

Introducing Individuality into Students' High School Timetables

Andreas Krystallidis¹[0009-0007-0183-3038] and Rubén Ruiz-Torrubiano¹[0000-0001-9314-0739]

IMC Krems University of Applied Sciences, Piaristengasse 1, Krems, 3500, Austria
andreas.krystallidis@fh-krems.ac.at
ruben.ruiz@fh-krems.ac.at

Abstract. In a perfect world, each high school student could pursue their interests through a personalized timetable that supports their strengths, weaknesses, and curiosities. While recent research has shown that school systems are evolving to support those developments by strengthening modularity in their curricula, there is often a hurdle that prevents the complete success of such a system: the scheduling process is too complex. While there are many tools that assist with scheduling timetables in an effective way, they usually arrange students into groups and classes with similar interests instead of handling each student individually. In this paper, we propose an extension of the popular XHSTT framework that adds two new constraints to model the individual student choices as well as the requirements for group formation that arise from them. Those two constraints were identified through extensive interviews with school administrators and other school timetabling experts from six European countries. We propose a corresponding ILP formulation and show first optimization results for real-world instances from schools in Germany.

Keywords: Educational Timetabling, High School Timetabling, Integer Linear Programming, XHSTT.

1 Introduction

As educational systems continuously evolve, crafting close-to-optimal high school timetables continues to pose a major challenge. An important aspect that contributes to this phenomenon is the increasing demand for personalized and flexible learning experiences by all stakeholders of the educational system (policy makers, teachers, students and society in general). This demand results in modular educational systems, where students can choose parts of their own curriculum individually. However, conventional approaches to the creation of timetables can not fulfill the diverse needs of students. This paper addresses the need for innovation in the encoding of constraints for high school timetables, aiming to support the flexibility that modern modular educational systems afford to students.

While traditional formats of constraint encoding do support most of the requirements that schools have on timetables including the possibility to work with individual students instead of classes, they are not able to represent the requirements of each student pursuing

diverse individual academic interests in an adaptive manner. In this paper, we present an extension of the XHSTT format [11] (introduced in the Third International Timetabling competition [10]) that incorporates two new constraints, making possible to encode flexible student course choices and class formation requirements while still supporting all previous instances developed for this format. The need for those two new constraint types became apparent through recent research that explored the timetabling demands for schools across central Europe [12] which shows that there is a trend of enabling students to follow their individual interests. Even though this new formulation features students choosing their own respective courses, the problem is still very different from University Course Timetabling Problems [8,7], which were already found to be solvable when transformed to the XHSTT format [4].

While, our new formulation has some similarities to the student choices featured in the Post Enrolment based Course Timetabling (PE-CTT) Problem, which was first presented in the second track of the ITC 2007 [6], the problem we are modeling here has some key differences: In the PE-CTT Problem students select a set of events that they want to attend without providing alternatives or preferences. It is a hard constraint that they visit all their selected events, while in our newly formulated Constraints it is possible to specify that only a subset of a flexible quantity should be attended. Some additional flexibility is provided in the University Course Timetabling Problem featured at the ITC 2019 [9]. While the students still attend a fixed amount of selected courses, the courses themselves can be split into structured subparts that could be used to model distributions of students to equivalent courses (e.g. Math_1_1 and Math_1_2). However, to the best of our knowledge, there is no constraint in any Educational Timetabling format that would enable us to model student choices on the individual level in the flexible manner that is desired in modular educational systems. Finally, since we are solving High School Instances and not University Instances, we use almost all the different constraints of XHSTT, so it is much more efficient to extend this format instead of an University Timetabling format that focuses on other qualities that we mostly do not need (e.g. differentiating between different weeks of the semester and complex orderings and structures of events).

We hope that by extending the standard high school timetable format we will be able to spur new original research that adapts and improves methods of automatic timetabling for modular high schools. We support these developments explicitly by providing an ILP formulation that extends one of the state-of-the-art ILP formulations by Kristiansen et al. [5] as well as 18 publicly available instances that include the new constraints. Additionally, we present first results for upper bounds using our ILP.

This paper is structured as follows: In Section 2, we introduce the new constraints as well as how they can be used and provide examples of how to encode various situations that may occur in modular school timetables. In Section 3, we provide a corresponding extension to the well-known ILP formulation by Kristiansen et al. [5]. In Section 4, we describe the new instances in our format and provide some first results. The instances are made publicly available as a benchmark set for modular high school timetabling. Finally, in Section 5, we provide an overview of our findings and describe possible directions of future research.

2 An Extension to the XHSTT Format

The XHSTT format [11] is the most widely used format for encoding the High School Timetabling Problem. It is versatile enough to encode many real-world instances of timetable requirements from various schools around the world accurately. However, in recent years there has been a trend of giving students more choices in what individual courses they wish to attend. While some of the requirements that arise from those choices can be modeled using the existing constraints, others are not supported by the XHSTT format. Since the goal of the XHSTT format is to provide a way to encode timetable requirements in a unified way, we find it important that those recent developments reflect themselves in the form of an extension of the format. This extension should be as small as possible while still being able to accurately encode the new constraints for timetables. Furthermore, we find that everything that can be encoded with the existing set of constraints should still be encoded using them (even when it is in a slightly roundabout way) in order not to put an unnecessary strain upon those who maintain and possibly want to extend existing methods and solutions for solving the High School Timetabling Problem. Finally, it is also important that the format only includes constraints that are actually useful for those in charge of creating the timetables. That is why the constraints proposed in this paper are chosen based on a study [12] where experts across Europe were asked about the challenges they face when creating timetables for high schools. The result of all of those requirements are three new constraint categories that encode student choices, class size requirements and class size balance. Of those three requirements, the class size requirements will be encoded using the existing constraints of the XHSTT format while the other two require one new constraint type each.

2.1 Student Choices

From the interviews conducted in the paper by Ruiz-Torrubiano et al. [12] it becomes apparent that many schools offer course choices to students in one form or another, especially in the respective upper cycles. Such choices can range from choosing a general direction (profile) for their studies, which usually results in scheduling all students with a given profile together, to individual course choices subscribing them to specific courses together with other students that made the same election. However, once there is a certain level of modularity it is usually in practice infeasible to schedule the courses in such a way that every student can attend exactly the courses they have selected. To deal with that problem schools have adopted two different methods to manage student course choices. Either they create a timetable first, and students choose lectures that fit into their individual schedules, or students give priorities and/or alternatives for their choices and the timetabler tries to fulfill those preferences to the best of their ability. The first possibility is already supported in XHSTT by simply creating courses with no classroom assignment together with some time preference constraints for said courses. However, the second method requires a possibility to model how many from a pool of courses can be attended on an individual resource level. So for example, a student that wishes to learn another language might choose to attend a Spanish course, if they do not get into that course they would like to learn Italian and if there is also no more room in the Italian course they might want to learn French. It is a hard requirement for that

student that they will exactly attend one of those courses. Another student may want to specialize in natural sciences, and they want to attend biology, physics and chemistry eventually while not caring exactly how many of those subjects they will attend in the following year as long as it is at least one. An example for how this constraint may look like can be seen in Figure 1. Additionally, the school might also support that students provide weights (preferences) to their choices. Note that while we designed this constraint to model course choices for students, it can also be used to model teaching preferences, which are also a common theme for many high schools.

All those requirements can be unified into a new constraint type which we call Student Choice Constraint. The constraint has the standard children that all XHSTT constraints share (Id, Name, Required, Weight, CostFunction) the AppliesTo tag consists of Resources and ResourceGroups children. Additionally, the constraint has the child categories EventGroups, Minimum and Maximum.

```

1 <StudentChoiceConstraint Id="StudentChoice_ST_Bob">
2   <Name>StudentChoice_ST_Bob</Name>
3   <Required>true</Required>
4   <Weight>50</Weight>
5   <CostFunction>Linear</CostFunction>
6   <AppliesTo>
7     <ResourceGroups />
8     <Resources>
9       <Resource Reference="ST_Bob" />
10    </Resources>
11  </AppliesTo>
12  <EventGroups>
13    <EventGroup Reference="Biology_10" />
14    <EventGroup Reference="Physics_10" />
15    <EventGroup Reference="Chemistry_10" />
16  </EventGroups>
17  <Minimum>1</Minimum>
18  <Maximum>3</Maximum>
19 </StudentChoiceConstraint>

```

Fig. 1: Student Choice Example

- **AppliesTo:** Each resource that is either part of the Resources child or is part of a resource group, which is mentioned in the ResourceGroups child, is a point of application for this constraint.
- **EventGroups:** All Event Groups mentioned in this child are relevant to the constraint.
- **Minimum:** Each resource this constraint applies to has to attend at least Minimum Event Groups from the relevant Event Groups for this constraint.

- **Maximum:** Each resource this constraint applies to can, at most, attend Maximum Event Groups from the relevant Event Groups for this constraint.

The deviation of the constraint is described as follows: For each resource part of the Resources and ResourceGroups the deviation is equal to the number of Event Groups that the resource attends, which exceed the Maximum or fall short of the Minimum. Note that we define attendance as visiting any subevent of a given Event Group.

This constraint can also be used to encode weighted preferences. Imagine a student who wants to attend one out of 3 courses A , B , and C but has a preference order of $A > B > C$. First, a hard constraint can be added where the EventGroups child contains all 3 courses, and the Minimum and Maximum are set to 1. We then add a soft constraint containing only Events A and B and another constraint containing only event A . Again we set the Minimum and Maximum to 1 for both soft constraints. Depending on the weights of the Soft constraints we can now adjust the importance of the student getting his first or second choice.

2.2 Class Sizes

Whether it is due to room limitations, pedagogic restrictions, or legal reasons (supervision duties), schools usually have limits on how large the classes for each course can be at most. In a system without student choices, this is usually enforced when the classes are put together before scheduling the individual lessons. Events where the classes are split and mixed are often modeled using one main Event (to which the whole class is assigned) and multiple subevents that are all linked to the main event under the assumption that none of the students attend multiple of those subevents. This works, for example, if one wants to split all students from one class level into two math groups, and every student has to choose a second foreign language. However, this method quickly becomes more complex the more individual the student choices become since it most likely won't be possible to build sets of subjects that have no student overlaps while also not creating many idle periods in student timetables and giving all of the students their preferred subjects, which means that there is a need to find some optimal balance between those constraints. Usually there are restrictions that don't allow for any idle time in student timetables during certain periods, while other periods are more flexible (e.g., in the afternoon). Building the classes as part of the optimization problem allows the solver to find which classes to group dynamically based on when it schedules them, also taking into account how important it is to fulfill each individual student's choice. We can model those restrictions using the existing XHSTT constraints, which makes it easier for existing approaches to adapt to those changes. In the following, we describe exactly how to model this class size problem because the translation into XHSTT constraints is not completely trivial. However, first, we want to recap how the XHSTT format works on a high level.

An instance in the XHSTT format consists of Times, Resources, Events, and Constraints. The Events have a duration, which specifies how many time slots must be assigned to them. How exactly those times are distributed over the week is specified as part of the Constraints. We say that each block of consecutive time slots that is

part of the final schedule builds a SolutionEvent (or subevent). Each of those SolutionEvents specifies some resource requirements that can either come in the form of a fixed AssignedResource or a flexible UnassignedResource. In the case of an UnassignedResource, there are often some constraints that restrict the pool of possible concrete resource assignments. Finally, the set of Constraints also consists of other types of Constraints that further impose limits on when, how often, and in what constellations Events, Resources, and groups of Events and Resources are scheduled.

With that basic understanding of the XHSTT format, we can now get into how we modeled the class sizes. First of all, we assume that three resource types exist (but it is possible to arbitrarily add more): Teachers, Rooms, and Individual Students. If fixed classes still exist for certain courses they can easily be modeled by assigning all students of that class directly to that course. For those courses that should be built by the solver, we create three event types:

1. One main event that will be used in all constraints that handle the time assignment of the lessons. This main event will also be used to either directly assign a room and teacher or model the resource preferences for those two resource types. Any restrictions on how the event should be split and distributed over the week will also be applied here
2. As a next step, we create one event for each student resource that is required to fulfill the minimum student number of the course s_{min} (e.g., if it needs at least 10 students to build a language class we create 10 events). They only have one student event resource which is usually not preassigned (can optionally be preassigned to a specific student if attendance is mandatory). We also need a hard Assign and Prefer Resource Constraint so that only those students who choose the course can be assigned and all events must have a student assigned.
3. Next, we need to add a hard Avoid Split Assignments constraint for each Event (modeled with an Event Group that only contains one student Event) which ensures that two subevents of the same Event can't have different student assignments.
4. Afterwards, we add a hard Link Events Constraint to the main Events so that all subevents must have the exact same time assignments as the main Event. Through hard Avoid Clashes Constraints on the individual students this also guarantees that each subevent must have a different student assigned. We will henceforth call Events of this type "minimum requirement events".
5. As a final step, we create Events that are very similar to the previous ones but contain those students that are optional from the Event perspective. We create a total of $s_{max} - s_{min}$ events of this type. All constraints are the same except that we do not use any Assign Resource Constraints since it is fine if no students are assigned to the Event (The Prefer Resource constraints have to stay so that if a student is assigned it must be one that chose the class). We will henceforth call events of this type "maximum requirement events"

Note that it would be possible to combine the student requirement events into a single event from a modeling perspective (using a separate Role for each student). We decided against this approach in case that some existing solvers might enumerate all combinations of resource assignments for each Event, which would lead to an exponential amount of such combinations. In all other aspects the approaches are to the

best of our knowledge equivalent, except that it would be possible to directly quantify over b variables in the two new proposed constraint types. However, quantifying over EventGroups instead of individual Events gives the advantage that we can use more "high level" constraints, for example if we go back to the example provided in Listing 1 Biology_10, Physics_10 and Chemistry_10 could each be an EventGroup that represents multiple courses (e.g. Biology_10_1, Biology_10_2 and Biology_10_3). We would then add three more Student Choice Constraints (one for each subject) for ST_Bob each with a Minimum set to 0 and Maximum set to 1. The result is that on the high level Bob will visit between 1 and 3 of his selected choices and on the lower level he will be assigned to exactly one course corresponding to the assigned subjects.

2.3 Class Size Balance

Sometimes one subject is taught in multiple courses handling exactly the same school material because the number of students is too big for one single classroom and teacher. One way to handle this using existing constraints is to simply use two teachers and rooms for the class while also scaling the student requirements. However, this has the restriction that both of those courses would need to happen in parallel, which takes away some flexibility, especially when the student schedules are very individual.

A better alternative would be to have two completely separate courses. With the help of Student Choice constraints, we can then model that a student can or must attend one of them. However, this could result in the undesirable property of possibly very unbalanced class sizes (e.g., two math classes, one with the bare minimum assignment of 10 students while the other is fully booked with 30 students). To prevent this, we introduce Balance Class Size constraints that, aside from the standard children that all XHSTT constraints share (Id, Name, Required, Weight, CostFunction), have the tags AppliesTo with the child EventGroups, Role and MaximumDifference. An example for how this constraint would be modeled in the case of two equivalent math classes can be found in Figure 2.

- **AppliesTo:** Each Event Group that is mentioned in the EventGroups child is relevant for this constraint.
- **MaximumDifference:** An integer that sets a limit of how much difference between the number of assigned resources can be between the Event Groups without causing a deviation.
- **Type:** Optional child. If a Type is given only resources with this type are counted towards the assigned resources of each event.

The deviation of this constraint is described as follows: For each Event Group that is part of the EventGroups child, the deviation is calculated as the Maximum Difference to the Event Group (part of the Event Groups Child) with either the most or the least assigned resources (depending on which difference is higher) minus the allowed MaximumDifference.

```

1 <BalanceClassSizeConstraint Id="BalanceClassSize_Math_5a">
2   <Name>BalanceClassSize_Math_5a</Name>
3   <Required>>false</Required>
4   <Weight>1</Weight>
5   <CostFunction>Linear</CostFunction>
6   <AppliesTo>
7     <EventGroups>
8       <EventGroup Reference="Math_1_5A"/>
9       <EventGroup Reference="Math_2_5A"/>
10    </EventGroups>
11  </AppliesTo>
12  <MaximumDifference>2</MaximumDifference>
13  <Type>Student</Type>
14 </BalanceClassSizeConstraint>

```

Fig. 2: Balance Class Size Example

3 ILP Formulation

In this section we introduce an ILP model that can be used to solve an instance of our extension to the High School Timetabling Problem. For this purpose we will extend the formulation from Kristiansen et al. [5] by our two new constraints as well as the relevant variables and linkings. In this section, we will only describe the new constraints. The full ILP formulation can be found in the Appendix. Note that for the moment our new instances and format only support linear and quadratic cost functions.

3.1 Sets

First, we introduce some sets of entities relevant to the extended modular XHSTT problem which are the same as used by the model for the original problem [5].

$t \in T$	ordered set of times	(1)
$tg \in TG$	set of time groups	(2)
$r \in R$	set of resources	(3)
$e \in E$	set of events	(4)
$eg \in EG$	set of event groups	(5)
$er \in e$	set of event resources of an event e	(6)
$se \in e$	set of subevents of an event e	(7)
$c \in C$	set of constraints	(8)
$p \in c$	points of application of constraint c	(9)
$d \in p$	deviations of a point of application p	(10)
$i \in I$	possible deviation values of deviations d	(11)

$j \in J$	possible deviation sum values at points p (12)
$T_{se,t}^{\text{start}}$	possible start times for se that occupy t (13)

3.2 Further Notation

Next, we need some further notation to express some of the constraints.

r_D	dummy-resource with meaning no resource assigned (14)
t_D	dummy-time with meaning no time assigned (15)
D_e	duration of event e (16)
D_{se}	duration of subevent se (17)
$e \in c$	constraint c applies to event e (18)
$r \in c$	constraint c applies to resource r (19)
$eg \in c$	constraint c applies to event group eg (20)
w_c	weight of constraint c (21)
$\rho(t)$	index of time t in ordered set T (22)
PA_{er}	is 1 if event resource er has a preassigned resource otherwise 0 (23)
\overline{B}_c	upper limit of constraint c (24)
\underline{B}_c	lower limit of constraint c (25)

3.3 Variables

Compared to the formulation by Kristiansen et al. [5], we added the variables b and c which are required to model if a resource is participating in Events or Event Groups.

$x_{se,t,er,r}$	binary, indicates if se starts at t and r is assigned to er (26)
$y_{se,t}$	binary, indicates that se starts at t (27)
$v_{t,r}$	integer, indicates how often r is used at t by any se (28)
$w_{se,er,r}$	binary, indicates if se is assigned r for er (29)
$b_{e,r}$	binary, indicates if r is assigned to any se of e (30)
$c_{eg,r}$	binary, indicates if r is assigned to any e of eg (31)
$s_{c,p,d}$	integer, deviation d at point of application p of c (32)
$s_{c,p,d,i}$	binary, indicates that d has value i at p of c (33)
$u_{c,p,j}^{\text{SquareSum}}$	binary, indicates that the sum of deviations at p is j (34)
u_{se}	binary, indicates whether se is active or not (35)
$q_{r,t}$	binary, indicates if r is busy at t (36)
pr,tg	binary, indicates if r is busy at some t in tg (37)

3.4 Functions

We also need to introduce some functions that will mainly be used to express the objective value of the problem.

$$f(s_{c,p,d}) = w_c \cdot \text{CostFunction}(s_{c,p,d}) \quad \text{cost of constraint } c \quad (38)$$

$$CF^{\text{Sum}} = \sum_{p \in c, d \in p} s_{c,p,d} \quad \text{Sum cost function} \quad (39)$$

$$CF^{\text{SumSquare}} = \sum_{p \in c, d \in p, i \in I} i^2 \cdot s_{c,p,d,i} \quad \text{SumSquare cost function} \quad (40)$$

$$CF^{\text{SquareSum}} = \sum_{p \in c, j \in J} j^2 \cdot u_{c,p,j}^{\text{SquareSum}} \quad \text{SquareSum cost function} \quad (41)$$

Some constraints have an upper limit and a lower limit. In this case we define the value of deviation V using the function $U_{\underline{B}_c, \bar{B}_c} V$ as follows:

$$s \geq U_{\underline{B}_c, \bar{B}_c} V \rightarrow \begin{cases} s \geq V - \bar{B}_c \\ s \geq \underline{B}_c - V \end{cases} \quad \text{deviation with upper and lower limit} \quad (42)$$

3.5 Updated Objective Function

The objective function consists of the sum of all cost functions of individual constraints. We can also split this objective function into separate values z_{hard} and z_{soft} to denote the costs of hard and soft constraints respectively. Compared to the objective function by Kristiansen et al. [5], we simply extended the function by adding the terms describing the deviation of our new constraint types.

$$\begin{aligned} \min z = & f(s_{c,er}^{\text{assignres}}) + f(s_{c,er}^{\text{assigntime}}) + f(s_{c,e}^{\text{spliteventamount}} + s_{c,e}^{\text{spliteventdur}}) \\ & + f(s_{c,e,er}^{\text{distsplitevent}}) + f(s_{c,er}^{\text{preferres}}) + f(s_{c,e}^{\text{prefertime}}) + f(s_{c,eg}^{\text{avoidsplit}}) \\ & + f(s_{c,eg,tg}^{\text{spreadevent}}) + f(s_{c,eg,t}^{\text{linkevent}}) + f(s_{c,r,t}^{\text{avoidclashes}}) + f(s_{c,r}^{\text{unavailabletimes}}) \quad (43) \\ & + f(s_{c,r}^{\text{idletimes}}) + f(s_{c,r}^{\text{clusterbusy}}) + f(s_{c,r,tg}^{\text{limitbusy}}) + f(s_{c,r}^{\text{limitworkload}}) \\ & + f(s_{c,eg}^{\text{balancesize}}) + f(s_{c,r}^{\text{studentchoice}}) \end{aligned}$$

3.6 Constraints

Added General Constraints

The model needs several linking constraints and other general constraints to make the variables express the above-described properties. We will only describe those general constraints that were added to the model of Kristiansen et al. [5] the remaining constraints can be found in the Appendix.

The following new constraints link variables $w_{se,er,r}$ and our new variables $b_{e,r}$:

$$w_{se,er,r} \leq b_{e,r} \quad \forall e \in E, se \in e, er \in e, r \in R \quad (44)$$

$$\sum_{se \in e} w_{se,er,r} \geq b_{e,r} \quad \forall e \in E, er \in e, r \in R \quad (45)$$

To link our new variables $c_{eg,r}$ to $b_{e,r}$ we need two more new constraints:

$$b_{e,r} \leq c_{eg,r} \quad \forall eg \in EG, e \in eg \quad (46)$$

$$\sum_{e \in eg} b_{e,r} \geq c_{eg,r} \quad \forall eg \in EG \quad (47)$$

Balance Class Size Constraint

Applies to: Event Groups

Point-of-application: Event Group

We use the parameter \bar{B}_c to denote the maximum class size difference specified in constraint $c \in \bar{C}$. The role parameter is optional and denotes that only resources of a specific role in the event should be considered. We use the variable $mr_{c,eg}$ to denote the number of resources allocated to the specified Event Group

$$\sum_{\substack{e \in eg, er \in e, \\ r \in er, type_r = type_c \setminus \{rD\}}} c_{eg,r} = mr_{c,eg} \quad \forall c \in \bar{C}, eg \in \bar{C} \quad (48)$$

$$mr_{c,eg} - mr_{c,eg2} - \bar{B}_c \leq s_{c,eg}^{balance_size} \quad \forall c \in \bar{C}, eg \in c, eg2 \in c, eg \neq eg2 \quad (49)$$

$$mr_{c,eg2} - mr_{c,eg} - \bar{B}_c \leq s_{c,eg}^{balance_size} \quad \forall c \in \bar{C}, eg \in c, eg2 \in c, eg \neq eg2 \quad (50)$$

Student Choice Constraint

Applies to: Resources

Point-of-application: Resource

We use parameters \underline{B}_c and \bar{B}_c to denote the minimum and maximum values specified in constraint $c \in \bar{C}$.

$$U_{\underline{B}_c, \bar{B}_c} \sum_{eg \in c, e \in eg, er \in e, r \in er, r=r_c} c_{eg,r} \leq s_{c,r}^{studentchoice} \quad \forall c \in \bar{C}, r_c \in c \quad (51)$$

3.7 Model Size Reductions

Using the model described above without any additions results in models that are too big to handle on our computing cluster with 64 GB of RAM for most of the instances. For that reason, we eliminated some variables that would never be used for an acceptable solution. Our variable eliminations consist of the following list:

- We eliminated all variables $x_{se,t,er,r}$ and $w_{se,er,r}$ for students r that did not select an Event e , $se \in e$. This means that we do not permit solutions where students who did not select an event are assigned to one of its subevents (which is supported by practice).
- We only generate subevents with a feasible duration if there is a hard split events constraint restricting the duration and/or amount of generated subevents. [3] This is the same technique that was used by Fonseca et al. [2] to reduce model sizes.

4 Evaluation

4.1 Instances

As a first benchmarking set¹, we chose 18 high schools with modular school systems from 6 different federal states in Germany. Note that the secondary educational systems in Germany can vary greatly depending on the particular federal state [12], which makes this set more diverse than instance groups from most other countries. The original anonymized instances were provided by Untis GmbH², an Austrian company that specializes in software that assists schools with their various scheduling problems. We implemented methods to automatically translate their format for encoding constraints to the new extended XHSTT format, which enables us to provide many more instances in the future (Untis collaborates with over 26,000 schools worldwide). However, it is important to note that the XHSTT instances are not a one to one match semantically with the original instances provided by Untis. This is due to the complex nature of the Untis specification that uses a lot of empirical experience to evaluate timetables on factors that can't be represented in a standardized format. We still managed to achieve an extended XHSTT formulation that matches the Untis formulation closely. Some statistics of the instances can be found in Tables 1 and 2. Table 1 describes how many resources of each resource type are used in each instance. Note that we do not include the minimum/maximum requirement events in the event count since we categorize them as part of the original event and they will always be scheduled together. However, we list the number of requirement events as well as other quantifiable properties that describe the modularity of each instance in Table 2. Note that the amount of requirement events is equal to the number of student assignments that can (but don't necessarily have to) happen to modular events. Table 2 lists how many Student Choice and Balance Class Size constraints each instance uses. The column Modular Events describes the cardinality of the subset of events that have minimum and/or maximum requirement events associated with them. However, there are many more factors that can have an impact on how complex the resulting instance will be. One factor of complexity is the number and type of constraints from the original XHSTT problem definition. Another factor that has a significant impact is the size of the student pool that is feasible for each requirement event. An instance will be much harder if it features some events where a high percentage of the total amount of students wants to participate in certain events. It is also noteworthy that compared to most other benchmark instances of the original XHSTT

¹<https://github.com/IMC-UAS-Krems/modularXHSTT>

²<https://www.untis.at>

problem, this instance set is more restrictive on the possible times for events and the available times for students and teachers. Specifically, there are constraints restricting how many primary subjects a student can attend per day/in a row, how many lessons a teacher may teach in a row without a break, minimum and maximum amounts of idle times per week for teachers, global constraints that define a time-window when a lunch break can/must happen for both students and teachers and hard restrictions that allow no student idle times in the hours before lunch. We made sure to choose schools of varying sizes and proportions of modular events so that it will be possible in the future to find out where the complexities of the problem lie. We also plan to extend this set of instances in the future with other modular schools from across Europe to cover more of the possible instance space.

Table 1: Amount of resources present in each of the new instances (by resource type).

Instance	Events	Students	Classes	Teachers	Rooms
GermanyRHPF1	1430	143	141	142	208
GermanyHAMB1	554	847	10	104	80
GermanyNRWE1	627	118	52	185	186
GermanyHAMB2	636	215	43	106	73
GermanyNRWE2	329	252	16	40	69
GermanyRHPF2	414	134	25	88	74
GermanyRHPF3	454	167	25	92	84
GermanyNRWE3	226	252	16	40	69
GermanyNRWE4	1045	331	112	185	310
GermanyRHPF4	545	894	0	83	106
GermanySAAR1	428	169	25	75	65
GermanyRHPF5	375	182	21	75	61
GermanyRHPF6	526	360	30	107	67
GermanyBAWU1	832	157	76	201	176
GermanyBAWU2	976	272	59	122	173
GermanyBAWU3	241	228	22	92	71
GermanyBAWU4	762	205	24	106	70
GermanyHESS1	705	181	49	177	225

4.2 ILP Evaluation

Based on the size of the new instances and the added complexity from the new constraints we expect the new instances to be more difficult to solve than previous benchmark sets for the XHSTT. Previous experiments [5] using ILP as an exact method have shown that the problem is still too challenging for modern ILP solvers when using bigger instances. Therefore, we expect to only produce weak upper bounds using the above-described ILP model. Nevertheless, we find it important to provide first results for the newly introduced problem, which should show whether the problem is trivial to solve or not.

Table 2: Quantifiable properties describing the modularity of the new instances.

Instance	Choice Constraints	Balance Constraints	Modular Events	Requ. Events
GermanyRHPF1	2378	34	142	2440
GermanyHAMB1	911	42	46	970
GermanyNRWE1	1254	60	73	1265
GermanyHAMB2	2053	62	112	2206
GermanyNRWE2	2408	56	106	2893
GermanyRHPF2	1476	40	93	1385
GermanyRHPF3	1718	50	109	1892
GermanyNRWE3	1677	35	73	1942
GermanyNRWE4	3433	79	179	3828
GermanyRHPF4	3984	100	241	4063
GermanySAAR1	1421	57	81	1596
GermanyRHPF5	1904	41	124	2622
GermanyRHPF6	2567	21	131	3255
GermanyBAWU1	1452	45	76	1764
GermanyBAWU2	3743	84	176	4112
GermanyBAWU3	2473	75	162	2496
GermanyBAWU4	2309	61	160	2347
GermanyHESS1	2947	24	47	1189

While employing ILP as an exact method might not yield practical solutions directly, past research suggests its potential when integrated into metaheuristic or matheuristic approaches for tackling the High School Timetabling Problem [1,2].

Table 3 displays the outcomes of executing the 18 instances, with a Memory Limit set at 64 GB, on an AMD EPYC 7252 processor utilizing the commercial ILP solver Gurobi version 10.0.1. Each instance was allotted a time limit of 6 hours for computation. This time frame notably exceeds the 1000-second constraint imposed during the ITC2011 competition. The objective value is split into two components of the form (hard, soft) constraint deviations.

It's worth noting that in practical scenarios, schools typically face fewer time constraints when devising their timetables. They are often willing to invest several hours, or even days, in computational time without significant pressure. However, it's crucial to strike a balance between computational resources and research accessibility. Excessive resource consumption could potentially limit accessibility to the problem, which runs counter to the goal of fostering open research.

The experiment reveals the complexity inherent in the problem, suggesting that exact methods may struggle to provide satisfactory solutions. Out of the 18 instances studied, solutions were only found for 10.

Furthermore, the integral solutions obtained from the experiment did not meet the criteria necessary for a viable school timetable. Despite leveraging Gurobi's optimization capabilities, the solutions fell short, underscoring the intricacies of the problem and the limitations of current methodologies.

Table 3: ILP results and statistics on 6-hour run.

Instance	Variables	Constraints	Objective Value
GermanyRHPF1	9007909	67145239	—
GermanyHAMB1	22276651	61355937	—
GermanyNRWE1	5282717	22736030	(152422, 2330)
GermanyHAMB2	6268589	19983356	—
GermanyNRWE2	12183297	41285751	—
GermanyRHPF2	3013943	9066705	(125430, 4663)
GermanyRHPF3	6422201	21112371	(157904, 86000)
GermanyNRWE3	8581837	30750272	(75295, 206299)
GermanyNRWE4	23744735	88931169	—
GermanyRHPF4	42657262	87422983	—
GermanySAAR1	4695284	14566661	(142059, 23465)
GermanyRHPF5	5870336	14030401	(117175, 379454)
GermanyRHPF6	20876398	53277518	—
GermanyBAWU1	7567498	27892830	(231653, 70076)
GermanyBAWU2	11729612	31872156	(257726, 348709)
GermanyBAWU3	8464935	18952959	(81008, 3416)
GermanyBAWU4	7740808	19141384	(193313, 8665)
GermanyHESS1	5737775	29194865	—

Additionally, Gurobi's inability to provide lower bounds within the designated time-frame prevents us from assessing the optimality gap. However, there is potential for progress as the schools that provided the instances have successfully created their own timetables. In general it is not always possible to find feasible solutions for the requirements encoded by schools. Untis deals with this problem by either leaving some hours unscheduled or reporting the problems with the final timetable that will then be manually resolved by the administrator, which then has to decide which hard constraints can be softened. Nevertheless those solutions may enable us to establish tighter upper bounds in the future, enhancing our understanding of the problem and potentially guiding optimization strategies.

5 Conclusion and Future Work

In this paper, we proposed an extension to the XHSTT format to address the constraints that are present in the ever-growing number of modular high schools. We explained the reasons that make those changes meaningful and how exactly they can be incorporated into the XHSTT format. Furthermore, we provided 18 new real-world instances from modular high schools of different regions in Germany together with an ILP model that can be used to find feasible schedules. Our experiments showed that using our ILP model as an exact method for finding solutions is not very effective, and even after 6 hours of runtime, it could only find solutions that are nowhere near satisfactory. However, based on the data from previous benchmarks those results were expected and should not

discourage us from finding more efficient methods for creating timetables for modular high schools. We believe that the real strength of the presented ILP will be revealed once it is used as part of a heuristic approach like Large Neighborhood Search (LNS).

In future work, we want to use different exact methods, like a SAT solver to inspect if the problem is really as hard as it seems to be or if ILP is simply not the right approach for this problem. There has already been previous work into possible cuts for the original ILP model and we plan to look into possible new cuts for our extension as well. It will also be interesting to see how the various heuristics developed for the original High School Timetabling Problem perform on this extension and if new heuristics can be found that might work even better for this extension. Specifically, we want to look into a more adaptive LNS-based approach to see if methods from Reinforcement Learning can be used to improve the process of finding good schedules. We will support those developments by providing more benchmark instances from all across Europe together with a publicly available tool for validation. Finally, we will also compare future findings with the timetables produced for actual schools to see if there is any gap between theory and practice.

Acknowledgements A. Krystallidis and R. Ruiz-Torrubiano acknowledge financial support from the Research Promotion Agency of Lower Austria (GFF) under project grant FTI21-A-002.

Appendix

Complete ILP Formulation

The variables, sets, functions and general further notation of the model can be found in Section 3. Here we provide the full set of constraints used in the model (with some repetition from Section 3 to avoid confusion), as well as some further additions to guarantee exact objective values.

Constraints

General Constraints Link variables $s_{c,p,d}$ and $s_{c,p,d,i}$:

$$\sum_{i \in I} i * s_{c,p,d,i} = s_{c,p,d} \quad \forall c \in C, p \in c, d \in c \quad (52)$$

Only one deviation indicator can be set per deviation:

$$\sum_{i \in I} s_{c,p,d,i} = 1 \quad \forall c \in C, p \in c, d \in c \quad (53)$$

Link variables $s_{c,p,d}$ and $u_{c,p,j}^{\text{SquareSum}}$:

$$\sum_{j \in J} j * u_{c,p,j}^{\text{SquareSum}} = \sum_{d \in p} s_{c,p,d} \quad \forall c \in C, p \in c \quad (54)$$

Only one deviation indicator can be set per point of application:

$$\sum_{j \in J} u_{c,p,j}^{\text{SquareSum}} = 1 \quad \forall c \in C, p \in c \quad (55)$$

Link variables $s_{c,p,d}$ and u_c^{StepSum} :

$$M \cdot u_c^{\text{StepSum}} \geq s_{c,p,d} \quad \forall c \in C, p \in c, d \in p \quad (56)$$

A subevent is assigned exactly one starting time and the number of assigned resources equals the number of event resources ($|er|_{se}$):

$$\sum_{t \in T, r \in er} x_{se,t,er,r} = 1 \quad \forall se \in SE, er \in se \quad (57)$$

$$\sum_{er \in se, r \in er} x_{se,t,er,r} = |er|_{se} \cdot y_{se,t} \quad \forall se \in SE, t \in T \quad (58)$$

Link variables $v_{t,r}$ and $w_{se,er,r}$:

$$\sum_{se \in SE, er \in se, t' \in T_{se,t}^{\text{start}}} x_{se,t',er,r} = v_{t,r} \quad \forall t \in T \setminus \{t_D\}, r \in R \quad (59)$$

$$\sum_{t \in T} x_{se,t,er,r} = w_{se,er,r} \quad \forall se \in SE, er \in se, r \in er \quad (60)$$

The following new constraints link variables $w_{se,er,r}$ and our new variables $b_{e,r}$:

$$w_{se,er,r} \leq b_{e,r} \quad \forall e \in E, se \in e, er \in e, r \in R \quad (61)$$

$$\sum_{se \in e} w_{se,er,r} \geq b_{e,r} \quad \forall e \in E, er \in e, r \in R \quad (62)$$

To link our new variables $c_{eg,r}$ to $b_{e,r}$ we need two more new constraints:

$$b_{e,r} \leq c_{eg,r} \quad \forall eg \in EG, e \in eg \quad (63)$$

$$\sum_{e \in eg} b_{e,r} \geq c_{eg,r} \quad \forall eg \in EG \quad (64)$$

A subevent can not be assigned a start time that does not have enough times after it to fit its duration:

$$y_{se,t} = 0 \quad \forall se \in SE, t \in T \setminus \{t_D\}, \rho(t) + D_{se} - 1 > |T| \quad (65)$$

Only a subset of the subevents are active at a time (since we create all possible subevents).

A subevent is considered active if it has a starting time or resource assigned:

$$\sum_{r \in er \setminus \{r_D\}} w_{se,er,r} \leq u_{se} \quad \forall se \in SE, er \in se, PA_{er} = 0 \quad (66)$$

$$\sum_{t \in T \setminus \{t_D\}} y_{se,t} \leq u_{se} \quad \forall se \in SE \quad (67)$$

$$\sum_{t \in T \setminus \{t_D\}} y_{se,t} + \sum_{\substack{r \in er \setminus \{r_D\}, \\ er \in se, PA_{er} = 0}} w_{se,er,r} \geq u_{se} \quad \forall se \in SE \quad (68)$$

The sum of the durations of the subevents of an event must equal the total duration of the event:

$$\sum_{se \in e} D_{se} * u_{se} = D_e \quad \forall e \in E \quad (69)$$

Linking the variables $q_{r,t}$ and $p_{r,tg}$ that indicate if a resource is busy:

$$|SE| \cdot q_{r,t} \geq v_{t,r} \quad \forall r \in R, t \in T \setminus \{t_D\} \quad (70)$$

$$q_{r,t} \leq v_{t,r} \quad \forall r \in R, t \in T \setminus \{t_D\} \quad (71)$$

$$p_{r,tg} \geq q_{r,t} \quad \forall r \in R, tg \in TG, t \in tg \quad (72)$$

$$p_{r,tg} \leq \sum_{t \in tg} q_{r,t} \quad \forall r \in R, tg \in TG \quad (73)$$

(74)

Events that have a given start time must have that time assigned (se^* , represents an arbitrarily chosen subevent that has the same duration as the event e):

$$y_{se^*,eTime} = 1 \quad \forall e \in E, eTime \neq None, se^*Duration = eDuration \quad (75)$$

Assign Resource Constraint

Applies to: Events

Point-of-application: Event resource

$$D_e - \sum_{\substack{se \in e, \\ r \in er \setminus \{r_D\}}} D_{se} \cdot w_{se,er,r} = s_{c,er}^{\text{assignres}} \quad \forall c \in \bar{C}, e \in c, er \in e, role_{er} = role_c \quad (76)$$

Assign Time Constraint

Applies to: Events

Point-of-application: Event

$$D_e - \sum_{t \in T \setminus \{t_D\}, se \in e} D_{se} \cdot y_{se,t} = s_{c,e}^{\text{assigntime}} \quad \forall c \in \bar{C}, e \in c \quad (77)$$

Split Events Constraint

Applies to: Events

Point-of-application: Event

We use the parameters $\underline{B}_c^{\text{amount}}$ and $\bar{B}_c^{\text{amount}}$ to denote the minimum and maximum amount of events respectively. Likewise, we use the parameters $\underline{B}_c^{\text{dur}}$ and \bar{B}_c^{dur} to denote the minimum and maximum duration of a subevent that is part of a given event, respectively. The full deviation of a constraint $c \in \bar{C}$ is given by $s_{c,e}^{\text{spliteventamount}} + s_{c,e}^{\text{spliteventdur}}$

$$U_{\underline{B}_c^{\text{amount}}, \bar{B}_c^{\text{amount}}} \sum_{se \in e} u_{se} \leq s_{c,e}^{\text{spliteventamount}} \quad \forall c \in \bar{C}, e \in c \quad (78)$$

$$\sum_{se \in e, \underline{B}_c^{\text{dur}} > D_{se}, \bar{B}_c^{\text{dur}} < D_{se}} u_{se} = s_{c,e}^{\text{spliteventdur}} \quad \forall c \in \bar{C}, e \in c \quad (79)$$

Distribute Split Events Constraint

Applies to: Events

Point-of-application: Event

We use the parameters \underline{B}_c and \bar{B}_c to denote the minimum and maximum number of subevents respectively. D_c denotes the duration for which the constraint applies.

$$U_{\underline{B}_c, \bar{B}_c} \sum_{se \in e, D_{se} = D_c} u_{se} \leq s_{c,e,er}^{\text{distsplitevent}} \quad \forall c \in \bar{C}, e \in c \quad (80)$$

Prefer Resources Constraint

Applies to: Events

Point-of-application: Event resource

$$\sum_{\substack{se \in e, r \notin c, \\ r \in R \setminus \{r_D\}}} D_{se} \cdot w_{se,er,r} = s_{c,er}^{\text{preferres}} \quad \forall c \in \bar{C}, e \in c, er \in e, PA_{er} = 0, role_{er} = role_c \quad (81)$$

Prefer Times Constraint

Applies to: Events

Point-of-application: Event

If D_c is given only sub-events of Duration D_c are considered. Otherwise, all sub-events are considered ($D_c = D_{se}$ is removed from sum).

$$\sum_{\substack{se \in e, t \notin c, \\ t \in T \setminus \{t_D\}, D_c = D_{se}}} D_{se} \cdot y_{se,t} = s_{c,e}^{\text{prefertime}} \quad \forall c \in \bar{C}, e \in c \quad (82)$$

Avoid Split Assignments Constraint

Applies to: Event Groups

Point-of-application: Event Group

We slightly simplify this constraint compared to the original formulation, since we can make use of our new c variables.

$$\sum_{\substack{er \in e, PA_{er}=0, \\ role_c = role_{er}}} w_{se,er,r} \leq k_{c,eg,r} \quad \forall c \in \bar{C}, r \in R, eg \in c, e \in eg, se \in e \quad (83)$$

$$\sum_{r \in R} k_{c,eg,r} - 1 \leq s_{c,eg}^{\text{avoidsplit}} \quad \forall c \in \bar{C}, eg \in c \quad (84)$$

Spread Events Constraint

Applies to: Event Groups

Point-of-application: Event Group

We use parameters $\underline{B}_{c,tg}$ and $\overline{B}_{c,tg}$ to denote the minimum and maximum number of sub-events of a given event that can be placed in time group tg of a constraint $c \in \bar{C}$

$$U_{\underline{B}_{c,tg}, \overline{B}_{c,tg}} \sum_{se \in e \in eg, t \in tg} y_{se,t} \leq s_{c,eg,tg}^{\text{spreadevent}} \quad \forall c \in \bar{C}, eg \in c, tg \in c \quad (85)$$

Link Events Constraint

Applies to: Event Groups

Point-of-application: Event Group

We define the binary variable $o_{e,t}$ that takes the value 1 if at least one sub-event of event

e is scheduled at time t , and 0 otherwise. We define the binary variable $l_{eg,t}$ that takes the value 1 if at least one event in event group eg is scheduled at time t , and 0 otherwise.

$$\sum_{t' \in T_{se,t}^{\text{start}}} y_{se,t'} \leq o_{e,t} \quad \forall e \in E, se \in e, t \in T \setminus \{t_D\} \quad (86)$$

$$\sum_{se \in e, t' \in T_{se,t}^{\text{start}}} y_{se,t'} \geq o_{e,t} \quad \forall e \in E, t \in T \setminus \{t_D\} \quad (87)$$

$$l_{eg,t} \geq o_{e,t} \quad \forall eg \in EG, e \in eg, t \in T \setminus \{t_D\} \quad (88)$$

$$l_{eg,t} - o_{e,t} \leq s_{c,eg,t}^{\text{linkevent}} \quad \forall c \in \bar{C}, eg \in c, e \in eg, t \in T \setminus \{t_D\} \quad (89)$$

Order Events Constraint

Applies to: Pairs of Events

Point-of-application: Pair of Events

The variables h_e^{first} and h_e^{last} represent the first and last time assigned to any subevent of event e . We use parameters \underline{B}_c and \bar{B}_c to denote the minimum and maximum number of times to separate the pair of events (e, e') which are specified in constraint $c \in \bar{C}$.

$$\rho(t) \cdot y_{se,t} + D_{se} \leq h_e^{\text{last}} \quad \forall c \in \bar{C}, e \in c, se \in e, t \in T \quad (90)$$

$$|T| - (|T| - \rho(t)) \cdot y_{se,t} \leq h_e^{\text{first}} \quad \forall c \in \bar{C}, e \in c, se \in e, t \in T \quad (91)$$

$$U_{\underline{B}_c, \bar{B}_c} (h_e^{\text{last}} - h_{e'}^{\text{first}}) \leq s_{c,(e,e')}^{\text{orderevent}} \quad \forall c \in \bar{C}, (e, e') \in c \quad (92)$$

Avoid Clashes Constraint

Applies to: Resources

Point-of-application: Resource

$$v_{t,r} - 1 \leq s_{c,r,t}^{\text{avoidclashes}} \quad \forall c \in \bar{C}, r \in c, t \in T \setminus \{t_D\} \quad (93)$$

Avoid Unavailable Times Constraint

Applies to: Resources

Point-of-application: Resource

$$\sum_{t \in c} q_{r,t} = s_{c,r}^{\text{unavailabletimes}} \quad \forall c \in \bar{C}, r \in c \quad (94)$$

Limit Idle Times Constraint

Applies to: Resources

Point-of-application: Resource

We define the binary variables $h_{r,tg,t}^{\text{before}}$ and $h_{r,tg,t}^{\text{after}}$ to indicate if any events are happening

before and after time t in the timegroup tg , respectively. We define the binary variables $h_{r,tg,t}^{\text{timeslot}}$ to indicate if time t is an idle time. We define the integer variable $h_{r,tg}^{\text{timegroup}}$ to indicate the total amount of idle times in timegroup tg . We use $|tg|$ to indicate the amount of times in a time group tg . We also use parameters \underline{B}_c and \overline{B}_c to denote the minimum and maximum values specified in constraint $c \in \overline{C}$.

$$q_{r,t_2} \leq h_{r,tg,t_1}^{\text{before}} \quad \forall r \in C_R, tg \in C_{TG}, t_1, t_2 \in tg, \rho(t_1) > \rho(t_2) \quad (95)$$

$$\sum_{t_2 \in tg, \rho(t_1) > \rho(t_2)} q_{r,t_2} \geq h_{r,tg,t_1}^{\text{before}} \quad \forall r \in C_R, tg \in C_{TG}, t_1 \in tg \quad (96)$$

$$q_{r,t_2} \leq h_{r,tg,t_1}^{\text{after}} \quad \forall r \in C_R, tg \in C_{TG}, t_1, t_2 \in tg, \rho(t_1) < \rho(t_2) \quad (97)$$

$$\sum_{t_2 \in tg, \rho(t_1) < \rho(t_2)} q_{r,t_2} \geq h_{r,tg,t_1}^{\text{after}} \quad \forall r \in C_R, tg \in C_{TG}, t_1 \in tg \quad (98)$$

$$h_{r,tg,t}^{\text{before}} - q_{r,t} + h_{r,tg,t}^{\text{after}} - 1 \leq h_{r,tg,t}^{\text{timeslot}} \quad \forall r \in C_R, tg \in C_{TG}, t \in tg \quad (99)$$

$$-q_{r,t} + 1 \geq h_{r,tg,t}^{\text{timeslot}} \quad \forall r \in C_R, tg \in C_{TG}, t \in tg \quad (100)$$

$$h_{r,tg,t}^{\text{before}} \geq h_{r,tg,t}^{\text{timeslot}} \quad \forall r \in C_R, tg \in C_{TG}, t \in tg \quad (101)$$

$$h_{r,tg,t}^{\text{after}} \geq h_{r,tg,t}^{\text{timeslot}} \quad \forall r \in C_R, tg \in C_{TG}, t \in tg \quad (102)$$

$$\sum_{t \in tg} h_{r,tg,t}^{\text{timeslot}} = h_{r,tg}^{\text{timegroup}} \quad \forall r \in C_R, tg \in C_{TG} \quad (103)$$

$$U_{\underline{B}_c, \overline{B}_c} \sum_{tg \in c} h_{r,tg}^{\text{timegroup}} \leq s_{c,r}^{\text{idletimes}} \quad \forall c \in \overline{C}, r \in c \quad (104)$$

Cluster Busy Times Constraint

Applies to: Resources

Point-of-application: Resource

We use parameters \underline{B}_c and \overline{B}_c to denote the minimum and maximum values specified in constraint $c \in \overline{C}$.

$$U_{\underline{B}_c, \overline{B}_c} \sum_{tg \in c} p_{r,tg} \leq s_{c,r}^{\text{clusterbusy}} \quad \forall c \in \overline{C}, r \in c \quad (105)$$

Limit Busy Times Constraint

Applies to: Resources

Point-of-application: Resource

We use parameters \underline{B}_c and \overline{B}_c to denote the minimum and maximum values specified in constraint $c \in \overline{C}$.

$$-|tg| \cdot (1 - p_{r,tg}) + U_{\underline{B}_c, \overline{B}_c} \sum_{t \in tg} q_{r,t} \leq s_{c,r,tg}^{\text{limitbusy}} \quad \forall c \in \overline{C}, r \in c, tg \in c \quad (106)$$

Limit Workload Constraint

Applies to: Resources

Point-of-application: Resource

The workload of a solution resource is given by $w_{e,se,er} = \frac{D_{se} \cdot L_{er}}{D_e}$ where L_{er} is an integer denoting the workload of event resource er . We use parameters \underline{B}_c and \overline{B}_c to denote the minimum and maximum values specified in constraint $c \in \overline{C}$.

$$U_{\underline{B}_c, \overline{B}_c} \sum_{e \in c, t \in T \setminus \{t_D\}, se \in e, er \in e} w_{e,se,er} \cdot x_{se,t,er,r} \leq s_{c,r}^{\text{limitworkload}} \quad \forall c \in \overline{C}, r \in c \quad (107)$$

Balance Class Size Constraint

Applies to: Event Groups

Point-of-application: Event Group

We use the parameter \overline{B}_c to denote the maximum class size difference specified in constraint $c \in \overline{C}$. The role parameter is optional and denotes that only resources of a specific role in the event should be considered. We use the variable $mr_{c,eg}$ to denote the number of resources allocated to the specified Event Group

$$\sum_{\substack{e \in eg, er \in e, \\ r \in er, type_r = type_c \setminus \{r_D\}}} c_{eg,r} = mr_{c,eg} \quad \forall c \in \overline{C}, eg \in \overline{C} \quad (108)$$

$$mr_{c,eg} - mr_{c,eg2} - \overline{B}_c \leq s_{c,eg}^{\text{balancesize}} \quad \forall c \in \overline{C}, eg \in c, eg2 \in c, eg \neq eg2 \quad (109)$$

$$mr_{c,eg2} - mr_{c,eg} - \overline{B}_c \leq s_{c,eg}^{\text{balancesize}} \quad \forall c \in \overline{C}, eg \in c, eg2 \in c, eg \neq eg2 \quad (110)$$

Student Choice Constraint

Applies to: Resources

Point-of-application: Resource

We use parameters \underline{B}_c and \overline{B}_c to denote the minimum and maximum values specified in constraint $c \in \overline{C}$.

$$U_{\underline{B}_c, \overline{B}_c} \sum_{eg \in c, e \in eg, er \in e, r \in er, r=r_c} c_{eg,r} \leq s_{c,r}^{\text{studentchoice}} \quad \forall c \in \overline{C}, r_c \in c \quad (111)$$

Further Additions for guaranteeing exact objective values

Using the model as described above will result in valid/optimal schedules if the ILP solver is run without a time limit. If there is a time limit the objective value provided by the ILP solver might be bigger than the actual objective value of the produced schedule. This is due to the formulation using the \leq operator in combination with the deviation variables. So while it does lead to a worse objective value the ILP solver can save computational time by setting the deviation variables higher than necessary. To mitigate this issue for the constraints that feature an upper and lower bound we can add binary indicators that tell the ILP solver which constraint to use based on the current assignments (e.g. if the constraints have the form $U_{\underline{B}_c, \overline{B}_c} x \leq s$) we set a binary variable $bin_{max} = 1$ if $x > \overline{B}_c$ and a binary variable $bin_{min} = 1$ if $x < \underline{B}_c$. We then use those indicators to add the following constraints:

$$x - \overline{B}_c = s \quad \text{if } bin_{max} \quad (112)$$

$$\underline{B}_c - x = s \quad \text{if } bin_{min} \quad (113)$$

We proceed similarly for constraints of the form $x_1 \leq s, x_2 \leq s, \dots, x_n \leq s$ (where x_i may represent an arbitrary linear expression). We can instead model them with the following constraint:

$$\max(x_1, \dots, x_n) = s \quad (114)$$

Note that while we use so-called general constraints (indicator constraints, max constraints etc.) the ILP solver automatically transforms those into a set of linear constraints.

References

1. Fonseca, G.H.G., Santos, H.G., Carrano, E.G.: Integrating matheuristics and metaheuristics for timetabling. *Comput. Oper. Res.* **74**, 108–117 (2016). <https://doi.org/10.1016/J.COR.2016.04.016>, <https://doi.org/10.1016/j.cor.2016.04.016>
2. Fonseca, G.H.G., Santos, H.G., Carrano, E.G., Stidsen, T.R.: Integer programming techniques for educational timetabling. *Eur. J. Oper. Res.* **262**(1), 28–39 (2017). <https://doi.org/10.1016/J.EJOR.2017.03.020>, <https://doi.org/10.1016/j.ejor.2017.03.020>
3. da Fonseca, G.H.G., Santos, H.G., Toffolo, T.Â.M., Brito, S.S., Souza, M.J.F.: GOAL solver: a hybrid local search based solver for high school timetabling. *Ann. Oper. Res.* **239**(1), 77–97 (2016). <https://doi.org/10.1007/S10479-014-1685-4>, <https://doi.org/10.1007/s10479-014-1685-4>
4. Fonseca, G., Santos, H.G., Carrano, E.G., Stidsen, T.: Modelling and solving university course timetabling problems through xhstt. In: PATAT. vol. 16, pp. 127–138 (2016)
5. Kristiansen, S., Sørensen, M., Stidsen, T.R.: Integer programming for the generalized high school timetabling problem. *Journal of Scheduling* **18**, 377–392 (2015)
6. Lewis, R., Paechter, B., McCollum, B., et al.: Post enrolment based course timetabling: A description of the problem model used for track two of the second international timetabling competition. Cardiff Business School Cardiff, Wales (2007)

7. Mansour, N., El-Jazzar, H.: Curriculum based course timetabling. In: Wang, H., Yuen, S.Y., Wang, L., Shao, L., Wang, X. (eds.) Ninth International Conference on Natural Computation, ICNC 2013, Shenyang, China, July 23-25, 2013. pp. 787–792. IEEE (2013). <https://doi.org/10.1109/ICNC.2013.6818082>, <https://doi.org/10.1109/ICNC.2013.6818082>
8. McCollum, B., Schaerf, A., Paechter, B., McMullan, P., Lewis, R., Parkes, A.J., Gaspero, L.D., Qu, R., Burke, E.K.: Setting the research agenda in automated timetabling: The second international timetabling competition. *INFORMS J. Comput.* **22**(1), 120–130 (2010). <https://doi.org/10.1287/IJOC.1090.0320>, <https://doi.org/10.1287/ijoc.1090.0320>
9. Müller, T., Rudová, H., Müllerová, Z., et al.: University course timetabling and international timetabling competition 2019. In: Proceedings of the 12th International Conference on the Practice and Theory of Automated Timetabling (PATAT-2018). vol. 1, pp. 5–31 (2018)
10. Post, G., Gaspero, L.D., Kingston, J.H., McCollum, B., Schaerf, A.: The third international timetabling competition. *Ann. Oper. Res.* **239**(1), 69–75 (2016). <https://doi.org/10.1007/S10479-013-1340-5>, <https://doi.org/10.1007/s10479-013-1340-5>
11. Post, G., Kingston, J.H., Ahmadi, S., Daskalaki, S., Gogos, C., Kyngäs, J., Nurmi, C., Musliu, N., Pillay, N., Santos, H.G., Schaerf, A.: XHSTT: an XML archive for high school timetabling problems in different countries. *Ann. Oper. Res.* **218**(1), 295–301 (2014). <https://doi.org/10.1007/S10479-011-1012-2>, <https://doi.org/10.1007/s10479-011-1012-2>
12. Ruiz-Torrubiano, R., Knopp, S., Wolf, L.M., Krystallidis, A.: A scheduling perspective on modular educational systems in europe (2024), <https://arxiv.org/abs/2403.05549>