Extracting Curriculum Rules: Formal Modelling of the Study Entry and Orientation Phase

Selina Reinhard
TU Wien, Austria
Selina.Reinhard@tuwien.ac.at

Tom Breuer
RWTH Aachen University, Germany
Tom.Breuer@rwth-aachen.de

René Röpke
TU Wien, Austria
Rene.Roepke@tuwien.ac.at

ABSTRACT

The transition into higher education is a critical phase. In Austria, it is structured through the Study Entry and Orientation Phase (StEOP), a formally mandated introductory phase. Despite its central role, corresponding regulations are specified in natural language, making them difficult to interpret, compare, and analyse. This paper addressed this challenge by proposing a formal model of StEOP regulations based on nested course pools and credit requirements. To operationalise the model, we introduce a semi-automated extraction approach. Results show that all but one of the analysed StEOPs can be consistently mapped to the proposed model. By introducing a formal model of the StEOP and empirically validating its applicability, this work enables the systematic analysis and comparison of curriculum designs, laying the foundation for data-driven evaluation of study entry phases.

Keywords

Curriculum Modelling, Formalisation, Constraint Satisfaction Problems, Rule Extraction, Data Mining

1. INTRODUCTION

The transition to higher education (HE) represents a critical phase for many students, often marked by uncertainty about study requirements and expectations. To facilitate this transition, universities have implemented various support mechanisms. In Austria a structured introductory phase – the **St**udy **E**ntry and **O**rientation **P**hase (StEOP) – has been introduced. The StEOP defines a set of courses with additional constraints that students must complete before they can enrol in more advanced courses. The aim of the StEOP is to provide an early overview of the study program's content and requirements, enabling students to make informed decisions about whether the chosen field of study aligns with their interests and abilities, and to serve as a regulatory mechanism.

Selina Reinhard, Tom Breuer, and René Röpke. Extracting Curriculum Rules: Formal Modelling of the Study Entry and Orientation Phase. In Anthony Botelho, Maria Mercedes T. Rodrigo, Adish Singla, Hiroaki Ogata, Hyojeong So, and Young Hoan Cho (eds.) Proceedings of the 19th International Conference on Educational Data Mining, Seoul, Republic of Korea, June, 2026, pp. 669–673. International Educational Data Mining Society (2026).

https://doi.org/10.5281/zenodo.21039746

While the StEOP is implemented nationwide and follows general regulatory guidelines, its concrete design and structure are defined individually for each study program. These specifications are described in textual form within the official curricula. However, there is no standardised formal model or machine-readable representation of these regulations. Without such a formal representation, systematically analysing and understanding how curriculum design – particularly different StEOP implementations – affects student progression and success remains challenging. This work addresses this gap by proposing a formal model of StEOP regulations, introducing a semi-automated extraction approach, and empirically evaluating the model's validity. Thus, we aim to answer the following research questions:

RQ1 How can the StEOP structure be formally modelled?

RQ2 To what extent can different StEOP implementations be accurately represented using the proposed model?

2. RELATED WORK

Our work relates to and extends existing work in the fields of curriculum analytics (CA) and educational data mining (EDM), especially prior research on curriculum modelling, study planning, and formal representations of study regulations: Study regulations are typically specified in natural language, which makes them accessible to humans but difficult to process computationally. This gap introduces challenges such as ambiguity and limited support for automated analysis and verification [3]. Therefore, modelling curriculum regulations (e.g., StEOPs) in a formal, machine-readable way remains an important topic in educational technology (EdTech) and EDM, where structured representations are a prerequisite for large-scale analysis. Prior work has approached curriculum modelling from different perspectives, including structural analysis [1], formal verification [6], and individual study planning [7]. One line of research formulates curriculum planning as a CSP, proposing a planning system based on a constraint-satisfaction approach for constructing valid study plans across multiple semesters [11]. Similarly, process mining approaches can be used to model prerequisite structures and learning pathways to support study planning [5, 10]. Other approaches focus on richer semantic representations of curricula, for instance, based on disciplines, competences, and specialisations [12], or model curricula as prerequisite graphs to support curriculum overview [8] and to verify structural properties such as soundness and the absence of cycles [6].

3. BACKGROUND

The StEOP was initially defined in §66 of the Austrian Universities Act (AUA) of 2002 and underwent significant revisions in the 2015 amendment (BGBI. I Nr. 131/2015). The regulation mandates that all Bachelor's programs, with a few exceptions, must incorporate a StEOP. Its primary goal is to "provide students with an overview of the essential contents of the respective degree program and its further progression, and create an objective basis for making a personal assessment of his or her choice of study" (§66 AUA 2002 (1)). The StEOP must comprise between 8 and 20 credits according to the European Credit Transfer System (ECTS), which have to be completed before students can continue with more advanced courses. An exception are the pre-reg courses, non-StEOP courses students can take up to a certain credit limit before completing the StEOP. By restricting access to advanced courses, the StEOP serves as a regulatory mechanism, imposing a barrier that encourages early completion of foundational courses to engage with the study program's core concepts and expectations at an early stage. At the same time, the pre-reg courses relax the definition by providing some flexibility.

As part of a nationwide analysis in Austria in 2020 [4], standardised indicators were compiled and evaluated to compare the StEOP across eleven institutions. Results showed that 26 % of all enrolled students did not complete any StEOP examinations. Further, among those who started the StEOP, only 47 % completed it within the first semester, and 72 % within two semesters. Still, surveys showed that the StEOP mostly provided a good overview of essential study content, clarified performance expectations, and positively reinforced students' choice of study.

4. AN ILLUSTRATIVE EXAMPLE

To illustrate the structure of StEOP regulations, we consider an example from the Mechanical Engineering (ME) program at TU Wien in 2025. The English translation of the corresponding part of the curriculum definition is given in Figure 1. As shown, the StEOP consists of ten courses organised into groups, accompanied by a natural-language specification that defines the conditions for completing the StEOP: Students must complete twelve credits while meeting various conditions, including completing all mandatory courses and choosing courses from different subgroups. In addition, the curriculum defines a set of pre-reg courses, of which up to 22 credits can be taken.

> As part of the Study Entry and Orientation Phase (StEOP) for the Bachelor's program in Mechanical Engineering (ME) at TU Wien, a total of 12 credits must be completed:
>
> **Mandatory courses:**
>
> 1.0 Introduction to Mechanical Engineering (INTRO)
> 2.0 Physics for Mechanical Engineering (PHY)
>
> **Pool 1: Mathematics Fundamentals**
>
> 6.0 Mathematics 1 (MATH)
> 3.0 Mathematics 1 Exercises (MATH EX)
>
> **Pool 2: Subject-Specific Fundamentals**
>
> 3.0 Fundamentals of Manufacturing Engineering (MANU)
> 3.0 Fundamentals of Business Administration (BA)
> 2.0 Technical Drawing / CAD (DRAW)
> 3.0 Fundamentals of Design Engineering (DESIGN)
> 3.0 Technical Drawing / CAD Exercise (DRAW EX)
> 2.0 Mechanics 1 Exercises (MECH EX)
>
> The STEOP is considered successfully completed once both mandatory courses have been passed, and a total of at least 9 credits have been obtained from Pool 1 and Pool 2 combined. At least one course must be selected from each of the two pools. Before completing the StEOP, students may take up to 22 credits of non-StEOP courses from [a list of pre-reg courses].

Figure 1: Description of the StEOP structure in ME.

5. CONCEPTUAL MODEL OF THE STEOP

To derive a formal representation of StEOP regulations, we conducted an exploratory analysis of multiple natural-language descriptions in the official curricular documents. Based on these observations, we first introduce a conceptual model of the StEOP, which we then formalise mathematically and apply to our example.

From analysing several StEOP definitions, we observe that they share the following structural elements:

(1) A set of courses included in the StEOP.
(2) A number of credits to be completed from these courses.

Additionally, they may include some of the following:

(3) A set of mandatory courses.
(4) A subset of courses from which a specified number of credits must be earned.
(5) A subset of courses of which at least one must be completed.

Based on this, we propose modelling the StEOP as a set of *pool rules*. Each pool consists of a *set of courses* and a *minimum number of credits* that must be completed from the set. By combining multiple pool rules (including nested ones) that must all be fulfilled, we can represent all the elements identified above. Elements (1), (2), and (4) are directly captured. A set of mandatory courses (3) can be modelled as a pool in which the required number of credits equals the sum of all courses' credits. Similarly, a subgroup of courses, of which at least one must be completed (5), can be represented as a pool in which at least one credit must be earned. Since courses can only be completed in their entirety, this effectively ensures that at least one full course is completed. In addition to the StEOP course pools, curricula may define a set of pre-reg courses that students can take before completing the StEOP. While this gives students some flexibility, a maximum credit limit applies to pre-reg courses, requiring students to complete the StEOP before taking even more advanced courses.

5.1 Mathematical Formalisation

This conceptual model can be mathematically formalised as follows: Let $\mathcal{C}$ be the universal set of all available courses, where each course $c \in \mathcal{C}$ is represented as a tuple $c = (\text{id}_c, \text{name}_c, \text{cr}_c)$ consisting of a unique course ID, a name, and its corresponding number of credits ($\text{cr}_c \in \mathbb{N}$). Let $\mathcal{X}$ be the set of all available curricula. A specific curriculum $x \in \mathcal{X}$ is defined as a tuple $x = (C_x, S_x, V_x)$ where $C_x \subseteq \mathcal{C}$ is the set of all courses belonging to this specific curriculum.

The StEOP S_x for curriculum x is defined as a set of pool rules:

$$S_x = \{(P_1, k_1), (P_2, k_2), \ldots, (P_n, k_n)\}$$

where each pool $P_i \subseteq C_x$ is a subset of courses, and $k_i \in \mathbb{N}$ specifies the minimum number of credits that must be earned from pool P_i. Let $P_{S_x} = \bigcup_{i=1}^{n} P_i$ denote the set of all StEOP courses across all pools in curriculum x.

The pre-reg rules, $V_x = (W_x, m_x)$, define the additional non-StEOP courses a student may take before completing the StEOP. $W_x \subseteq C_x$ is the set of eligible pre-reg courses and $m_x \in \mathbb{N}$ is the maximum allowed credit limit from those. These courses must not overlap with any StEOP pool:

$$W_x \cap P_{S_x} = \emptyset$$

Let $\mathcal{T}$ be the set of all students. We introduce two boolean predicates: $\text{Passed}(t, c)$ evaluates to true if a student $t \in \mathcal{T}$ has successfully completed course c, and $\text{Taken}(t, c)$ is true if the student t has already passed or is currently registered for course c. We can now define the predicate $\text{Steop}(t, x)$, which indicates whether student t has fulfilled all StEOP requirements for curriculum x:

$$\text{Steop}(t, x) \iff \forall (P_i, k_i) \in S_x : \sum_{\{c \in P_i \mid \text{Passed}(t,c)\}} \text{cr}_c \geq k_i$$

This states that for all pool rules in the StEOP, the sum of the credits of all passed courses within that pool must be at least the defined minimum k_i. From a theoretical perspective, this formalisation can be interpreted as a constraint satisfaction problem (CSP) over sets of courses [2]. Each pool rule corresponds to a constraint on the accumulated credits within the sets, and the StEOP is completed when all constraints are fulfilled.

Further, we can define the predicate $\text{CanEnroll}(t, c, x)$, which defines whether a student t can enroll in course $c \in C_x$ within curriculum x as:

$$\text{CanEnroll}(t, c, x) \iff \text{Steop}(t, x) \lor c \in P_{S_x} \lor \left(c \in W_x \land \sum_{\{c' \in W_x \mid \text{Taken}(t,c')\}} \text{cr}_{c'} + \text{cr}_c \leq m_x \right)$$

This states that a student can enrol in the course if they have already completed the StEOP, the course is part of the StEOP pools, or the course is a pre-reg course, and taking it does not cause the sum of credits of all taken pre-reg courses to exceed the maximum allowed credit limit.

5.2 Applying it to our Example

Applying this model to the ME study program at TU Wien, the StEOP can be modelled using four pools, as shown in Figure 2. The mathematical and subject-specific fundamentals are represented as two pools (each requiring 1 credit), while their combination forms an elective pool requiring 9 credits. The mandatory component is modelled as a pool with a 3-credit requirement, equal to the sum of all included courses. While an overarching pool with a 12-credit requirement could be added, this constraint is already implicitly covered. This formal model precisely maps the textual requirements shown in Figure 1.

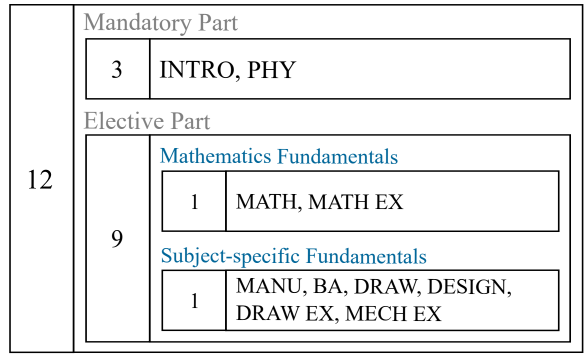

Figure 2: Visualised formalisation of the StEOP in ME.

6. MODEL VALIDATION

To validate the proposed model, we aim to formalise the StEOPs of all curricula at TU Wien respectively. Due to the ambiguity of many pre-reg rules, we focused only on the StEOP rules.

6.1 Semi-Automated Rule Extraction

Since StEOP regulations are provided exclusively in textual form and do not follow a strictly standardised structure, we opted for a semi-automated, interactive extraction approach. First, all relevant curriculum documents (PDFs) are automatically collected from the institution's website. Due to their broadly consistent structure, these documents can be processed uniformly, allowing the StEOP sections to be automatically identified and extracted. The extracted StEOP descriptions are presented to the user in an interactive command-line interface, where the user then encodes the corresponding pool rules through a sequence of simple commands. A pool is created by specifying its name, whether it is mandatory, and the minimum number of credits required. Subsequently, the courses in the pool are entered line by line in a dedicated input mode, including the course name, course type (e.g., lecture or exercise), and number of credits. These course descriptions can often be copied directly from the StEOP description, as they most often follow the same pattern. Lastly, the course names are automatically matched to the corresponding course identifiers. The described semi-automated pool rule extraction process was performed independently by two researchers for all 20 study programs at TU Wien since 2017.

6.2 Results

Across the 20 study programs and their curriculum versions at TU Wien, a total of 46 distinct StEOP regulations were found. Extracting the rules took about two minutes per StEOP. All but one could be fully modelled using the proposed formalisation, achieving a coverage of 98 %. The exception being the StEOP for Computer Engineering from 2023, which contains an elective pool rule that prohibits combining three included courses. Since the proposed formalisation only allows specifying minimum credit requirements, it cannot represent such exclusions. Capturing this rule would require extending the model to include exclusion criteria.

Comparing the extracted pool rules showed that both researchers always identified the same underlying structures. Some differences emerged in naming conventions for pools: one researcher adhered more closely to the terminology used in the official curriculum, while the other preferred a more functional naming scheme. Further, cases in which at least one course from a subgroup has to be completed were sometimes interpreted as requiring the exact minimum number of credits for a course, whereas in other cases they were modelled as a single-course constraint requiring one credit. Similarly, mandatory courses were sometimes additionally included in the elective pool, leading to slight variations in how credit requirements were parameterised. Despite these differences in representation, the resulting models were semantically equivalent and thus equally valid.

During the modelling process, several practical challenges emerged, including missing information (e.g., credit values), underspecified courses, and inconsistencies between course names in the StEOP descriptions and the module handbook. Additionally, some descriptions were inherently ambiguous. For example, the requirement: *"Completion of at least 10.5 credits from the StEOP pool, including at least 3.5 credits from the sub-pool 'Mathematics Basics', and successful completion of the course 'Introduction to Mathematical Work'"* raises the question of whether the latter course is included in the 10.5 credit requirement or must be completed in addition. However, both researchers reached the same conclusion about how to interpret this ambiguity.

A preliminary analysis of the 46 investigated StEOP regulations identified a total of 131 pool rules, of which 43 are mandatory and 88 elective. Each StEOP comprises between one and five pool rules, with most consisting of multiple interacting requirements rather than a single constraint (on average 2.85 pools). Despite this variability, recurring structural patterns can be observed: Four StEOPs consist exclusively of mandatory pools and therefore offer no choice to students, while three consist solely of an unrestricted elective pool, providing maximum flexibility. Ten StEOPs follow a simple structure that combines one mandatory and one elective pool. The most common structure, observed in 20 cases, extends this combination by introducing additional constraints within the elective pool via subpools. These typically require students to complete at least one course from a specific topic area, such as mathematics, or from a particular course type (e.g., lecture or exercise). Furthermore, nine StEOPs combine a mandatory pool with two non-overlapping elective pools. These structures enforce coverage of different subject areas while still allowing individual choice, often balancing subject-specific and foundational or mathematics components. In six of these cases, additional constraints are imposed within the elective pools.

6.3 Discussion

The analysis shows that the pool rule abstraction reliably captures the core structure of StEOP requirements. Only a single special case required an additional constraint that could not be directly modelled with the proposed StEOP formalisation, but which is already addressed by broader curriculum rules. This demonstrates that the proposed formalisation is generally applicable and robust, with the identified exception highlighting a potential extension point for the model without undermining its overall usefulness. Further, the high level of agreement between the two researchers indicates that the proposed modelling approach is robust, even when applied to partially ambiguous curriculum descriptions. While minor differences in naming conventions and modelling decisions were observed, these did not affect the semantic equivalence of the resulting representations.

In addition, the analysis highlights the structural complexity of StEOP regulations. Rather than consisting of a single requirement, most StEOPs comprise multiple interacting constraints, typically combining mandatory and elective components. These are often further restricted by additional conditions, such as selecting at least one course from a specific type or topic area. This structure allows for combining foundational knowledge with some flexibility for individualisation. Although the StEOP regulations are defined independently by different study programs, recurring patterns can be observed, suggesting that they follow implicit design principles.

The formalisation process also revealed challenges. Natural language curriculum descriptions can lack precision, leading to ambiguities in interpretation. While both researchers arrived at semantically equivalent models, some cases allowed multiple plausible interpretations, requiring subjective decisions. This highlights the need for clearer, more standardised, and machine-readable formulations of curriculum rules [9]. While the results demonstrate the general applicability of the proposed model, several limitations remain. The analysis is based on a relatively small sample of 46 StEOP regulations and relies on manual annotation by two researchers, which may limit generalisability. In addition, certain regulatory aspects, such as exemptions due to previously completed courses, are not captured by the current model. Addressing these limitations provides directions for future work. In particular, automating the extraction of the regulations and validating the model against real student data would enable generalisability and allow assessment of how well the formalised rules reflect actual enrolment behaviour. In addition, the proposed formalisation enables several practical applications. For instance, it allows automatically validating whether a planned set of courses satisfies curriculum constraints, thereby supporting study planning tools and reducing invalid course selections. In addition, the model can be used for automated study-planning tools and recommender systems to generate valid study paths under given constraints or to analyse how different curriculum designs affect student progression. By making regulatory structures explicit, it also facilitates their integration into learning analytics systems, where they can be used to contextualise and interpret student behaviour.

7. CONCLUSION

This paper introduces a formal model for representing Study Entry and Orientation Phase (StEOP) regulations as constraint satisfaction problems (CSP), based on nested pool rules over sets of courses with credit-based requirements (RQ1). The model is evaluated on all 46 unique StEOPs at TU Wien since 2017 by applying a semi-automated extraction approach. The results show that all but one analysed regulations can be consistently mapped to the proposed representation, despite being only specified in natural language and lacking standardisation, demonstrating the expressiveness of the model (RQ2). Beyond this, the preliminary analysis revealed recurring structural patterns in curriculum design that are not directly visible in the textual descriptions alone, underlining the need for more standardised and transparent curriculum formulations. The presented formalisation of these curriculum rules not only helps identify such patterns and enable their systematic analysis, but it also supports improvements in curriculum design and enables downstream applications, such as study planning tools. Overall, by formalising the StEOP and evaluating its applicability, this work represents a first step toward transforming curriculum regulations into machine-readable artefacts, facilitating their analysis and integration into EdTech. In this way, the proposed approach contributes to more transparent and data-informed support for students, particularly in early study phases.

8. REFERENCES

[1] P. R. Aldrich. The curriculum prerequisite network: Modeling the curriculum as a complex system. *Biochemistry and Molecular Biology Education*, 43(3):168–180, 2015.

[2] S. C. Brailsford, C. N. Potts, and B. M. Smith. Constraint satisfaction problems: Algorithms and applications. *European journal of operational research*, 119(3):557–581, 1999.

[3] D. Giannakopoulou, T. Pressburger, A. Mavridou, and J. Schumann. Automated formalization of structured natural language requirements. *Information and Software Technology*, 137, 2021.

[4] S. Humpl, K. Scherz, J. Fischer, A. Dibiasi, and M. Unger. Evaluierung der studieneingangs-und orientierungsphase (steop) nach § 66, 2020.

[5] M. Pechenizkiy, N. Trcka, P. De Bra, and P. Toledo. Currim: curriculum mining. In *Proceedings of the 5th International Conference on Educational Data Mining 2012*, pages 216–217. International Educational Data Mining Society (IEDMS), 2012.

[6] E. Roanes-Lozano, A. Martínez-Zarzuelo, and M. J. Fernandez-Diaz. An application of knowledge engineering to mathematics curricula organization and formal verification. *Mathematical Problems in Engineering*, 2020(1), 2020.

[7] R. Roepke, S. Judel, and U. Schroeder. Study path analyses for quality assurance and support of study planning. *Informatik Spektrum*, 47:97–104, Sept. 2024.

[8] M. Trippel and R. Röpke. Developing a Graph-based Visualization of Elective Courses to Support Course Selection in Higher Education. In V. Köppen and S. Strickroth, editors, *23. Fachtagung Bildungstechnologien (DELFI 2025)*, pages 89–95, Bonn, 2025. Gesellschaft für Informatik e.V.

[9] M. von der Heyde, C. Otunuya, M. Goebel, D. Zoerner, and U. Lucke. Automatic and interactive validation of study regulations in accreditation processes of higher education institutions. In *Proceedings of European University Information Systems Congress 2023*, volume 95 of *EPiC Series in Computing*, pages 320–331. EasyChair, 2023.

[10] M. Wagner, H. Helal, R. Roepke, S. Judel, J. Doveren, S. Goerzen, P. Soudmand, G. Lakemeyer, U. Schroeder, and W. M. P. van der Aalst. A Combined Approach of Process Mining and Rule-Based AI for Study Planning and Monitoring in Higher Education. In M. Montali, A. Senderovich, and M. Weidlich, editors, *Process Mining Workshops*, Lecture Notes in Business Information Processing, pages 513–525, Cham, 2023. Springer Nature Switzerland.

[11] K. Wu and W. S. Havens. Modelling an academic curriculum plan as a mixed-initiative constraint satisfaction problem. In *Conference of the Canadian Society for Computational Studies of Intelligence*, pages 79–90. Springer, 2005.

[12] A. V. Zykina, O. N. Kaneva, and V. V. Munko. Developing a curriculum modeling approach. In *CEUR Workshop Proceedings*, volume 2642, 2020.