

The Flexible Scheduling Paradigm: The Prototype School

<https://doi.org/10.3991/ijet.v14i05.9683>

Yaakov Snyder^(✉)
Baltimore, Maryland, USA
ysnyder@gmail.com

Yale T. Herer
Israel Institute of Technology, Haifa, Israel

Michael Moore
Israel Institute of Technology, Haifa, Israel

Avishai Catane
Bet Shemesh, Israel

Richard M. Novak
La Jolla, California, USA

Abstract—The flexible scheduling paradigm (FSP) improves student learning by dynamically redeploying teachers and other pedagogical resources to provide students with customized learning conditions over shorter time periods called ‘mini-terms’ instead of semesters or years. By conceptualizing the school curriculum as a physical map, we customize the routing of students through curriculum using a core curriculum-targeted mastery-based approach. FSP increases deployed teacher effectiveness by making customized mentoring part of teacher’s regular schedules and by deploying teachers to their strengths. We establish a *prima facie* case for FSP by building comparative simulations of various schools as they are currently run (the Present Schools) and the same schools as they would be run with FSP (the Schools of the Future). Statistical results of the simulations confirmed that using FSP can increase key educational metrics including graduation rates, final course grades, mean grades in core curriculum, average teacher effectiveness, and the quality of teacher deployed expertise.

Keywords—Education reform, educational technology, mastery learning, flexible scheduling paradigm, operations research, simulation

1 Introduction

The Flexible Scheduling Paradigm (FSP) is a modern approach for organizing schools, which dynamically redeploys teachers and other pedagogical resources to provide students with customized learning conditions over shorter time periods called ‘mini-terms’ instead of semesters or years. FSP conceptualizes school curriculum as a physical map, routes students through curriculum using a core curriculum-targeted

mastery-based approach, and increases teacher effectiveness through customized mentoring and teacher deployment [22], [23], [24], [25]. To compare FSP with current school organizational practices, we built comparative simulations of various schools as they are currently run (the Present Schools) and the same schools as they would be run with FSP (the Schools of the Future). We also validated our models through extensive interviews and written feedback from Baltimore City, Maryland (USA) public school teachers, administrators and executives.

Under FSP, each round of unit-based teaching will be implemented in mini-terms, whose length can be any reasonable period of time. FSP exploits natural dependencies that exist in curriculum by constructing curricular maps and tracking student assessment at the level of granularity of a curricular unit. Students still complete courses, but typically over compressed or extended periods of time rather than semesters or years. FSP can be integrated with alternative/modern types of learning such as professional tutoring, peer-to-peer tutoring, small group learning, video-based or assisted learning, independent learning, e-learning, and blended or virtual learning environments [22]. FSP can also potentially end the pernicious phenomenon of social promotion [28] by regularly customizing school schedules directly to student needs in place of the current typical school system of organization, which encourages social promotion by forcing students to remain in rigid course and grade level structures.

FSP uses modern operations research techniques [8], [18] to enhance teacher effectiveness [1], [4], [6], [19] through targeted training and the prioritized deployment of teachers. See Snyder, Herer & Moore [24] for details on how FSP integrates elements of educational paradigms: mastery learning [11], flexible modular scheduling [27], block scheduling [3] and curriculum mapping [10]. Our research also benefits theoretically from Hutchins [9]. Hutchins, using navigation as an example, made a brilliant general case for using an entire system as the unit of analysis rather than any one of its component parts. In education, for too long, the ‘classroom’ has been the dominant unit of analysis of educational research but it is just one component of the larger school or even units of analysis broader than a school.

2 Research Goals and Methodology

Our primary research goal was to make a *prima facie* case that the FSP model provides superior educational results to today’s typical model of school organization. To achieve this goal, we built a simulation model for schools as are typically run today using a yearlong system of organization, called the ‘Present School’. We then built a model for the same schools reorganized using FSP’s unit-based, mini-term scheduling called the ‘School of the Future’ (SF). We identified parameters to compare the performance of PS and SF, such as graduation rates. We checked the validity of both PS and SF simulations by interviewing public school teachers, administrators, and executives. We note our case is only ‘*prima-facie*’ because we have not yet tested the FSP model in a real school. Indeed we feel that simulation testing is a necessary precursor for real-world testing, the next natural step. As a result we only tested the simulated quality of learning and instruction.

The primary tool that we use is simulation. Herer [7] describes the modeling process in five steps:

- See the real world (see what is important)
- Model the real world (include what you want)
- Solve the model (the model, not the real world – optimally or heuristically)
- Take the model solution to the real world (Does it make sense? Even if the solution is optimal for the model, is it a heuristic for the real world?)
- Implement the solution in the real world

In this initial work, we implement steps 1 – 4. Simulation methodology has the major advantage of being able to vary key parameters at will. Sometimes one may wish to simulate a variety of entities in the real world such as school size or curricula. Other times one may wish to simulate a range of values for a parameter that is not yet known with accuracy in the real world. An example of such a parameter would be the amount of variation within individual teachers in their capabilities to teach particular units. Such ‘sensitivity analysis’ is an extremely valuable tool in simulation. In addition, though we did our best in creating our model, simulation work is extremely flexible and we could vary or add any major factor to our models in future work.

We chose Microsoft Visual Studio, SQL Server database, and Excel to build our simulations and store our data. We conducted statistical analyses of results using Microsoft Excel and SPSS. We found these platforms more advantageous for our needs than simulation platforms such as ARENA [5] due to their speed and flexibility.

2.1 Definition of sigma

Bloom [2] used the term ‘sigma’ to refer to the impact of a particular factor (in units of standard deviation) upon the performance of students as measured by standardized tests. So for a teaching approach to achieve a one sigma improvement, it must produce performance results in students that are one standard deviation higher than the mean. Henry et al. [6] used a similar sigma-based approach to measure the changing effectiveness of teachers over their first five years, and Walberg [29] also measured statistically a variety of inputs into educational success into this sigma-based approach. We incorporated this general approach in our model, adapting it to build formulae to forecast individual students’ grades.

2.2 Definitions of key parameters

The following are key parameters used in the simulations:

- **SSC**–Student subject competency represents a combination of how talented and motivated a student is in a given subject or more formally: the impact of students’ own capacity in sigma on their learning results in a given subject.
- **TUE**–Teacher unit effectiveness measures how well a teacher can teach a particular unit of material in terms of the sigma impact on students’ learning results. This concept is closely related to the teacher effectiveness of a particular teacher but

measured at the unit level rather than overall. See Goldhaber & Anthony [4] for one example of the vast literature on teacher effectiveness (TE), which we are specifying with even greater detail at the unit level.

- **PTUE**—Perceived teacher unit effectiveness. A modified value where “noise” (*PTUEM*, see below) is added to *TUE* to model a school’s limitation to accurately measure a teacher’s capability to teach a unit. *PTUE* reflects how the school perceives a teacher’s effectiveness, rather than the real *TUE* value.
- **PTUEM**—Perceived teacher unit effectiveness modifier. The “noise” that is added to *TUE* values to produce *PTUE* values. *PTUEM* reflects the inevitable inaccuracy of an attempt to measure *TUE*.
- **RTUE**—Realized teacher unit effectiveness measures how well teachers are deployed to teach units in terms of the sigma impact on students’ learning results. A teacher has to actually teach a unit for their *TUE* value in that unit to count as *RTUE*. We measure *RTUE* at the student level, so a teacher with a high *TUE* in a unit with 30 students has more overall impact than one with 20.
- **LG**—Last grade. A student’s most recent grade in a unit on a percentage scale. This value is relevant to forecasting a student’s grade when a unit is repeated.
- **MP**—Mastery of prerequisites. A student’s grade average across all prerequisite units, on a percentage scale; relevant to forecasting a student’s grade in a unit. See Shapiro [20] for a seminal paper on this domain.

2.3 Model description

When building a model one must choose parts of the ‘real world’ to model [12] and ignore certain aspects of the system being modeled. Since the goal of our study is to compare the PS and SF models, we simplify, wherever possible, in such a way as to minimally impact the comparative analysis.

Simulated schools: We simulated high schools with a six-period day, all academic periods. Course grades are determined by averaging the final grades in units. Midterm and final exams are not simulated. Students have up to six years to complete graduation requirements. A ‘drop-out’ is a student who has had a full six years to graduate and failed to do so. We did not simulate students who drop-out in the middle of their high school careers. SF prioritizes scheduling students in unit classes with teachers with higher *TUE* values, which PS does not. SF implements in classroom teacher training as part of the scheduling, which PS lacks.

Simulated curriculum: Both PS and SF simulations had the same 21 course graduation requirement, based upon a reasonable midpoint of American high schools [15]. Courses were designated as mandatory or elective to match graduation requirements. Courses in all subjects are split into eight equal units. Admittedly, this may be a difficult task to achieve in some subject areas, and more complex scheduling approaches may be adopted to deal with less uniform curricular structures.

We represent real curriculum in math, history, and chemistry. These curricular units were designed by professional teachers [23]. For other subjects, ‘simulated curricula’ without specific content were created by designing prerequisite structures which ‘look reasonable’ for those subjects. For example, English curricular structures were set up

to be less linear than history but more than math. We do not include curriculum for special education or gifted students.

We use the same curricular units for both PS and SF models, though PS uses course-based dependencies and SF uses unit dependencies and unit designations as core, mandatory, or elective. Core units are also mandatory and are prioritized higher in SF. The PS approach has no way of scheduling based upon unit importance, since it is scheduling a course at a time. Elective units may be pruned from a course in SF; students can pass a course if they pass all core and mandatory units and have a course average above the needed passing grade. Students in PS courses will always be exposed to all units in a course, including elective units.

Simulated teachers: Teachers are mirrored across PS and SF so every new teacher in PS starts with a mirror teacher in SF teaching the same subject with the same starting *TUE* values. These parallel teachers retire at the same time across PS and SF. In this way we improve the accuracy of our comparison of PS and SF. We simulate teachers gaining competence at teaching each year by implementing functions that change their set of *TUE* values, depending on the number of times they teach a unit, if they are mentored in a unit, a gain from acquiring general teaching experience, and a noise value. As teachers are scheduled differently, *TUE* values of parallel teachers across PS and SF will naturally tend to diverge over time.

Teachers teach only one subject but can teach all of its units. Teachers teach a random number of years within a total maximum duration. We represent teacher effectiveness at the unit, rather than the subject or general level. At the start of a simulation and each time a new teacher enters a simulation, the teacher receives a random distribution of *TUE* values in the units within their subject. We first calculate a randomly generated Teacher Effectiveness (*TE*) value to represent the overall effectiveness of the teacher. We then calculate individual *TUE* values out of this *TE* value using a second randomly generated Teacher Unit Effectiveness Modifier (*TUEM*) value. The purpose of this two-step process is to enable us to control variation in *TUE* values across and within teachers separately by having independent control of the means and standard deviations of the *TE* and *TUEM* curves to test a wide variety of possibilities that may exist in the real world (this type of data is not definitive from the literature).

Since no school can ever have a perfect system for assessing *TUE* values, we implement the middle layer of *PTUE* between *TUE* values and the scheduler. Every year teachers have a *PTUE* value calculated for every one of their *TUE* values using a randomly generated noise factor (*PTUEM*). We use the *PTUE* values for SF scheduling decisions (since those are the values schools would have) and *TUE* values for PS and SF grading forecasts (since *TUE* represents the real effectiveness of a teacher in a unit). We vary the degree of noise in different simulation experiments to represent better or worse teacher assessment systems.

Simulated students: We represent student academic competence at the subject level using one *SSC* value per subject for each student. We generate these values for the student population using a set mean and standard deviation plus a noise factor, each of which may be varied across simulation experiments to represent different types of student populations.

Every student in PS has a mirror student in SF with the same *SSC* values. These remain fixed throughout a simulation run, though in the real world it is possible for students to improve their underlying competence in a given subject. We permit student ages in classes to vary freely in both PS and SF simulations, so students from different grades may be combined.

2.4 Flow of a simulation

For both PS and SF, we begin by initializing simulation parameters and simulate the system from year to year with the end of one year feeding into the next, until the total number of desired years is completed. See Figure 1 for one year of a PS or SF simulation. At the beginning of each simulated year we first gather information on teacher and students from the previous year, add the students who are entering grade 9, and similarly we add new teachers. With this information, the school schedules the classes and students learn and are evaluated. At the end of the year, teacher and student information is updated. Teachers gain general experience and retiring teachers are removed from the simulation. Each student graduates, drops out, or is passed on to the next year. Figures 2 and 3 expand on the scheduling and grading sub processes in Figure 1. See Figure 2 for PS scheduling and grading for one year and Figure 3 for SF scheduling and grading for one year. In general these processes take student and teacher data and schedule the school as it would be scheduled in real-life, i.e. the focus is on what a student can learn and what a student needs to learn to continue with their education and graduate. At the end of instruction grades are determined based on the students' and teachers' abilities and are updated based on what they taught. Primary differences between the processes represented in Figure 2 and Figure 3 include that SF scheduling is carried out eight times a year versus once per year for PS; students are scheduled based on units for SF and courses for PS; and SF implements the FSP training model. These flowcharts are meant to give the reader a general sense of the overall flow rather than specifics of the process.

2.5 Scheduling description

We now describe the major factors involved in creating a PS and SF schedule. PS and SF both use a student 'need list'. For PS this consists of all courses a student is ready to take; for SF this consists of all units a student is ready to take. We used a similar algorithm for PS and SF to prioritize students reaching graduation to assure that the scheduling algorithm does not cause a difference in dropout rates between PS and SF. Each course (PS) or unit (SF) a student is eligible to take is given a value measuring its 'slack', or how many years (PS) or mini-terms (SF) the course or unit could be delayed and still allow a student graduate on time. This slack calculation is based on course or unit prerequisites defined in the curricular model. Using this information, the algorithm generates a prioritized course list for PS and a prioritized unit list for SF. SF prioritizes the assignment of teachers to units by highest *PTUE* and schedules teachers with gaps in their schedules to be mentored by higher *PTUE* teachers in scheduled units.

PS randomly assigns qualified teachers to courses. Both PS and SF algorithms use the above data to do their best to ensure students graduate on time.

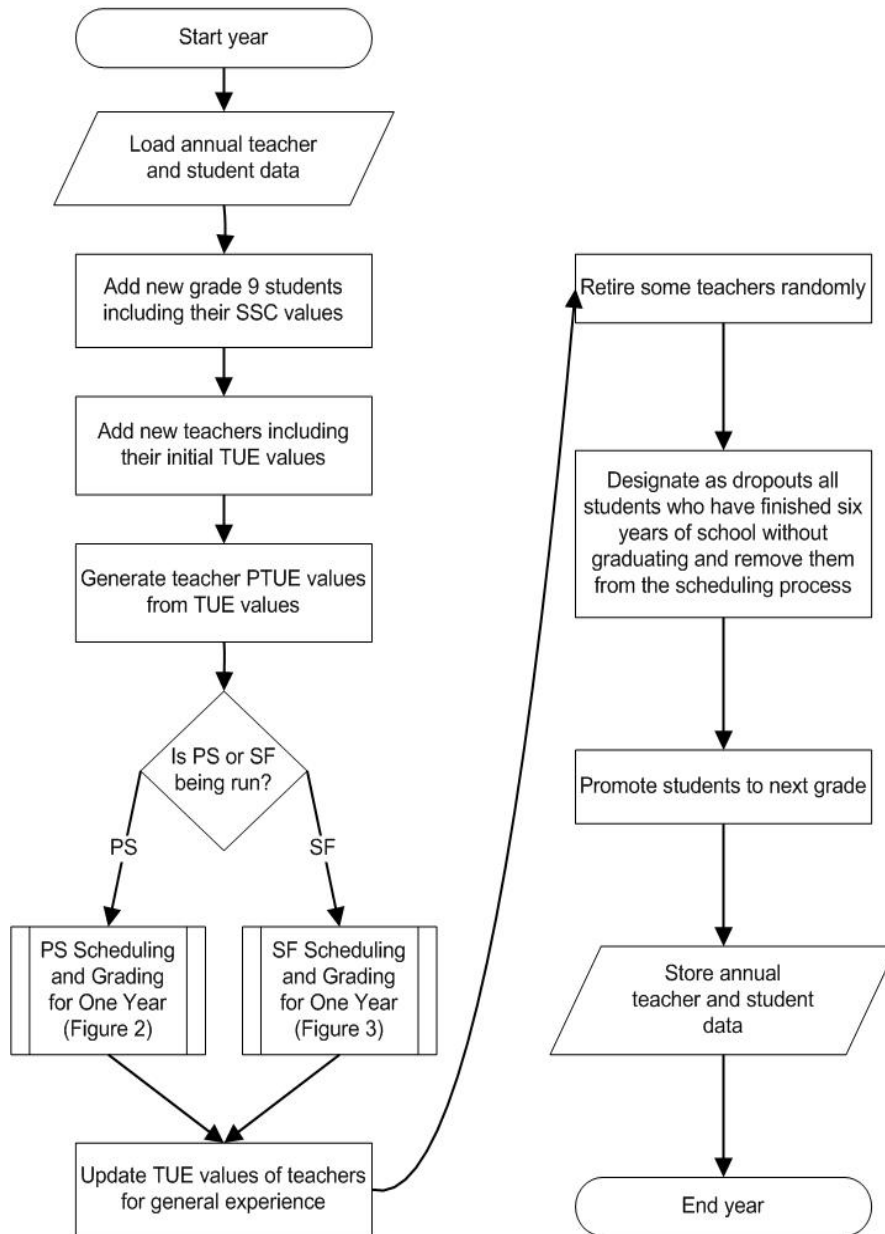


Fig. 1. One Year of a Simulation (PS or SF)

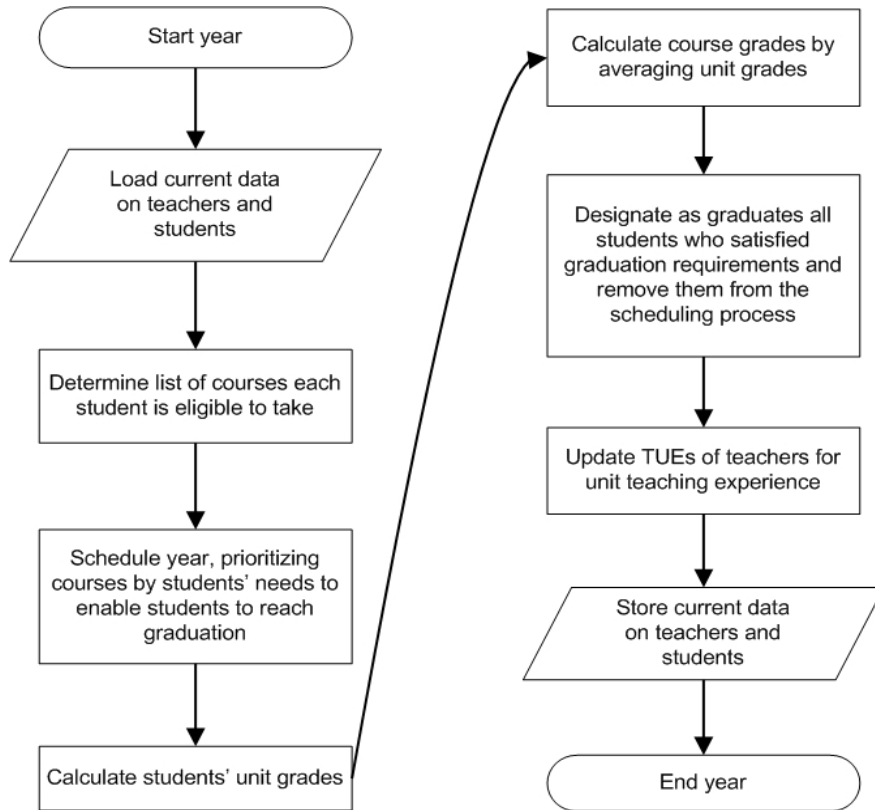


Fig. 2. PS Scheduling and Grading for One Year

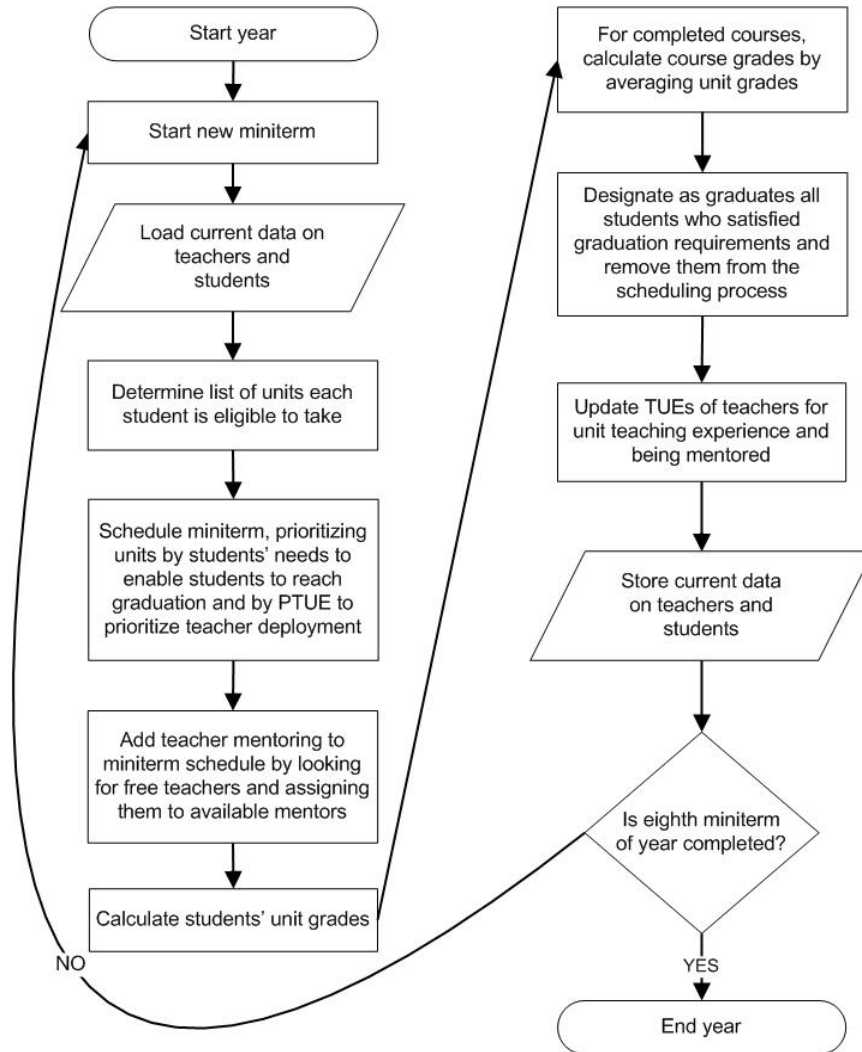


Fig. 3. SF Scheduling and Grading for One Year

2.6 Predictive models

Central to our PS and SF simulations are predictive models for how students perform in units and how *TUE* changes over time. For both PS and SF models, grades are assigned at the unit level and course grades are determined by averaging all unit grades in a course.

Predicting student grades: Since a primary goal is a valid PS and SF comparison we predict grades using the same method in both schools. The grading scale of 0-100

is used throughout. Our grade prediction model begins with a value for a nominal grade (NG) and the value of one standard deviation (SD). The nominal grade is a ‘starting grade’ for grade calculations and not an average grade, since many other factors affect a student’s grade.

In our simulation a student’s grade in a unit is forecast using the following four factors:

- Teacher Unit Effectiveness (TUE): the quality of the teacher to teach a particular unit
- Student Subject Competency (SSC): how skilled the student is in a particular subject
- Mastery of Prerequisites (MP): how well the student knows the prerequisites if applicable
- Last Grade (LG): how well the student did the last time he took the unit if applicable

It is critical to note that simulations such as those we have built may readily be expanded to include many other factors to student learning.

The number of standard deviations a forecasted unit grade is above or below the nominal grade is determined by first combining these four factors into a single sigma value. TUE and SSC are already measured in terms of sigma. In contrast, MP and LG are measured in grade percentage and so must be converted to a measure in terms of sigma.

MP is converted into a sigma value by subtracting NG and dividing the result by SD . If a student’s mastery of prerequisites (MP) is less than the nominal grade (NG), the impact upon the grade is negative and vice versa.

LG is converted into a sigma value by dividing it by the nominal grade (NG). This value will only be positive. The idea here is to calculate a sigma proportional to the nominal grade, but which is only positive, because prior experience learning a unit should only produce an advantage. We acknowledge that it is theoretically possible for a very poor learning experience to produce a net negative effect on a student’s future learning in a unit. Such a rare instance could be a result of a teacher poorly teaching a student with such an inaccurate conceptual model that a future teacher would have to ascertain the deficit and remediate the student. We consider this to be a relatively rare anomaly and therefore do not account for it in our model.

Before the four factors are combined, TUE and SSC are weighted more heavily than MP and LG due to their relatively greater impact. We then combine the TUE , SSC , MP and LG sigma factors into a single combined sigma value σ_C . This combined effect determines how many standard deviations above or below the nominal grade a student is expected to achieve. To combine these values, we use a formula of our own design which keeps the combined effect of the factors within a certain maximum range (typically between ± 2.5 sigma). The formula first combines positive factors and negative factors separately in a sub additive manner. Making the combination additive (without interaction) would produce unreasonably high or low resultant σ_C values. For example, it seems unlikely that a teacher of $TUE =$ one sigma combined with a student of $SSC =$ one sigma would raise a student’s grade two sigma above the nominal grade.

The two resultant positive and negative sigmas are considered independent of each other, and so we sum them, creating a cancellation effect. Finally, we calculate the predicted grade using the following formula:

$$\text{Unit Grade} = NG + (\sigma_c * SD) + \text{noise} \quad (1)$$

This calculation uses the nominal grade as a starting point, adds in the effects of the various factors (i.e., the combined effects (σ_c) multiplied by the standard deviation), and then adds in a Gaussian noise term. The noise term is meant to simulate the fact that in the real world two students taking a unit with the exact same factors would likely achieve different grades in a unit due to un-simulated factors and unpredictable local conditions.

We set the parameters NG , SD , and the maximum combined effect of the factors appropriately in a simulation and verify that the results across students generate a reasonable bell curve. For example, setting $NG = 75$, $SD = 10$, and the maximum combined effect of the factors to ± 2.5 sigma would yield a likely potential range of values for $NG + (\sigma_c * SD)$ of 50-100. The noise term may raise this range a bit further.

The real interactions between our four chosen grading factors (and many others, such as class size) are likely to be highly complex, so we encourage continued development of grade forecast modeling and studies gathering the underlying relevant data in the field. The particular factors affecting grades and the mathematical model for combining these factors may be readily adjusted in future work as better data emerges, as is the power of simulation.

Modeling change in teacher unit effectiveness: We model the following critical factors for TUE change for both PS and SF: general teaching experience and how often a teacher has taught the unit in question. Additionally for SF, when a teacher is mentored in a unit, we use the difference in TUE values between mentor and mentee in order to determine the improvement of the mentored teacher.

Our general goal is a curve which approximates the TE findings from *Science* [6]. This research shows that larger increases in TE occur in the earlier years of teaching. To do this, we set a maximum career improvement in TE for a teacher, and use the factors described above to calculate a TUE improvement as a percentage of the remaining improvement possible. In this way we model the idea that a new teacher just starting a career has the most to gain in teacher effectiveness. Although the *Science* article only covers the first five years of teaching, our simulations cover entire careers. We do not model the subject-specific TE increases as described in *Science*, but rather treat all subjects equally.

3 Design of Experiment

In this section we describe our experimental design. We begin by describing some aspects of our experimental design that is related to the fact that our experiments are based on simulations.

3.1 Replications

Due to the fact that no two student or teacher bodies are the same (even if they are drawn from the same distribution) and simulations involve the pervasive influence of preceding events upon subsequent events, it is necessary for statistical purposes to run multiple replications of each simulation and combine the results to get a statistically robust sample. We therefore run multiple pairs of PS / SF simulations.

The number of replications needed [21] is determined with an appropriate statistical calculator. We found that five replications of each PS / SF pair were always sufficient from a statistical viewpoint. To be conservative, we used ten replications each for all of our simulation runs. When we report results below, we report the average across all replications in PS or SF. For example, when we refer to the mean unit grade in a PS simulation, we are describing the overall mean from all the replications, each using a different pseudorandom number stream for input.

3.2 Parallelism

In order to maximize the statistical accuracy of comparing PS and SF simulations, we employed parallelism in our simulations. This parallelism has the technical name ‘common random number’ and is an accepted variance reduction technique [12]. We use the Mersenne Twister approach to generate the pseudorandom numbers [13], [26].

Within each pair of PS / SF simulations, the same stream of pseudorandom numbers are duplicated for PS and SF for each of a variety of factors that vary in the real world, such as initial *SSC*, *TE* and *TUE* values, and noise in the grading or *TUE* change functions. This parallelism allows our comparisons to focus on the effect of the PS versus SF models. For example, a paired set of identical randomly generated students is run through PS and SF with the exact same *SSC* values. We can think of these paired simulations as the same group of students learning in a PS versus an SF organized school. Similarly, we use a paired set of randomly generated teachers having identical initial *TE* and *TUE* values and the same teacher retirement year, though these values will diverge over time due to differences in the PS and SF models. We can think of these paired sets of teachers as being the same person teaching in a PS versus an SF organized school.

Across each of the ten pairs of PS / SF simulation runs, a different stream of pseudorandom numbers is employed for all of the above described factors to simulate 10 different pairs of schools, each with a parallel PS and SF version. We average the results across the 10 pairs for our various educational metrics.

3.3 Ramp up and ramp down years

We run each simulation for 55 school years. Since our simulations start with an empty school with all new teachers, we need to simulate our schools for several years before starting to examine the behavior of the school. Since teachers in our simulations work between 1 and 20 years before they retire and students graduate after a maximum of six years we decided to start collecting data for analysis after an initial ramp up time

of 20 years. Similarly, we strip the last five student classes out of the data, since these cohorts do not have a full six years to graduate or drop out per our simulation model. Given these procedures, we end up with 30 entering classes of students (from grade 9 to graduation or dropout) of usable data per simulation.

3.4 Educational metrics

The most common method to compare two systems is to choose meaningful metrics and our PS / SF comparison is no exception. We used the following educational metrics:

- Mean Final Course Grade
- Mean Unit Grade and Percentage of Units Passed for:
 - All Units
 - Core Units
 - Mandatory Units
 - Elective Units
- Graduation Rate
- Mean *TUE* and Mean *RTUE*
- Mean Number of Units per Course, Mean Number of Units Passed, Mean Number of Units per Student.

All of the above were measured by student class except for Mean *TUE* and Mean *RTUE*, which were measured year to year. A ‘student class’ is the entire set of students that begin together in grade nine. Both PS and SF simulations give students up to six years to graduate. However they can, and normally will, graduate earlier. Moreover, SF students may graduate after any mini-term whereas PS students can only graduate at the end of a year.

For Mean Final Course Grade, we used the final completed course grades for PS and SF. In PS, if a course is failed the entire course is taken again and a new course grade is assigned. We use the last course grade assigned for each student. For SF, courses are typically not ‘failed’. Instead, students simply retake failed units within courses. We therefore use the set of all the last unit grades a student received to determine the student’s course grade.

All Mean Unit Grade measures include failed or repeated units. Measures of Mandatory Units do not include core units, which are also technically mandatory (but measured separately).

To compute Mean *TUE*, each year all *TUE* values of individual teachers are first averaged into a single mean *TUE* value for each teacher and all the teachers’ means are averaged into an overall mean *TUE*. To compute Mean *RTUE*, the *RTUE* values students encounter with teachers (one value per student) are averaged into a mean each year. Recall that *TUE* values measure the competency of the teaching staff while Mean *RTUE* values measure how effectively teachers are deployed.

Mean Number of Units per Course measures how many units students take of the eight possible in each course. Whereas in PS, students always take eight units per course, SF will produce a lower mean number as SF can prune elective units. Mean

Number of Units Passed measures how many unique units students pass over their school career, whether they graduate or dropout. Mean Number of Units per Student measures how many units students take over their school career, whether they graduate or dropout. Units that are taken multiple times are counted every time they are taken. For example, if a PS student has to repeat a course, each of the eight units he repeats for the course would be counted.

3.5 Prototype school

Because there are a very large number of possible schools we can simulate, we chose a prototypical school to evaluate for our present analysis. We call this the ‘Prototype School.’ Our Prototype School is designed to be fairly close in its features to an average U.S. public school. For the Prototype School comparison, we measure the educational metrics described in Section 3.4. Since we use parallelism across PS and SF in order to simulate the same teachers and students across each pair, the proper test to use for a comparison of the Prototype School for PS and SF is the one-tailed, paired *t*-test. We also measure the relative frequency of student teacher interaction in response to a suggestion from a validation interviewee.

Prototype school settings: The prototype school size is set to 220 students per grade, or about 880 students in grades 9-12 plus some residual students in the two extra years allowed before dropout occurs. The average American public high school in 2010-11 had about 846 students [14].

We use a total of 60 teachers distributed across subjects in rough proportion to the number of courses in each subject. The average ratio of students to teachers in U.S. public schools in 2007-8 was 16.4 [15]. Since there are 880 students in grades 9-12, when a simulated school has 104 (of the 440 maximum possible) students continuing in their 5th or 6th year, the school would match the US national average student teacher ratio. The number of physical classrooms is set to 50. Teachers are permitted to teach up to 5 of 6 periods. They may receive mentoring in the 6th period, but in practice this is extremely rare.

We set the permitted class size range for both PS and SF to 15-35 students, with the target set at 25, or about the average in American public schools today. For grade prediction purposes $NG = 75\%$; $SD = 10\%$, and the relative weight of *MP* and *LG* are set to half of *TUE* and *SSC*. This means, e.g., that a student’s mastery of prerequisites is half as important as their teacher’s effectiveness in terms of forecasting their grade. The passing grade for courses in PS and for all units in SF is set to 65%. Teacher to teacher mentoring is set up to occur in SF where pairings can be found between teachers of at least a 0.2 sigma difference in *PTUE*. Weights for the relative importance of the three factors used to update *TUE* (teaching a specific unit, being mentored in a specific unit, and general teaching experience) are set equally. The maximal possible increase in teaching effectiveness in a teacher’s career is set at 2.5 sigma.

The settings for key parameters for the Prototype School are determined randomly (using a truncated normal distribution) as shown in Table 1.

Table 1. Setting Parameters for Prototype School

	Mean (Mu)	Standard Deviation (Sigma)	Upper Limit	Lower Limit
SSC	0	0.5	2.5	-2.5
TE	-0.2	0.4	0.8	-1.2
TUE Modifier (TUEM)	0	0.3	0.8	-0.8
PTUE Modifier (PTUEM)	0	0.3	none	none

The *SSC* settings produce a curve that reflects a reasonable spread of student abilities. The maximal student *SSC* value of 2.5 takes a student from the nominal grade of 75% to a 100%, though that grade can be made lower by a poor teacher, other negative factors, and random noise.

The combination of *TE* (Teacher Effectiveness) and *TUEM* (Teacher Unit Effectiveness Modifier) ranges produces starting *TUE* values for teachers between -2 and 1.6 sigma.

3.6 Beyond the prototype school

For the purposes of this paper, we only report results from the Prototype School comparing PS and SF. However, we have also run a wide variety of other simulations which vary key parameters from the Prototype School and use ANOVA statistical analyses in place of *t*-tests. We varied School Size, *PTUEM* Sigma, *SSC* Mu, *TE* Mu, *TE* Sigma, and *TUEM* Sigma in these studies, keeping the Prototype School as the middle value to provide a useful comparison point [25]. For example, by varying School Size up and down from the Prototype School, we were able to simulate how effective FSP is relative to the number of students in a school. By varying *PTUE* Sigma, we simulated schools with superior and inferior teacher assessment systems. By varying *SSC* Mu, we simulated schools with stronger or weaker student populations. By varying *TE* Mu, we simulated schools with superior and inferior starting teaching staff capabilities. Since *TUEM* Sigma, the variation of *TUE* values within individual teachers, is currently unknown, we ran simulations varying *TUEM* Sigma higher and lower as well. These results will be reported in future publications.

4 Results

4.1 Prototype school, PS / SF comparison, *t*-tests

See Table 2 for our *t*-test results. In the table we present for each of the educational metrics the sample mean and standard deviation for PS and SF along with the *p*-value for a one tailed *t*-test.

Table 2. Significance of Comparing Prototype School: PS Versus SF

Educational Metric	PS Mean	PS σ	SF Mean	SF σ	<i>p</i> -value
Mean Final Course Grade	78.5	0.357	81.2	0.234	<0.001
Mean Unit Grade	76.0	0.406	79.9	0.308	<0.001
Mean Core Unit Grade	75.8	0.539	80.0	0.355	<0.001
Mean Mandatory Unit Grade	76.1	0.414	80.2	0.295	<0.001
Mean Elective Unit Grade	75.9	0.443	75.8	0.793	0.312
Percentage of Units Passed	83.3	0.012	93.9	0.009	<0.001
Percentage of Core Units Passed	85.2	0.015	94.3	0.012	<0.001
Percentage of Mandatory Units Passed	83.2	0.012	94.4	0.007	<0.001
Percentage of Elective Units Passed	81.8	0.014	86.9	0.020	<0.001
Graduation Rate	88.3	0.029	98.0	0.004	<0.001
Mean <i>TUE</i> (by Year)	-0.026	0.029	0.069	0.026	<0.001
Mean <i>RTUE</i> (by Year)	0.126	0.039	0.468	0.029	<0.001
Mean Number of Units per Course	8.00	0.00	7.27	0.008	<0.001
Mean Number of Units Passed	173	0.910	172	0.185	0.183
Mean Number of Units per Student	215	2.76	184	1.78	<0.001

4.2 Frequency of Teacher-Student assignments in classes together

We measured the frequency of teacher-student assignments in classes together in PS versus SF. PS had a much higher frequency of assignments. No teacher in the PS model ever has a student for fewer than eight units simply because every course in PS has eight units and teachers are always scheduled for an entire course at a time in the PS model. For PS simulations, the relative frequency of teacher-student assignments was 8 units (one course) for 85.7% of the time, 16 units (2 courses) for 12.7% of the time, and 24 or more units (3 or more courses) for the remaining 1.6% of the time. For SF simulations, the relative frequency of teacher-student assignments was 1 unit for 20.7% of the time, 2 units for 20.5% of the time, 3 units for 19.1% of the time, 4 units for 15.7% of the time and 5 or more units for the remaining 24% of the time, a substantial difference.

4.3 RTUE distribution graph

Because sometimes a picture is worth a thousand words, we include in Figure 4 a visual on the distribution of RTUE across PS and SF simulations. One can see that higher RTUE values are realized in SF on a percentage of units taught basis. For example, teachers with TUE values of -1 sigma teach PS students 1.8% and SF students 0.7% of the time. In contrast, teachers with TUE values of 1 sigma teach PS students 3.4% and SF students 4.8% of the time. One can see visually from this graph how SF students are exposed to superior teaching more often than PS students.

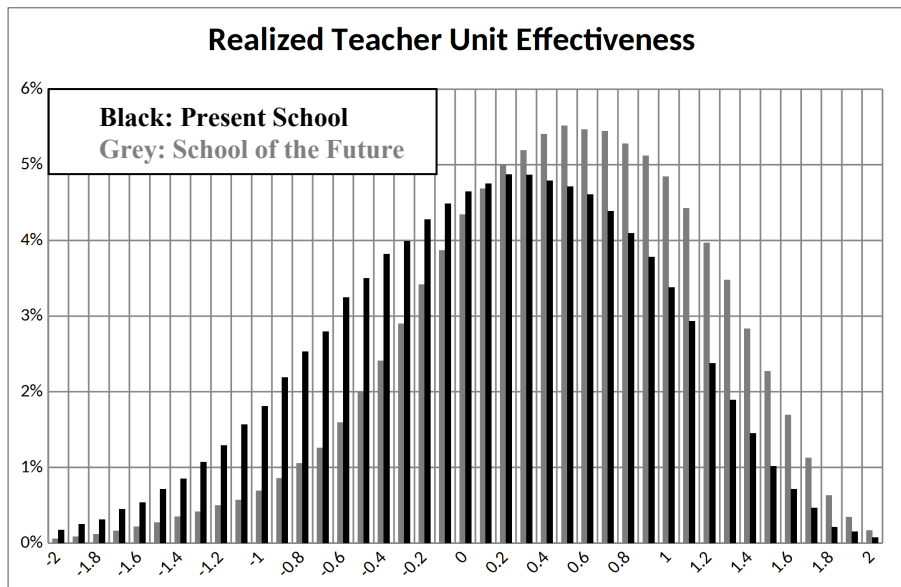


Fig. 4. RTUE Distribution across PS and SF

5 Discussion

We believe we have demonstrated that FSP can improve the simulated quality of instruction by enabling higher *RTUE* values and by customizing schedules for students on a unit by unit basis. We provided evidence that FSP can improve the simulated quality of learning by comparing PS and SF across a variety of metrics such as final course grade and graduation rates. In the sections below we highlight some key issues.

5.1 Unit and course grades

SF does substantially better than PS in core and mandatory curriculum and about the same in elective curriculum. The mean core unit grade in the SF Prototype School was 80.0% versus 75.8% for PS. Mandatory (but non-core) mean grades were superior for SF at 80.2% versus 76.1% for PS. In contrast, the mean unit elective grades were statistically indistinguishable at 75.9% for PS and 75.8% for SF. Indeed, this is a deliberate part of the design of FSP and makes sense since FSP under prioritizes elective units versus mandatory and core units. This superior performance in core curricula for students is a key result for FSP, as core curricula are widely recognized to underpin student mastery of subject areas. Mean final course grades were also demonstrably superior for FSP at 81.2% versus 78.5%.

5.2 Percentage of units passed

SF far outperforms PS in percentage of units passed for all categories of units, whether they be core, mandatory, and even in elective units, though as designed FSP is most striking in its improvement in core and mandatory unit areas. The dramatic gains for SF in all ‘percentage of units passed’ categories is due to its superior deployment of teaching talent as well as FSP’s insistence that students be ‘ready to learn’ a unit before they take it. It makes sense that the percentage of units passed for core and mandatory units is relatively superior to that of elective units in SF, since SF prioritizes core and mandatory over elective units. The overall mean percentage of units passed of 93.9% for SF versus 83.3% for PS is a key gain for FSP, and means that the time of both students and teachers is much more efficiently used both economically and pedagogically with FSP.

5.3 Graduation rates

The statistically significant outperformance of SF of 98.0% versus PS of 88.3% in graduation rate is striking. We feel this result is due to FSP’s innate flexibility. For example, a student can never fail a ‘grade level’ in FSP as occurs in so many of today’s schools. Similarly, the higher percentage of units passed results for SF eventually compound into a higher graduation rate as students are repeating material less often.

5.4 TUE and RTUE Improvements

SF statistically outperforms PS in both mean *TUE* and mean *RTUE* as expected due the combination of its training model and targeted teacher deployment. Central to FSP is the idea of realized teacher unit effectiveness. By deploying teachers to their stronger units, we found significant gains for FSP along a variety of key measures, such as unit grades, percentage of units passed, final course grades and graduation rates. These effects were strong and consistent across the variety of statistical tests we performed and also held up when *PTUE* was varied to simulate greater or lesser noise in the teacher assessment system although, as expected, SF does better when the teacher assessment system is more accurate. In addition, the FSP teacher training model was highly corroborated by our study. SF gained substantially in base TUE values and then also in RTUE values from teacher training. There are subtleties that could be added to the modeling of this process. For example, interviewees provided feedback that some degree of co-teaching is superior to simple observation as a training method. If the trainee is teaching at times, this may bring down the RTUE value of the unit somewhat. Yet, as we argued above regarding deploying teachers to moderately strong units, a supervised trainee would have much to gain for the general good, and the supervision would additionally likely produce a better result for students than if the trainee taught on their own.

5.5 Number of units learned

We see from the Mean Number of Units per Course (8 out of 8 units for PS vs. 7.27 out of 8 units for SF) that FSP is regularly pruning elective units from courses in order to prioritize other objectives for students. The Mean Number of Units Passed (173 for PS vs. 172 for SF) was not statistically different across PS and SF. The Mean Number of Units per Student (215 for PS vs. 184 for SF) indicates that SF does a far more efficient job than PS at teaching students content as this statistic includes repetitions of units. The PS model must repeat all units in failed courses whereas the SF model only needs to repeat failed mandatory or core units. This may make FSP more cost-effective.

5.6 Frequency of Teacher-Student assignments in classes together

We and our educator validators were both concerned [25] about SF's relatively poor frequency of teacher-student assignments in classes together. Note that this difference between PS and SF could be reduced or even potentially raised for SF above PS by specifically programming into the SF algorithm either:

- A 'stickiness' factor that keeps particular students and teachers together more often, possibly based upon past success of the student with the teacher, or the preference of a student for a particular teacher [22].
- A general 'matching' factor that tends to place teachers and students with similar teaching and learning styles, or general compatibility, together over time. This approach would require a categorization system of such styles and compatibilities to be implemented, but such categorizations are available, and could make the teaching yet more effective in FSP. The nuanced reality is that rather than simply having an average TUE value, each teacher has a unique TUE value relative to each individual student, depending on their relative compatibility as teacher and student. Were students assigned to teachers based on that compatibility as a real priority, students would not just learn better but would also find themselves with a more consistent set of teachers over time. The same types of customizations are also readily done in FSP for specific learning formats, including those without teachers, that are conducive to student learning [22].

6 Conclusion

We have demonstrated that such gains we have predicted for FSP are potentially quite feasible. Our goal was to make a convincing case for the prototyping, testing, and widespread distribution of FSP technology to schools around the world. By both freeing up curricula (curricular maps) and time (flexibly created schedules), we intend to customize education to students more greatly than ever, in a manner affordable for schools throughout the world. Victor Hugo is noted for having said 'There is nothing so powerful as an idea whose time has come.' We argue in the spirit of Hugo that the flexible scheduling paradigm is an idea whose time has come, and may be irresistible.

The paradigm shift we are describing is an idea whose time has come. By integrating the best of many of the advances described in our literature review and some creative innovation, we offer the potential of building schools that can truly customize to student learning needs as well as teacher training needs. While we may not reach Bloom's 2-sigma goal [2], the gains in student performance and teacher capabilities could be very great. We recognize that our simulations (and validation process [25]) provided only a prima facie argument for FSP, and that the real test comes in implementation. We believe the FSP approach may move things very far forward.

7 References

- [1] Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95–135. <https://doi.org/10.1086/508733>
- [2] Bloom, B. (1984). The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4-16. <https://doi.org/10.3102/0013189X013006004>
- [3] Canady, R.L., & Rettig, M.D. (1995). *Block scheduling: A catalyst for change in high schools*. Princeton, NJ: Eye on Education.
- [4] Goldhaber, D. & Anthony, E. (2007). Can teacher quality be effectively assessed? National Board certification as a signal of effective teaching. *The Review of Economics and Statistics*, MIT Press, 89(1), 134-150. <https://doi.org/10.1162/rest.89.1.134>
- [5] Hammann, J.E. & Markovitch, N.A. (1995). Introduction to ARENA. Proceedings of the 1995 Winter Simulation Conference, 519-523. <https://doi.org/10.1145/224401.224674>
- [6] Henry, G.T., Fortner, C.K. & Bastian, K.C. (2012). The effects of experience and attrition for novice high-school science and mathematics teachers. *Science*, 335(6072), 1118-1121. <https://doi.org/10.1126/science.1215343>
- [7] Herer, Y.T. (2007). Lecture Notes from Introduction to Industrial Engineering, Technion, Israel Institute of Technology.
- [8] Hillier, F.S. & Lieberman, G.J. (2010). *Introduction to operations research, 9th edition*. New York, NY: McGraw-Hill Companies, Inc.
- [9] Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
- [10] Koppang, A. (2004). Curriculum mapping: building collaboration and communication, *Intervention in School & Clinic*, 39(3), 154-161. <https://doi.org/10.1177/10534512040390030401>
- [11] Kulik, C.C., Kulik, J.A. & Bangert-Drowns, R. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60(2), 265-306. <https://doi.org/10.3102/00346543060002265>
- [12] Law, A. M. & Kelton, W. D. (2000). *Simulation modeling and analysis, 3rd edition*. Boston, MA: McGraw-Hill.
- [13] Matsumoto, M. (2004). Mersenne twister: a very fast random number generator. Retrieved September 5, 2012, from <http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html>.
- [14] National Center for Educational Statistics (2012). Number and types of public elementary and secondary schools from the common core of data: school year 2010-2011. Retrieved September 5, 2012, from <http://nces.ed.gov/pubs2012/2012325.pdf>.

- [15] National Center for Educational Statistics (2010). Student teacher ratios in public schools. Retrieved September 5, 2012, from http://nces.ed.gov/programs/coe/indicator_qpt.asp.
- [16] National Center for Educational Statistics (2006). State requirements for a standard high school diploma: 2006. Retrieved September 2, 2012, from http://nces.ed.gov/programs/digest/d07/tables/dt07_158.asp.
- [17] National Education Commission on Time and Learning. (1994). *Prisoners of time: report of the national education commission on time and learning*. Washington, DC: U.S. Government Printing Office.
- [18] Pinedo, M.L. (2009). *Planning and scheduling in manufacturing and services*. New York, NY: Springer, Inc. <https://doi.org/10.1007/978-1-4419-0910-7>
- [19] Rivkin, S.G., Hanushek, E.A. & Kain, J.F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458. <https://doi.org/10.1111/j.1468-0262.2005.00584.x>
- [20] Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal*, 41(1), 159-189. <https://doi.org/10.3102/00028312041001159>
- [21] Snedecor, G.W. & Cochran, W.G. (1989). *Statistical methods*. Ames, IA: Iowa State University Press.
- [22] Snyder, J. and Robert, A. (2001) Patent application number PCT/IL01/00376 “System for Scheduling Classes and Managing Educational Resources.” US patent number 8,052,426 granted 2011. Note: the first author goes by his Hebrew name, which is Yaakov, but his English name is Jonathan.
- [23] Snyder, J. (2003). National Science Foundation Grant No. 0232833 given to Highmarks, Inc., of which the author is a Founder. Grant proposal and final reports are unpublished but available from the author. Grant proposal was titled “Highmarks Advanced Planning and Scheduling System for Schools.” Retrieved September 2, 2012, from <http://nsf.gov/awardsearch/showAward.do?AwardNumber=0232833&version=noscript>. Note: the author goes by his Hebrew name, which is Yaakov, but his English name is Jonathan.
- [24] Snyder, Y., Herer, Y.T. & Moore, M. (2011). A new engine for schools: the flexible scheduling paradigm. *Journal of Educational Technology Systems* 40(1) 3-18. <https://doi.org/10.2190/ET.40.1.b>
- [25] Snyder, Y. (2013). Flexible scheduling paradigm: a new middle and high school model. (Unpublished doctoral dissertation, Technion – Israel Institute of Technology).
- [26] Troschuetz, S. (2007). .NET random number generators and distributions. Retrieved September 5, 2012, from <http://www.codeproject.com/Articles/15102/NET-random-number-generators-and-distributions>.
- [27] Trump, J.L. (1958). An image of the future in improved staff utilization. *NASSP Bulletin*, 42, (237), 324-329. <https://doi.org/10.1177/019263655804223764>
- [28] U.S. Department of Education (1999). Taking responsibility for ending social promotion: A guide for educators and state and local leaders. Retrieved September 2, 2012, from <http://www2.ed.gov/PDFDocs/socialprom.pdf>.
- [29] Walberg, H.J. (1984). Improving the productivity of America’s schools. *Educational Leadership*, 41(8), 19-27.

8 Authors

Yaakov Snyder, Ph.D., is an independent researcher and math and science teacher in Baltimore, Maryland, USA. His email is ysnyder@gmail.com.

Yale T. Herer is an Associate Professor of Industrial Engineering at the Technion – Israel Institute of Technology in Haifa, Israel.

Michael Moore is an Associate Professor in Education in Science & Technology at the Technion – Israel Institute of Technology in Haifa, Israel.

Avishai Catane is a Senior Programmer in Bet Shemesh, Israel.

Richard M. Novak is a Senior Programmer in La Jolla, California, USA.

Article submitted 2018-10-10. Resubmitted 2018-11-17. Final acceptance 2018-11-21. Final version published as submitted by the authors.