DOCTORAL DISSERTATION THESIS BOOKLET

Dömsödi Balázs

Advancing Automated Exam Generation: Toward Scalable and Adaptive Solutions

Corvinus University of Budapest
Doctoral School of Economics and Business Informatics
(Business Informatics Doctoral Program)

Supervisor: Láng Blanka, Ph.D.

Budapest, 2026

TABLE OF CONTENTS

I. RESEARCH BACKGROUND AND JUSTIFICATION FOR THE SELECTION OF THE TOPIC

I.1. Topic Overview

The manual construction of examination papers has long been recognized as a labor-intensive and cognitively demanding process. Designing assessments that are simultaneously balanced in difficulty levels, comprehensive, and aligned with curricular objectives requires a high level of precision and sustained pedagogical judgment, therefore educators engaged in this process must navigate several interrelated dimensions.

The challenge further rises when multiple parallel versions of the same examination are required, as is commonly the case in large-enrollment courses or where institutional policy demands variation in test forms as a safeguard against academic dishonesty. Generating several versions of an assessment that are equivalent in content coverage, difficulty, and structural balance, while simultaneously ensuring sufficient variation to prevent direct replication, represents a level of combinatorial and pedagogical complexity that places substantial strain on even experienced educators.

This burden can cause meaningful costs on educators' time and intellectual resources, time that might otherwise be directed toward instructional design, student feedback, or other activities with a higher ratio of added personal value.

In response to these challenges, there has been growing scientific and practical interest in the development of automated systems capable of generating high-quality educational assessments, attracting increasing research attention over recent decades.

Automated assessment generation systems are designed to perform the computationally intensive and rule-governed aspects of exam construction algorithmically, drawing upon structured repositories of questions and applying parameterized selection procedures to compose assessments that satisfy defined pedagogical and structural criteria. When implemented effectively, such systems offer the potential to introduce greater objectivity, consistency, and scalability into assessment practices, qualities that are difficult to guarantee through manual processes alone.

A fundamental tension present in the design of automated assessment systems concerns the relationship between pedagogical richness and operational scalability. On the one hand, a system with high pedagogical fidelity would be capable of modeling fine-grained parameters and applying these distinctions to produce assessments of demonstrably high educational validity. On the other hand, the deeper and more elaborate the parameterization of such a system, the greater the computational and data-management demands it places on its infrastructure.

Despite the theoretical promise of automated assessment generation, the field has not yet produced systems that are universally regarded as satisfactory replacements for human-designed assessments.

This dissertation aims to examine the current state of automated educational software, with a particular focus on automated assessment generation. It seeks to identify existing research gaps and contribute to the advancement of the field through the further development of an existing system, with the goal of bringing it closer to practical deployment in real-world

educational environments. The automated examination generation system investigated in this study is the Exercise Generation Algorithm+ (EGAL+) (Láng & Kardkovács 2016; Láng 2019; Láng 2020), which serves as the primary subject of both analysis and implementation.

Within the specific context of EGAL+, these broader limitations manifested as concrete research problems requiring systematic investigation. At the time of the present study's commencement, the system existed as a promising but experimental prototype, exhibiting both notable capabilities and identifiable constraints. Specifically, the balance between EGAL+'s deep pedagogical parameterization and its scalability had not been formally examined or systematically addressed; the system's performance in real-world educational settings had not been subjected to rigorous empirical benchmarking against manually produced assessments; and the trajectory for its continued development had not been formally articulated in relation to the broader field. These gaps collectively defined the research problem motivating this dissertation and formulated its governing research questions.

I.2. Research Questions

1. **What structural and functional limitations exist in current automated exam generation systems?**

This question aims to critically examine the current landscape of automated exam generation technologies by identifying their inherent structural and functional shortcomings. It seeks to uncover systemic constraints that restrict the effectiveness of these systems in diverse educational contexts. By synthesizing insights from existing approaches, this analysis defines the problem space.

2. **How can the trade-off between deep pedagogical parameterization and scalability in EGAL+ be systematically addressed?**

This research question focuses on a central design tension within EGAL+: the balance between rich, fine-grained pedagogical control and the practical need for scalability and efficiency. It investigates how increased complexity in modeling parameters impacts system performance and usability. The question further explores methodological and architectural

strategies developed during the research to reconcile this trade-off, aiming to demonstrate how EGAL+ can achieve both pedagogical sophistication and operational viability.

3. **How does EGAL+ perform in real-world educational contexts compared to manual exam compilation in terms of quantitative assessment quality metrics and operational efficiency?**

This question evaluates the practical effectiveness of EGAL+ by benchmarking its performance against traditional, manually constructed exams. It emphasizes empirical comparison using measurable indicators, such as assessment quality, time efficiency, and consistency. The goal is to determine whether EGAL+ can not only match but potentially exceed human performance in exam generation tasks within authentic educational settings. This analysis provides critical evidence regarding the system's readiness for adoption and its potential impact on educational practice.

4. **Which future research and development directions logically follow for EGAL+, and what do they imply for the field at large?**

This final question is forward-looking, aiming to establish a trajectory for future research and development based on the current findings. It considers both the current limitations of EGAL+ and the broader opportunities for extending its framework. The goal is to articulate a vision for how this work may serve as a foundation for continued innovation in automated assessment and contribute to the evolving landscape of automated educational systems.

II. THE METHODS USED

II.1. Research Design

The research presented in this dissertation was conducted according to a multi-phase, multi-methodological design of coherent and logically sequenced investigations.

Rather than adhering to a single research paradigm, the study drew upon a combination of theoretical analysis, algorithmic investigation, software engineering practice, and empirical evaluation, each phase building upon the findings of the preceding one. This sequential design was deliberate: the complex, multi-dimensional nature of the research questions

spanning system design, computational performance, pedagogical validity, and real-world applicability necessitated a framework capable of generating evidence at multiple levels of abstraction.

The overarching research strategy can be described as a mixed-methods design with an applied systems orientation. Quantitative methods were employed to produce measurable, comparative evidence regarding computational performance and educational outcomes, while qualitative methods provided contextual insight into the experiences and perceptions of end users.

The research unfolded across four broad phases, which corresponded closely to the four research questions formulated in the preceding chapter. The first phase comprised a systematic review of the relevant literature, establishing the theoretical and empirical landscape of automated assessment generation and situating EGAL+ within it. The second phase involved a deep internal analysis and comprehensive reconstruction of the EGAL+ system, addressing both its algorithmic foundations and its software architecture. The third phase consisted of a

controlled empirical deployment of the redesigned system in an authentic university examination setting, generating quantitative and qualitative data against which its practical performance could be evaluated. The fourth and final phase synthesized the findings of all preceding phases into an integrated set of conclusions and forward-looking recommendations. This sequential, cumulative design ensured that each methodological component generated findings directly informative to the research questions and that the overall study was both coherent and robust.

II.2. Applied Methods

The literature review employed a systematic methodology for the identification, selection, and analysis of relevant research within the domain of automated assessment design. The review focuses on contributions published during the past 15 years and is based on a corpus of more than 40 publications identified through structured searches conducted using Google Scholar. The search strategy combined keywords related to automated assessment, question generation, metaheuristic optimization, and machine learning, thereby

ensuring broad yet focused coverage of the field. Inclusion criteria emphasized methodological relevance, empirical rigor, and citation impact in order to capture both influential and technically substantive studies. Publications lacking a clear methodological contribution or falling outside the defined scope of the review were excluded.

The core methodology underlying EGAL+ is the Harmony Search algorithm, a metaheuristic optimization technique that serves as the foundation for the system's development and enhancement. Metaheuristics are high-level problem-independent algorithmic frameworks especially fitting for complex combinatorial problems (Sörensen & Glover, 2018). The Harmony Search algorithm was first conceptualized and formally introduced by Geem et al. in 2001 and has since been applied to a wide range of optimization problems. Its unique characteristic is its analogy of musicians remembering successful note choices in increasing the probability of generating similarly high-quality solutions in subsequent iterations (Lee & Geem, 2005). In the present study, the algorithm was selected as the principal framework upon which additional capabilities and improvements were developed.

Furthermore, EGAL+ was completely reimplemented in C++, a programming language particularly fitting computation intensive tasks (Stroustrup, 2013), replacing its previous PHP-based architecture to improve performance.

The benchmarking evaluation of EGAL+ was conducted through a methodology that involved executing both the original and the enhanced versions of the software multiple times under identical input conditions. The resulting outputs were subsequently compared using a set of predefined key performance indicators, enabling a systematic assessment of the improvements introduced in the revised implementation.

To evaluate the real-world applicability of the proposed approach, a case study was conducted at Corvinus University of Budapest during the autumn semester of 2024. The study involved four distinct student groups, with two groups receiving examinations generated by EGAL+ and the remaining two groups receiving manually constructed examinations.

Both qualitative and quantitative measures were employed in the evaluation. Qualitative assessment focused on comparing the preparation time required by instructors for

manual and automated examination generation and students' perceived difficulty of the assessment. Quantitative analysis examined specific characteristics of the generated examinations, including diversity of included tasks between exams and following educator's preferences in exam composition, compared these between manually created and automatically generated examinations.

III. SCIENTIFIC RESULTS OF THE DISSERTATION

- The literature review identified a critical research gap concerning the trade-off between pedagogical richness and operational scalability in automated assessment systems and positioned EGAL+ within the field.
- Completely restructured EGAL+ and reimplemented from PHP to C++, overcoming major scalability limitations while preserving optimization quality and assessment consistency requirements.
- Designed and implemented an advanced difficulty-balancing mechanism for automated exam generation with EGAL+.

- Achieved execution time reductions of over 99% in large-scale benchmarking scenarios, significantly expanding the system's applicability to large cohorts and complex examination settings.
- Developed and prototyped a hybrid framework integrating generative AI with metaheuristic optimization, demonstrating a viable pathway toward partially automated assessment ecosystems with reduced dependence on precompiled question banks.
- Empirically validated the system in a real university examination context, where EGAL+-generated exams consistently outperformed manually composed exams in structural diversity, preference satisfaction, and overall optimization fitness.
- Demonstrated practical benefits for educators through substantial reductions in exam preparation time while maintaining assessment quality.
- Confirmed the pedagogical reliability of EGAL+ through student evaluations, which indicated comparable perceptions of difficulty and fairness

between automated and traditionally created examinations.

- Established EGAL+ as a significant contribution to the field of automated assessment generation, addressing a clear research gap while providing a foundation for future research on automation in educational workflows.
- Derived several directions for future research in automated assessment generation, including scalable optimization architectures for large-scale deployment, hybrid heuristics-AI workflows for assessment creation, automated and more user-friendly configuration of assessment parameters, enhanced interoperability with learning management systems and broader educational technology ecosystems.

IV. MAIN REFERENCES

Geem, Z. W., Kim, J. H., & Loganathan G. V. (2001) A New Heuristic Optimization Algorithm: Harmony Search. SIMULATION. 76(2), pp. 60-68. https://doi.org/10.1177/003754970107600201

Láng, B. & Kardkovács, T. Zs. (2016). Solving exercise generation problems by diversity oriented meta-heuristics, In: Shuang, Cang; Yan, Wang (eds.) SKIMA: 2016 10th International Conference on Software, Knowledge, Information Management & Applications: University of Information Technology, December 15- 17 2016, Chengdu, China, pp. 49-54, https://doi.org/10.1109/SKIMA.2016.7916196

Láng, B. (2019). Solving Exercise Generation Problems Using the Improved EGAL Metaheuristic Algorithm. SEFBIS Journal, 13, pp. 23-31.

Láng, B. (2020). Solving Exercise Generation Problems Using the Improved EGAL Metaheuristic Algorithm with Precedence Constraints. In: Auer, M., Hortsch, H., Sethakul, P. (eds) The Impact of the 4th Industrial Revolution on Engineering

Education. ICL 2019. Advances in Intelligent Systems and Computing, 1135. https://doi.org/10.1007/978-3-030-40271-6_56

Lee, K. S., Geem, Z. W. (2005). A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory and practice. Computer Methods in Applied Mechanics and Engineering, 194(36–38), pp. 3902-3933. https://doi.org/10.1016/j.cma.2004.09.007

Sörensen, K., Sevaux, M., & Glover, F. (2018). A History of Metaheuristics. In: Handbook of Heuristics, pp. 791-808. https://doi.org/10.1007/978-3-319-07124-4_4

Stroustrup, B. (2013). The C++ programming language

V. LIST OF OWN (OR CO-AUTHORED) PUBLICATIONS ON THE TOPIC

Láng, B., & Dömsödi, B. (2022). Development of the Improved Exercise Generation Metaheuristic Algorithm EGAL+ for End Users. *International Journal of Emerging Technologies in Learning (iJET)*, *17*(11), pp. 210–224. https://doi.org/10.3991/ijet.v17i11.28099

Dömsödi, B., & Láng, B. (2022). Improvements of the EGAL+ metaheuristic algorithm for solving exercise generation problems with a significantly larger number of elements. In: OGIK'2022: Országos Gazdaságinformatikai Konferencia, pp. 55-56.

Dömsödi, B., & Láng, B. (2023). Exploring Hybrid Solutions in Educational Software: The Integration of AI and Metaheuristics. In: OGIK'2023: Országos Gazdaságinformatikai Konferencia, pp. 15.

Láng, B., & Dömsödi, B. (2024). Integration of AI and Metaheuristics in Educational Software: A Hybrid Approach to Exercise Generation. *International Journal of Emerging Technologies in Learning (iJET)*, *19*(06), pp. 38–51. https://doi.org/10.3991/ijet.v19i06.49829

Dömsödi, B., Láng, B., & Kovács, L. (2024). Integrating AI and Metaheuristics in Educational Software: A Case Study in Live University Exam Settings. In: OGIK'2024: XX. Országos Gazdaságinformatikai Konferencia: Az előadások összefoglalói, pp. 53.

Dömsödi, B., Láng, B., & Kovács, L. (2025, November 13). Innovative educational systems through the fusion of metaheuristics and AI [Conference presentation]. Day of Hungarian Science, Budapest University of Economics and Business, section: Interdisciplinary approaches to addressing the opportunities and challenges posed by digitalization and artificial intelligence. https://webapi.uni-bge.hu/api/v1/files/display/documents/bueb-section-brochure-2025.pdf#page=6

Láng, B., Kovács, L., & Dömsödi, B. (2026). AI-Enhanced Exam Generator Program: A Case Study in Live University Exam Settings. *Central-European Journal of New Technologies in Research, Education and Practice*, *8*(1), 1–24. https://doi.org/10.36427/CEJNTREP.8.1.12403