономическое мышление структурно интегрировано в две системы высокоуровневого математического оценивания: SAT и национальный сертификационный экзамен. Несмотря на то, что экономические контексты часто используются в стандартизированных математических заданиях, сравнительных исследований, анализирующих различия в глубине структурного моделирования и направленности на оптимизацию, остается недостаточно. Опираясь на исследования математического мышления и когнитивной нагрузки, данная работа предлагает Framework оценки экономического мышления (ERAF), предназначенный для классификации заданий с экономическим содержанием по четырем уровням структурной сложности.

В исследовании был проведен качественный анализ документов, включающий 20 экзаменационных заданий (12 заданий SAT и 8 заданий национального сертификационного экзамена). Полученные результаты выявили четкий структурный контраст между двумя системами. В выборке SAT 75% заданий были сосредоточены на уровнях 1 и 2, акцентируя внимание на прикладном пропорциональном мышлении и моделировании на основе ограничений, тогда как только 8% заданий относились к структурам оптимизации уровня 4. В свою очередь, 50% заданий национального сертификационного экзамена были классифицированы как уровень 4, что отражает



выраженную ориентацию на максимизацию прибыли и формальное аналитическое мышление.

Полученные результаты свидетельствуют о том, что обе системы используют экономические контексты, однако различаются по глубине и структурной направленности экономического мышления, которое они формируют. Framework ERAF предоставляет системный инструмент для сравнительного анализа структуры заданий в стандартизированных математических оцениваниях.

Ключевые слова: экономическое мышление; когнитивная нагрузка; математическое моделирование; стандартизированное оценивание; оптимизация; структура заданий.

1. Introduction

Standardized mathematics examinations play a decisive role in shaping academic trajectories, institutional access, and educational policy decisions. In high-stakes contexts, these assessments function not only as measures of procedural fluency but also as instruments intended to evaluate higher-order reasoning and problem-solving competence. A notable proportion of mathematics test items embed economic contexts, including profit optimization, revenue modeling, percentage growth, cost functions, and resource allocation scenarios. Such tasks require students to interpret quantitative relationships within structured real-world settings rather than merely perform routine calculations.

Research in mathematics education has emphasized that reasoning is a central yet conceptually complex construct (Kollosche, 2021). Similarly, studies on cognitive demand demonstrate that the structure of assessment tasks significantly influences the depth of reasoning elicited from students (Hsu & Yao, 2023; Ni et al., 2018). Within modeling research, scholars have highlighted the importance of examining how contextualized mathematical problems are designed and assessed, particularly in high-stakes environments (Turner et al., 2022; Geiger et al., 2018). These strands of literature collectively suggest that the structure of assessment items plays a critical role in determining the type and level of reasoning measured.

Despite this growing body of research, limited attention has been devoted to the structural treatment of economic reasoning across different standardized examination systems. While international and national assessments frequently incorporate economic scenarios, comparative analyses of how such tasks operationalize reasoning depth remain scarce. In particular, there is insufficient investigation into whether high-stakes examinations differ in their emphasis on structural modeling, constraint systems, dynamic reasoning, and optimization-based analysis.

This study addresses this gap by conducting a comparative structural analysis of economic reasoning tasks drawn from the SAT and a national certification mathematics examination. Rather than focusing on student performance outcomes, the analysis centers on task design characteristics, cognitive complexity, and modeling orientation. By examining how economic reasoning is embedded and structured within assessment items, the study aims to contribute to broader discussions on cognitive demand, modeling competence, and high-stakes mathematics assessment design. This study addresses the following research question: How do the structural and cognitive characteristics of economic-context tasks differ between the SAT and a national certification mathematics examination?

2. Theoretical Framework

2.1 Mathematical Reasoning as a Structured Cognitive Activity

Mathematical reasoning is widely recognized as a central objective of mathematics education. However, what counts as “reasoning” is not always clearly defined. Kollosche (2021) argues that reasoning in mathematics education is often used as a general label without sufficient theoretical clarity. According to this perspective, reasoning should not be reduced to procedural



correctness; instead, it reflects structured ways of justifying, interpreting, and validating mathematical relationships.

Drawing on the idea of “styles of reasoning,” Kollosche (2021) suggests that different forms of reasoning operate within mathematics, including deductive reasoning, modelling-oriented reasoning, statistical reasoning, and experimental reasoning. These styles shape how problems are approached and what is considered a valid solution. This distinction is important for assessment analysis because the structure of a test item influences which style of reasoning is activated.

In the context of economic problems embedded in mathematics examinations, reasoning may involve interpreting functional relationships, identifying constraints, analysing variable interactions, or determining optimal outcomes. Such tasks go beyond computation and require students to organize quantitative information within structured models. Therefore, understanding reasoning as a structured cognitive activity provides a necessary foundation for analysing assessment items.

2.2 Cognitive Demand and Task Structure

The level of reasoning elicited by a task is closely related to its cognitive demand. Research on mathematical tasks has shown that problems differ significantly in the type and depth of thinking they require (Hsu & Yao, 2023). Some tasks primarily assess procedural fluency, while others demand relational understanding, justification, and strategic decision-making.

Empirical research further indicates that tasks of higher cognitive demand are associated with deeper conceptual engagement (Ni et al., 2018). When students are required to analyze relationships, justify conclusions, or work within constrained systems, the cognitive complexity of the task increases. This suggests that examining the structural properties of test items can provide insight into the level of reasoning being assessed.

For high-stakes examinations, this distinction becomes particularly important. If assessment systems differ in how they design and structure tasks, they may implicitly prioritize different forms of reasoning. Therefore, analysing cognitive demand is not only a matter of difficulty but also of structural orientation.

2.3 Mathematical Modeling in Assessment Contexts

Mathematical modeling research emphasizes the importance of connecting mathematical structures to real-world situations (Geiger et al., 2018). In assessment contexts, modeling tasks often require students to translate contextual information into formal representations such as equations, systems of constraints, or functional relationships.

Turner et al. (2022) highlight that the assessment of modeling competence depends on how tasks are designed and how much autonomy students are given in constructing representations. Some examination systems may provide highly structured prompts, while others require students to identify relevant variables and relationships independently.

Economic-context problems are particularly relevant in this regard. Revenue functions, profit optimization, growth models, and allocation problems are inherently modeling-based. However, the depth of modeling required may vary across examination systems. In some cases, economic tasks remain largely procedural, while in others they demand structural interpretation and strategic reasoning.

3. Methodology

3.1 Research Design



This study employs a qualitative comparative document analysis design. The primary objective is to examine how economic reasoning is structurally embedded within mathematics examination items across two high-stakes assessment systems: the SAT and a national certification mathematics examination. Rather than evaluating student performance data, the analysis focuses on task structure, modeling orientation, and cognitive characteristics of the items themselves.

A document analysis approach is appropriate because the unit of analysis is the assessment item. The study investigates how economic contexts are constructed, what types of quantitative relationships are required, and how constraints and optimization structures are presented within the tasks.

3.2 Data Description

The dataset for this study consisted of 20 economic-context mathematics examination items drawn from publicly available sources. Twelve items were selected from SAT sample materials, and eight items were selected from a national certification mathematics examination.

Items were included if they explicitly incorporated economic variables such as profit, revenue, cost, production quantity, allocation, or growth over time. Tasks involving purely procedural arithmetic without contextual modeling were excluded. The final dataset was constructed to ensure representation of modeling-based, constraint-based, and optimization-oriented problem structures.

All items were classified according to the Economic Reasoning Assessment Framework (ERAF). To enhance analytical consistency, each item was reviewed multiple times to ensure alignment with structural criteria.

3.3 Data Selection and Sampling Criteria

The dataset consists of mathematics examination items drawn from publicly available SAT sample materials and officially released national certification examination tasks. From the identified pool of publicly available items, tasks were selected according to the following criteria:

1. The task explicitly includes an economic context (e.g., profit, revenue, cost, growth, allocation, percentage change).
2. The task requires interpretation of quantitative relationships rather than simple arithmetic substitution.
3. The item structure allows classification according to modeling or optimization characteristics.

Items that involved purely procedural computation without contextual interpretation were excluded. The final dataset includes a structured sample of economic-context tasks from both examination systems to enable comparative analysis.

3.4 Analytical Framework: ERAF Classification

To systematically examine structural differences, the study applies the Economic Reasoning Assessment Framework (ERAF), developed for this research. The framework categorizes economic-context mathematics tasks into four levels based on structural and cognitive characteristics:

- Level 1: Fundamental Quantitative Economic Skills (basic proportional reasoning, percentage calculations, simple linear relationships).



- Level 2: Structural Constraint Modeling (systems of equations, variable interdependence, explicit constraints).
- Level 3: Dynamic and Functional Reasoning (exponential growth, time-dependent change, functional interpretation).
- Level 4: Strategic Optimization and Analytical Decision-Making (profit maximization, quadratic modeling, comparative decision structures).

Each selected item was independently analyzed and classified according to these structural criteria. The classification emphasizes task design features rather than student difficulty levels.

3.5 Comparative Analysis Procedure

After classification, the distribution of items across ERAF levels was examined for each examination system. The analysis focuses on identifying structural patterns, such as:

- The prevalence of modeling-based tasks,
- The presence of optimization structures,
- The degree of constraint interdependence,
- The balance between procedural and structural reasoning.

This approach enables a structural comparison of how economic reasoning is operationalized across assessment systems.

4. Results

4.1 Distribution of ERAF Levels in SAT Items

The analysis of SAT economic-context items ($n = 12$) indicates a concentration in Level 1 and Level 2 categories. Four items (33%) were classified as Level 1, involving proportional reasoning, percentage change, and basic linear relationships in applied economic settings.

Five items (42%) were categorized as Level 2, requiring structural modeling of interdependent variables, including systems of equations and explicit constraints. These tasks demanded interpretation of relationships rather than simple substitution.

Two items (17%) were classified as Level 3, incorporating dynamic or exponential reasoning within economic growth contexts. Only one item (8%) met Level 4 criteria, involving structured optimization reasoning within a quadratic framework.

4.2 Distribution of ERAF Levels in National Certification Items

In contrast, the national certification dataset ($n = 8$) demonstrated a stronger emphasis on higher-level reasoning structures.

One item (13%) was classified as Level 1. Two items (25%) fell under Level 2, involving constraint-based modeling. One item (13%) was categorized as Level 3.

Notably, four items (50%) were classified as Level 4, requiring explicit optimization reasoning such as profit maximization or quadratic function analysis. These tasks emphasized formal analytical structures and strategic decision-making.

4.3 Comparative Structural Patterns

The comparative distribution reveals distinct structural orientations between the two examination systems.

The SAT predominantly emphasizes applied structural modeling (Levels 1–2 = 75%), with limited representation of optimization-based reasoning (Level 4 = 8%). In contrast, the national



certification examination demonstrates a substantially higher proportion of optimization-oriented tasks (Level 4 = 50%).

These findings suggest that while both systems incorporate economic contexts, the structural depth and analytical orientation of reasoning differ in meaningful ways.

Table 1. Distribution of ERAF Levels Across Examination Systems

Examination	Level 1	Level 2	Level 3	Level 4	Total
SAT	4	5	2	1	12
National	1	2	1	4	8

5. Discussion

The findings indicate a clear structural contrast. Whereas 75% of SAT items were concentrated in Levels 1 and 2, the national certification examination allocated 50% of its items to Level 4 optimization structures. This contrast highlights different assessment priorities in the operationalization of economic reasoning.

The SAT dataset demonstrates a stronger concentration in structural modeling tasks, particularly those requiring interpretation of linear relationships and constraint-based systems. This pattern aligns with research emphasizing modeling-oriented reasoning and contextual interpretation in mathematics assessment (Geiger et al., 2018; Turner et al., 2022). The predominance of Level 1 and Level 2 items suggests that economic reasoning in the SAT is frequently operationalized through applied structural relationships rather than formal optimization procedures.

In contrast, the national certification examination exhibits greater representation of optimization-based reasoning. Tasks involving quadratic profit functions and maximum or minimum value identification reflect a stronger emphasis on analytical formalization. From a cognitive demand perspective, this distribution corresponds to higher levels of structured reasoning (Hsu & Yao, 2023; Ni et al., 2018), particularly when students must analyze constrained systems and justify optimal decisions.

These structural differences may reflect broader educational orientations. International standardized tests such as the SAT often prioritize transferable modeling competence within applied contexts, while national examinations may emphasize algebraic rigor and formal optimization skills. However, the present analysis does not evaluate the effectiveness of either approach; rather, it demonstrates that economic reasoning can be operationalized through distinct structural pathways.

Importantly, the ERAF classification framework provides a systematic lens for identifying these differences. By focusing on task design rather than performance outcomes, the study contributes to ongoing discussions about cognitive demand, modeling depth, and structural complexity in high-stakes mathematics assessment.

These findings suggest that assessment systems may implicitly prioritize different forms of economic reasoning depending on their structural design principles. This observation reinforces the importance of examining task architecture rather than relying solely on difficulty indices.

6. Conclusion

This study examined how economic reasoning is structurally embedded within two high-stakes mathematics examination systems. By applying the Economic Reasoning Assessment Framework (ERAF), the analysis identified clear differences in the cognitive and structural orientation of economic-context tasks across the SAT and a national certification examination.

The findings indicate that while both systems incorporate economic scenarios, they operationalize reasoning in distinct ways. The SAT predominantly emphasizes applied structural modeling within bounded algebraic contexts, whereas the national certification examination demonstrates stronger representation of optimization-oriented and analytically formalized tasks.



These differences suggest that assessment design plays a central role in shaping the type of reasoning that examinations prioritize.

Rather than evaluating which system is superior, this study highlights how structural characteristics influence the nature of economic reasoning measured in high-stakes contexts. The ERAF framework offers a structured analytical lens that may support future comparative research on assessment design and cognitive demand in mathematics examinations.

Several limitations should be acknowledged. The dataset is restricted to selected publicly available examination items, and the classification relies on structural analysis rather than empirical performance data. Future research may extend this framework by incorporating larger datasets, cross-national comparisons, or quantitative validation procedures.

Despite these limitations, the study contributes to ongoing discussions in mathematics education by demonstrating that economic reasoning in standardized assessments is not uniform but structurally shaped by examination design principles.

7. Data Availability Statement

The data analyzed in this study consist of publicly available examination items from SAT sample materials and officially released national certification examinations. No new datasets were generated. The analyzed materials are accessible through their respective official sources.

8. AI Usage Disclosure

The author used ChatGPT (OpenAI, GPT-5 architecture) for language refinement and structural editing assistance during manuscript preparation. All analytical decisions, framework development, data classification, and final interpretations were independently conducted and verified by the author.

References

Geiger, V., Stillman, G., Brown, J., & Galbraith, P. (2018). Using mathematics to solve real world problems: The role of enablers. *Mathematics Education Research Journal*, 30(1), 7–19. <https://doi.org/10.1007/s13394-017-0217-3>

Graff Zivin, J., Song, Y., Tang, Q., & Zhang, P. (2020). Temperature and high-stakes cognitive performance: Evidence from the national college entrance examination in China. *Journal of Environmental Economics and Management*, 104, 102365. <https://doi.org/10.1016/j.jeem.2020.102365>

Hsu, H.-Y., & Yao, C.-Y. (2023). A review of the mathematical tasks framework and levels of cognitive demand. In J. Cai, G. J. Stylianides, & P. A. Kenney (Eds.), *Research studies on learning and teaching of mathematics (Research in Mathematics Education)*. Springer. https://doi.org/10.1007/978-3-031-35459-5_10

Kollosche, D. (2021). Styles of reasoning for mathematics education. *Educational Studies in Mathematics*, 107, 471–486. <https://doi.org/10.1007/s10649-021-10046-z>

Maxwell, M., & Gleason, J. (2018). Item efficiency: An item response theory parameter with applications for improving the reliability of mathematics assessment. *International Journal of Mathematical Education in Science and Technology*, 49(2), 216–243. <https://doi.org/10.1080/0020739X.2018.1492038>

Turner, E. E., Roth McDuffie, A., Bennett, A. B., Aguirre, J. M., & Drake, C. (2022). Mathematical modeling in the elementary grades: Developing and testing an assessment. *International Journal of Science and Mathematics Education*, 20, 1387–1409. <https://doi.org/10.1007/s10763-021-10195-w>

Warsitasari, W. D., & Rofiki, I. (2023). Utilizing GeoGebra for solving economic mathematics problems: Promoting logical reasoning in problem-based learning. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 12(3), 3445–3456. <https://doi.org/10.24127/ajpm.v12i3.7300>

