Pearson

Technical Research Report

MyPedia – India

MyPedia classroom practices and their relationship with student achievement.

20 March, 2019

Authors
Guido Gatti, MA
Priyanka Sharma, PhD
Derek Fay, PhD

Pearson

Table of contents

Pearson

Executive summary

Overview of MyPedia

MyPedia is an integrated, multi-subject system of learning for early and late elementary schools. Developed initially for modern aspiring private schools in India, it aims to support measurable improvement in learners' thinking skills.

MyPedia consists of:

- various teaching and learning tools, including print materials, digital resources and assessment analytics, designed to help learners become active participants in their own learning, not passive recipients of information
- the MyPedia Excellence Program, which provides schools with customized lesson plans, professional development and school support visits, to make sure the teaching and learning tools work as intended for learners

The hope is that — through an academic partnership that strengthens product implementation, teachers' classroom practice, student progress monitoring, and access to learning — MyPedia schools can build a strong academic reputation.

Each lesson in MyPedia is built on the 5i model of instruction, which structures the learning process into five phases: inception, ideation, inculcation, inscription, and inquisition. As students cycle through these phases, they explore novel questions and examples, apply their developing knowledge and skills, and receive feedback on their learning. The 5i model is based on the renowned BSCS 5E model, itself grounded in educational theory and supported by empirical research (Bybee, et al, 2006).

Pearson

Intended outcomes

Here's what success looks like for MyPedia, in terms of learners' and teachers' experiences using the system.

MyPedia first aims to develop a positive attitude toward learning among students. To achieve this, students need to have an engaging, positive learning experience.

1. Learners have a positive experience and are engaged in learning.

Further, it has been hypothesized that learner achievement can be affected by improvements in pedagogy brought about by the professional development of teachers. It makes sense to next study whether teachers' knowledge and skills improve and, ultimately, whether those improvements are associated with increases in student achievement.

2. Teachers improve their subject knowledge, pedagogy, assessment and digital skills.
3. Learners improve their reading ability, language, mathematical understanding and scientific literacy.
4. Learners improve their discipline-specific problem-solving and critical thinking skills.

In the present study, we address learner outcomes #1, #2 and #3.

Research questions

For this study, local expert consultants observed teachers and students using MyPedia in typical private schools across India. As well as these classroom observations, we also analyzed student achievement test scores and the results of teacher and parent surveys.

The main research questions we set out to answer with this study are:

1. What are teachers' and parents' opinions of MyPedia?
2. How are MyPedia's core components implemented in classrooms by teachers and students?
3. How have classroom practices changed during the implementation of MyPedia from the 2017–18 to 2018–19 school years?
4. Does student academic improvement vary with classroom implementation?

Pearson

Key findings

In the context of this study, we can make the following correlational statements about the efficacy of MyPedia (i.e., Findings related to research question 4).

Higher observed MyPedia teaching quality and expected teacher impact ratings are associated with better students' summative scores.

- A one point increase in the MyPedia teaching quality rating is related to a 0.44 standard deviation increase (i.e., 17 percentile points) in students' summative test scores.[1]
- A one point increase in MyPedia teacher impact rating is related to a 0.71 standard deviation increase (i.e., 26 percentile points) in students' summative test scores.[2]

This finding addresses research question 4, Does student academic improvement vary with classroom implementation?

In the same context, we can also make several descriptive statements about the efficacy of MyPedia. These statements are as follows.

Findings related to research question 1

Overall, teachers and parents had high opinions of MyPedia in 2017 and 2018.

- When asked how they would rate the overall quality of MyPedia as a solution for their school, 95% of teachers (2017 = 169/181, 2018 = 158/165) and 93% of parents (2017 = 335/368, 2018 = 173/179) rated MyPedia as either Good, Very Good or Excellent.
- 94% (170/181) of the 2017 and 90% (149/165) of the 2018 teachers agreed MyPedia contributes to students' positive attitude toward learning.

Findings related to research question 2

Averaged across all observed classrooms, overall MyPedia teaching quality ratings increased each quarter from 2017 to 2018 moving from "does not meet standards" to "meets standards".

- The percentage of teachers whose average MyPedia teaching quality rating indicated they were "meeting standards" increased from 50% in quarters 1 and 2 of 2017 to 67% in quarter 2 of 2018.

1. For example, if a teacher's MyPedia teaching quality rating was to increase from 1.5 to 2.5, this, on average, is expected to lead to a 17 percentile point increase on students' end-of-year test scores (These examples are for illustrative purposes only, and not subject to independent assurance.)

2. For example, if a teacher's MyPedia teacher impact rating was to increase from 1.5 to 2.5, this, on average, is expected to lead to a 26 percentile point increase on students' end-of-year test scores (These examples are for illustrative purposes only, and not subject to independent assurance.)

MyPedia teacher confidence rating increased each quarter between 2017 and 2018.
18% of teachers were rated "very confident" in quarters 1 and 2 of 2017, whereas 27% were "very confident" in quarter 2 of 2018.

Findings related to research question 3

Averaged MyPedia teacher impact ratings increased each quarter from 2017 to 2018.

- The percentage of teachers whose average MyPedia teacher impact rating indicated they were "likely to have a positive impact on learners" increased from 36% in quarters 1 and 2 of 2017 to 72% in quarter 2 of 2018.

Results

The results of this study suggest that the schools, teachers, and learners MyPedia is designed for, both enjoy MyPedia and are able to use it effectively.

Teachers' confidence in teaching with MyPedia increased each quarter from 2017 to 2018. Averaged across classrooms, ratings for MyPedia teaching quality and expected impact on learning also increased over this period. Higher observed MyPedia teaching quality and expected impact ratings are associated with better student test scores.

Teachers and parents also expressed high opinions of MyPedia in both 2017 and 2018.

Next steps

We are already planning more research to see if we can replicate the results of this study under stricter conditions. This should allow us to reliably measure student achievement gain and higher order thinking skills, and find out how they are connected to the way teachers are using the MyPedia tools.

Recommendations

MyPedia schools' classroom practice should continue to be observed through the 2019–20 school year and that data should be shared with the research team. This will allow for the replication of results from this study and provide information on how well schools maintain the progress seen in their implementation.

A more rigorous quasi-experimental study is currently being planned that will compare MyPedia to similar schools under typical classroom and professional development practice.

Pearson

Introduction

Background

India has one of the largest school systems in the world. Since gaining Independence in 1947, India has worked successfully, achieving over 90% Universal Primary Education. More recently, the focus has been on improving the *quality* of education. The new education policies, Right to Education, and increasing number of enrolments in secondary education are steps in the right direction. The remaining challenges faced by the Indian education system are high pupil–teacher ratio, lack of professionally trained teachers, and poor levels of student learning (both scholastic and co-scholastic) resulting in weak learning outcomes (British Council, December 2014).

India has the world's largest population in the age bracket of 6–17 years. There is 14.6% growth in rural consumption, a middle class market of 350 million Indians (Saha, D, IndiaSpend, April 17, 2017), and the number of private unaided schools has risen by 35% between the 2010–11 and 2015–16 school years (Kant, et al, LiveMint, January 23, 2018). It is not surprising that currently in India, there is a rise of private unaided schools, serving the middle income bracket with a regional language background. These schools aspire to position themselves as modern, "international" schools with a strong academic reputation.

Currently, Indian private schools may be separated into three tiers:

- 15% are well known schools with good technology infrastructure, larger student numbers, higher fee points, and more experienced staff, and reside in the outskirts of large cities.
- 75% are aspirational schools with 300–1,000 students, a middle fee point, and adequate technology infrastructure, located in district cities or towns.
- The remaining 10%, also with a middle fee point, may have little technology infrastructure (for example, a very small computer lab, no wifi/internet). The students will typically be first generation learners and the parents and school leadership may not be as committed to modern technology-based education.

The MyPedia schools fall into the middle and largest segment of private schools.

Typical MyPedia schools charge a modest fee and are located in district towns, serving 300–1,000 students. They tend to have a high pupil–teacher ratio of around 40+ students per section, and offer limited but modern infrastructure (for example, a computer lab). MyPedia schools are typically staffed with young, less experienced teachers, with three or fewer years' teaching experience, possessing the requisite content knowledge but needing improved pedagogical understanding and experience.

The schools' leaders are looking for an academic partner that can provide the necessary professional development programs for their teachers, a student focused classroom with increased teacher-to-student and student-to-student interactions, increased access to learning for families, and progress monitoring aided by better implementation of technology.

Description of MyPedia India

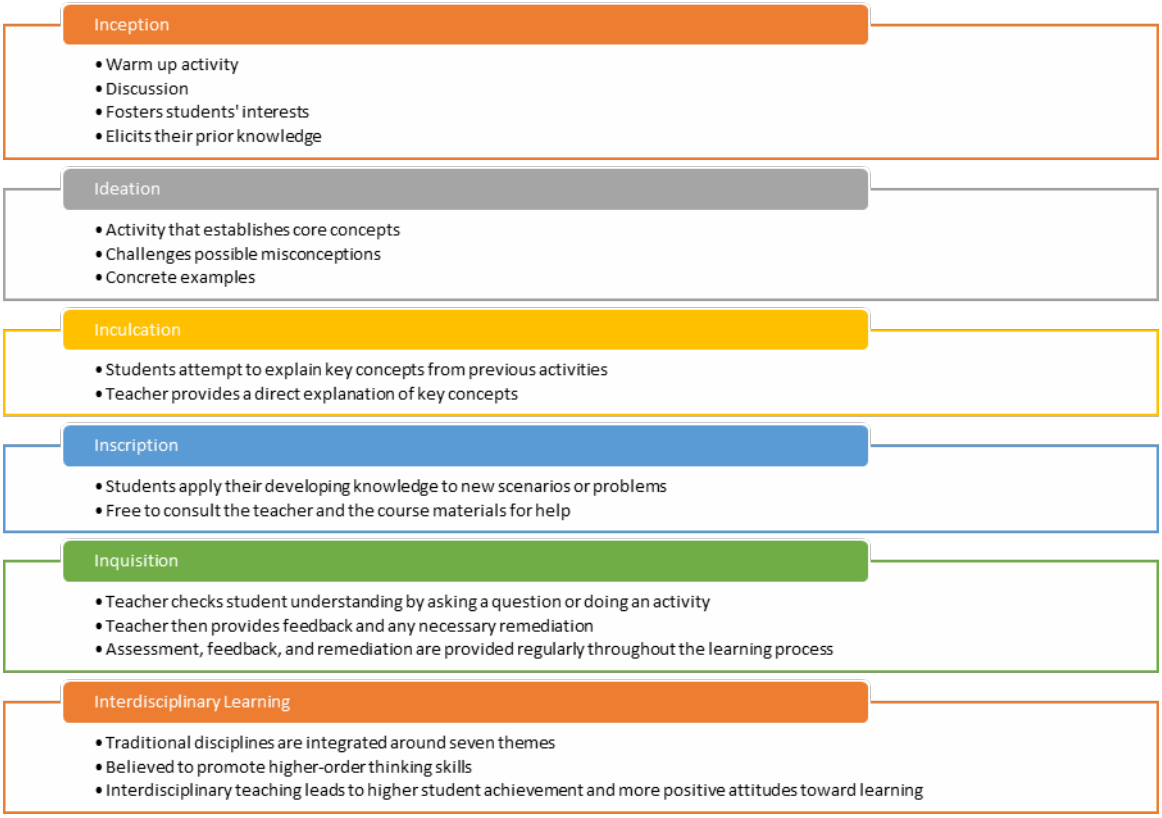
MyPedia was developed with modern private schools in India, and the families they serve, in mind from the beginning. In particular, the goal is to for MyPedia to help schools build strong academic reputations by improving teachers' classroom practice, student progress monitoring, and access to effective learning opportunities.

MyPedia is an integrated, multi-subject, cross-curricular, early and late elementary learning (whole school) solution that aims to support measurable improvement in the thinking skills of the learner. MyPedia is designed so that every class enables learners to be active participants in their own learning instead of just being passive recipients of information. This is achieved by integrating various learning tools like:

- course books designed on an interdisciplinary approach
- worksheets to strengthen application of concepts
- content mapped digital resources
- Home App-based video content and practice assessments for learners
- cutting edge assessment analytics for learners, teachers, and school leaders

To ensure these components are accessible to users as intended, the MyPedia Excellence Program engages and partners with schools in their journey through period-wise customized lesson plans, a professional development program for teachers (focused on improving teachers' subject knowledge, pedagogy, assessment, and digital skills), a school leadership program, and school support visits.

Further, each lesson in MyPedia is built upon the 5i model of instruction, which structures the learning process into five phases: inception, ideation, inculcation, inscription, and inquisition. As students cycle through these phases, they explore novel questions and examples, apply their developing knowledge and skills, and receive feedback on their learning. The 5i model is based on the renowned BSCS 5E model (Bybee, et al, 2006), an instructional model founded in educational theory and supported by empirical research. The 5 "i"s in 5i are inception, ideation, inculcation, inscription, and inquisition. Plus MyPedia also includes a sixth: interdisciplinary learning.



What is inception?

The teacher leads a warm-up activity or discussion that fosters students' interest and elicits their prior knowledge. For example, a 2nd grade lesson on the differences between living and non-living things begins with students discussing with one another how much they have grown over the past few years, prompting them to think about how they grow and change over time.

Why is inception effective?

Engaging students' prior knowledge — here, about how living things grow — is a crucial component of teaching and learning, because learning fundamentally involves adding to and reorganizing existing knowledge according to new information. Additionally, fostering students' interest supports learning by increasing attention and positive emotion. Principals and teachers who use MyPedia notice its effect on students' emotion, with 94% of those surveyed agreeing that MyPedia contributes to students' positive attitudes towards learning. See Ausubel, D. P. (1968), Mayer, R. E. (1979), Roschelle, J. (1995), and van den Broek, P. (2010), for more on activating students' prior knowledge, and see Hidi, S. (2006) and Krapp, et al (1992) for more on sparking students' interest.

Pearson

What is ideation?

Students complete an activity that establishes core concepts, challenges possible misconceptions, and introduces a set of concrete examples that form the basis for further instruction. For example, the lesson on living versus non-living things continues with a discussion of whether certain objects in the classroom, such as the tables, grow over time. Comparing students and classroom objects establishes concrete examples of living and non-living things.

Why is ideation effective?

Learning new concepts, such as the properties of living things, can require students to think in the abstract. Research suggests that one effective approach to teaching abstract concepts is to begin with concrete examples and slowly progress to more abstract thinking. Learning also often involves confronting existing misconceptions, which is itself a productive part of the learning process. Establishing and discussing examples can help uncover and challenge misconceptions. Please see the references from Fyfe, et al (2014) and Smith, et al (1993).

What is inculcation?

The teacher directs students' attention to key concepts from the previous activities, often asking students to attempt to explain them. The teacher then provides a direct explanation of key concepts. For example, students attempt to explain why humans are living things and classroom objects like the tables are not. The teacher then explains the primary differences between living and non-living things.

Why is inculcation effective?

When students develop their own explanations, they deeply engage their prior knowledge and relate it to what they have just learned or experienced. Doing so leads to greater learning than simply being presented with the teacher's explanation. In addition, students' explanations can be used by the teacher to adjust their teaching to address any specific misconceptions. See the following references regarding student self explanation: Pressley, et al (1992) and Fonseca, B. A., & Chi, M. T. H. (2011).

What is inscription?

In this phase, students apply their developing knowledge to new scenarios or problems. They are free to consult the teacher and the course materials for help. For example, the students complete a worksheet that presents a tree, an airplane, and a cat and asks students to use what they have just learned to determine why each is either a living or non-living thing.

Pearson

Why is inscription effective?

A central goal in learning is *transfer*, or the ability to apply knowledge to new situations. Completing worksheets and other activities gives students practice in transferring their knowledge. Because students can consult the teacher and course materials, they also learn how to use available resources to help them, which is an important part of being able to transfer knowledge effectively. See these references on knowledge transfer: Bransford, J. D., & Schwartz, D. L. (1999) and Schwartz, et al (2005).

What is inquisition?

The teacher checks students' understanding by asking a question or doing an activity. The teacher then provides feedback and any necessary remediation. Assessment, feedback, and remediation are provided regularly throughout the learning process. In the lesson on living and non-living things, the teacher reviews the worksheet with the students, assessing students' answers and clearing up any misunderstandings that those answers reveal.

Why is inquisition effective?

Assessment is not only a means to measure the outcome of the learning process — it also forms the basis for providing feedback and personalizing instruction. Research shows that regularly providing learners with feedback is crucial to learning. See these references on formative assessment: Bennett, R. E. (2011), Shute, V. J. (2008), and Wiliam, D. (2011).

What is interdisciplinary learning?

MyPedia utilizes an interdisciplinary approach to teaching and learning. In this approach, the traditional disciplines are integrated around seven themes, such as *Animals and Plants* and *Innovation and Technology*, through which students are encouraged to draw connections across disciplinary boundaries. For example, an 8th grade *Myself and My Family* science lesson teaches about sound waves. At the end, students link this lesson to English and social studies by writing a letter to the editor of a hypothetical newspaper that proposes how to reduce noise pollution in a crowded city. This exercise requires students to draw connections between what they know about the science of sound, the social aspects of cities, and writing to develop and clearly explain a solution to the problem. Interdisciplinary teaching is believed to promote higher order thinking skills and research suggests that, relative to traditional teaching approaches, interdisciplinary teaching leads to higher student achievement and more positive attitudes toward learning. To read more on interdisciplinary teaching see Jacobs, H. H. (Ed.) (1989) and Vars, G. F. (1997).

In addition to offering teachers instructional materials, MyPedia also attempts to change classroom practice through supporting the teachers in improving their MyPedia instruction. The MyPedia Teacher Development (MTD) program is essential to successful implementation of the MyPedia solution for two reasons:

- teachers must master the whys and hows of the key materials and applications (that is, print materials, digital assets, assessment and reports, and teaching plans)
- teachers are expected to become proficient in using modern pedagogies in their classroom teaching

MTD is a core responsibility of the Academic Relationship Managers (ARMs). The program begins with a two day orientation prior to the school's start of the new academic year. The focus is on training the teachers in using the materials and applications in their classroom teaching. Teachers begin to implement MyPedia the following 6–8 weeks, during which time the ARM visits the school to record classroom observations and facilitate implementation. About two months after the orientation, another training workshop is held in the school, and this time the focus is on making the teachers familiar with the student progress planning and monitoring.

After this training, ARMs continue to visit the school every 6–8 weeks to ensure effective delivery of the MyPedia program, focusing on continually developing pedagogy and improving areas of weakness. The MTD for renewal schools may focus more on this continual development. Here there is give and take, because these schools tend to have high teacher turnover. If most of the teachers are already familiar with the components and pedagogy of MyPedia, they are introduced to the Online professional development applications and advanced digital tools offered to students, such as Actual Reality–Virtual Reality, STEM and scratch programming.

The ARMs undergo 8–10 days of training to upskill them for their role. This training focuses on:

- product training in all components of MyPedia
- pedagogy training (that is, classroom management, Constructivist learning methods, formative assessment, and student progress planning)
- facilitation, coaching, and mentoring skills

In 2018, they received specialized training on facilitating online professional development, which is expected to roll out in 2019 to their MyPedia schools. They are supported with virtual sessions across the year, and often with any special requests from some of their schools. As such, professional development works on a continuous cycle of feedback and improvement between the ARM, the teachers, and the product and delivery teams, with the overarching aim of ensuring that overall classroom teaching is improved.

Pearson

Summary of the learner outcomes for MyPedia

MyPedia was designed to be learner focused, skills oriented, and professional developing. As such, listed here is a summary of the intended learner outcomes for the MyPedia program.

1. Teachers will improve in subject knowledge, pedagogy, assessment and digital skills.
2. MyPedia students have a positive experience and are engaged in learning. MyPedia aims to develop interest among learners, and develop a positive attitude toward learning, through a positive learning experience and engagement.
3. Learners improve performance in reading ability, language, mathematical understanding and scientific literacy.
4. Learners improve in discipline specific problem solving and critical thinking skills. MyPedia has embedded discipline specific problem solving and critical thinking skills in a very natural and learner-friendly manner with the goal of measurable progress of students in these areas.

In the present study, we will be addressing learner outcomes 1, 2, and 3.

The present study

The purpose of this study is:

- to understand how MyPedia components are being used in classrooms by teachers and students
- to examine if teachers using MyPedia demonstrate improvement in pedagogical skills and classroom practice
- to determine if a measurable relationship exists between the strength of implementation of MyPedia and student achievement

To answer these questions, teachers and students were observed (by local expert consultants) using MyPedia in typical private schools across India to investigate how well MyPedia components were implemented in these classrooms. More importantly, we determined if the use of MyPedia is associated with improved classroom practices in these schools over the school year.

The study used a mixed methods approach to gather a range of evidence types, including:

- classroom observations — to inform about implementation and changes in MyPedia teaching quality
- student achievement test scores — to inform about impact on learner outcomes
- teacher and parent surveys — to speak to users' experience, satisfaction and brand loyalty

Pearson

Method

In this research effort we employ a confirmatory implementation study design, which will attempt to connect program implementation to product outcomes. Information on MyPedia implementation and teaching quality, teachers' attitudes and confidence, as well as teachers' expected impact on learning, will be presented to understand the experiences of students and teachers using MyPedia. Because one of MyPedia's core components is professional development aimed at improving MyPedia teaching quality, we capture the strength of MyPedia implementation and its impact via classroom observations that depict teachers' use of effective teaching practices related to things like classroom management and formative assessment/feedback. In addition, the study will examine the relationship between MyPedia teaching quality and students' achievement after taking into account their baseline knowledge.

Participants

Fifty schools in India, that began using MyPedia, were randomly selected to be part of an implementation study spanning the 2017–18 school year and the first half of the 2018–19 school year. These schools were selected from a pool of early adopting schools. This sample of schools had a combined enrollment in excess of 20,000 students in kindergarten through 8th grade.

Over the course of time, some original schools left the study and new schools were added as logistics permitted. Schools left the study for several reasons, either declining to have teachers' lessons observed and reported to the research team, students not being available for testing, or simply discontinuing with the MyPedia program. This ultimately led to 74 schools having been observed at some point during the study period. The 74 schools may be categorized into three groups according to their involvement in the study. The categories are:

- 22 of the original 50 schools which were followed from the 2017–18 into the 2018–19 school year: schools observed at least once in 2017–18 and/or 2018–-19 school years and students tested in the beginning of the 2017–18 and 2018–19 school years
- 28 of the original 50 schools which were only followed in the 2017–18 school year: students not tested in the 2018–19 school year
- 24 schools added in 2018–19: started MyPedia in the 2018–19 school year, observed in 2018–19 school year but students not tested in 2017–18 school year

The group of 22 schools is of special interest because for these schools we have the necessary information to estimate the relationship between implementation and student achievement (related to research question 4).

Pearson

Basic information on the study schools' characteristics was collected from trusted online databases. As seen from Table 1, about half of the original MyPedia participating schools were located in Tier 3 population density areas or minor cities and towns. The remaining schools were split between Tier 2 (small city, population greater than 1,000,000) and Tier 1, or the largest cities in India. A similar breakdown is seen with the total sample of 74 schools. The group of 22 schools from the original 50 had more Tier 2 schools, or schools from small cities, and equally fewer Tier 1 and Tier 3 locations. The shift away from Tier 3 could mean this sample has larger enrollments, more trained teachers, and better access to educational technology. It would not necessarily mean the students in the sample receive better instruction and education materials. All 74 schools use the same materials and applications school wide and they receive similar ongoing MyPedia training and professional development.

Table 1: Participating MyPedia schools by population tier

School Sample	Schools	Tier 1 Percent	Tier 2 Percent	Tier 3 Percent
Original 2017 starters	50	28.0%	20.0%	52.0%
Followed 2017–18 and 2018–19	22	18.2%	40.9%	40.9%
Not tested 2018–19	28	35.7%	3.6%	60.7%
New in 2018–19	24	29.2%	37.5%	33.3%
Total	74	28.4%	25.7%	45.9%

Note: Tier 1 denotes metro or large city, Tier 2 denotes small city, and Tier 3 denotes minor cities and towns (population < 1,000,000).
Note: All rows sum to 100% without rounding error.
Note: All rows are not mutually exclusive. The Original 50 and the New 24 sum to the 74 total. Also the groups of 22 and 28 sum to the Original 50.

Of the 74 participating schools, 42 (57%) were found currently listed on the Central Board of Secondary Education (CBSE) website (see Table 2). The CBSE is a certification board managed by the national government of India. The over 19,000 CBSE affiliated schools are expected to follow the NCERT curriculum. Four other schools are affiliated with the Indian Certificate of Secondary Education (ICSE) and one is state certified. The ICSE is not managed by the Indian government, is a nonprofit, and offers school wide examinations in English. The original sample of 50 schools, and the group of 22, follow a similar breakdown of CBSE and non-CBSE affiliated schools. All the schools, including those with no official affiliation, followed a similar set of process and content standards since they were MyPedia schools and participated in the study.

Pearson

Table 2: Participating MyPedia schools by affiliation

School sample	Schools	No official affiliation	CBSE	ICSE	State certified
Original 2017 starters	50	34.0%	60.0%	6%	0%
Followed 2017–18 and 2018–19	22	31.8%	63.3%	4.5%	0%
Not tested 2018–19	28	35.7%	57.1%	7.1%	0%
New in 2018–19	24	41.7%	50.0%	4.2%	4.2%
Total	74	36.5%	56.8%	5.4%	1.4%

Note: All rows sum to 100% without rounding error.
Note: All rows are not mutually exclusive. The Original 50 and the New 24 sum to the 74 total. Also the groups of 22 and 28 sum to the Original 50.

The original 50 schools had an average total enrollment of 447 students in 2017. The group of 22 schools had similar enrollment numbers. The enrollment numbers were even more similar across 1st through 5th grade (see Table 3).

Pearson

Table 3: Enrollment for participating MyPedia schools

Group	Schools	Mean	Std. deviation	Median	Min	Max
Total enrollment	Original 50	447	357	347	68	2,334
	Group of 22	431	559	335	104	2,334
Enrollment in 1st through 5th grade	Original 50	345	234	335	45	1,384
	Group of 22	346	274	323	104	1,384

Note: 1st through 5th grade enrollment was only available for 48 of the original 50 schools.

Pearson

Of the 50 original schools, 30 were listed on the official CBSE website. The website had information on the availability of computer labs for 18 schools and the number of trained 1st through 5th grade teachers for 17 schools. The group of 22 schools had 14 schools listed on the CBSE website and 9 schools with this information. Only the CBSE website was found to share this information. For the schools that did have the information, the ratio of 1st through 5th grade students to trained teachers was similar for the original 50 and group of 22 schools (see Table 4).

Table 4: 1st through 5th grade student to trained teacher ratio

Schools	N	Mean	Std. deviation	Median	Min	Max
Original 50	17	56.51	34.75	46.13	12	153.78
Group of 22	9	58.37	46.48	41.63	12	153.78

The study schools tended to have a single computer lab, as seen in Table 5. Only one school, a member of the group of 22, has no lab. Three schools across the sample of original 50, new starters, and the 22 group have more than one. The limited access to technology is the reason MyPedia offers the home and mobile student applications.

The yearly fees are known for 14 of the schools in the group of 22. The average fee for attending one of these schools per school year is $442 US dollars (standard deviation = $129, median = $449, minimum = $210, maximum = $674). This fee structure and technology availability is indicative of the wider class of aspiring private schools MyPedia is designed for.

In summation, from the available data we can say that the schools in the analytic sample (that is, the group of 22 schools) appear to be similar in important ways to the original 50 schools. These schools had similar total and 1st through 5th grade enrollment as the original 50 school sample, and for the schools that had such data, the ratio of 1st through 5th grade students to trained teachers was very similar. Further, the original and analytic samples had the same percentage of CBSE certified schools, schools in both groups tended to have a single computer lab, and the yearly fees are modest.

Twenty-two of the original 50 schools ultimately participated in the study from the start of the 2017–18 school year. These schools have both the student achievement scores and the classroom observation ratings necessary to estimate the relationship between implementation and student achievement. Tables F1 through F4 show the availability of classroom observations and their alignment to subject area and grade level for these schools.

Pearson

The classroom observation protocol and student achievement testing are described more fully in the Measures section. Student test scores were collected in May–June 2017 (n = 9,844) and again in May–July 2018 (n = 12,537) for the same students in the subsequent grade (n = 6,334). To assess the relationship between classroom observation ratings and test scores, it was necessary to have both 2017 and 2018 scores for a student, as well as observation ratings for that student's grade level on a subject area tested (that is, English, math, or science). Ultimately, 3,596 student test score and classroom rating pairings from 22 schools allowed for this relationship to be tested.

Table 5: Number of computer labs for participating CBSE MyPedia schools

School sample	Schools	Schools with data	No computer lab	One computer lab	Two computer labs	Three computer labs
Original 2017 starters	50	18	1	15	1	1
Followed 2017–18 and 2018–19	22	9	1	6	1	1
Not tested 2018–19	28	9	0	9	0	0
New in 2018–19	24	7	0	6	1	0
Total	74	25	1	21	2	1

Note: All rows are not mutually exclusive. The Original 50 and the New 24 sum to the 74 total. Also the groups of 22 and 28 sum to the Original 50.

Measures

Implementation of product in study

The participating study schools used the MyPedia teaching–learning program across all elementary grades and course subject areas as their primary curriculum, student instruction and progress monitoring plan, and professional development system. MyPedia is an integrated teaching–learning program, meaning that the learning design principles and the content framework are common for all the MyPedia components. This is evident in the seven common themes that are woven across all subjects and the Course Book's references to cross-curricular links (for example, when students are studying about large numbers in elementary school math, they are asked to research the distance of the planets from the sun and refer to their Social Science book for more information). Both the Teaching Plans and the Assessments are based on the same Learning Objectives framework. The MyPedia Interim and Term Reports map students' progress against these Learning Objectives.

The Teaching Plans are at the heart of the MyPedia program. They guide the teacher on the classroom pedagogies and the lesson delivery best suited for achieving the Learning Objectives. Teachers receive training on how best to deliver the teaching plans. They are encouraged to adapt and contextualize the Teaching Plans to their class needs. However, the Learning Objectives cannot be changed and the Learning Objectives stated in the original plan need to be achieved. Teachers undergo a comprehensive professional development program prior to the start of the academic year to introduce them to all MyPedia components and their delivery. They undergo hands-on training on the assessment portal and digital assets. They are guided on the pedagogies necessary to deliver the Teaching Plans and are encouraged to reflect on their delivery. The ARMs observe teaching in action and share their feedback.

The MyPedia program has been constructed to engage students in their learning, improve their application of understanding and skills, and provide a measurable path for their cognitive development. Typical MyPedia classes are interactive and their assignments encourage independent learning skills, whether through the end-of-chapter tasks or the worksheets. This is further enabled through the digital assets, which are learning videos linked to the topic and available to each student at home through the Home App.

A typical MyPedia month

Let's take a look at a typical month for a MyPedia teacher and students.

The teacher would have prepared to deliver the content effectively by reading the Teaching Plans in advance (at least a week prior to delivery), and understanding the concept flow between the Course Book, Application Book and digital assets. They would need to arrange for the materials and any other requirements for the classroom teaching. For example, if the teaching plans suggest that students view a digital asset such as a video, the teacher may need to connect with the ICT teacher or book the AV room, depending on the school's infrastructure. The teacher would connect with other subject teachers to discuss the cross-curricular segments indicated in the teaching plans. The teacher would also plan which of the worksheets would be assessed by the student, their peers, or their teacher, and which would be better suited to individual or group work. In addition, the teacher would be prepared to assign the formative assessments primarily found in the Course Books, worksheets and Home App quizzes.

In a month, usually two main concepts and related concepts are completed. The students are introduced to the new concepts through warm-up activities that draw on their prior knowledge. Lots of class discussions, question-and-answers, and simple group activities are encouraged at this stage in the Teaching Plans. This usually takes one or two class periods. Teachers then lead the students' learning to deeper conceptual understanding, and assign group work and worksheets. This forms the bulk of between three and five class periods. There are formative, oral assessments during these classes, while in the last six to eight lesson plans, the written assignments and assessments help measure student learning. The teacher is able to identify which students require more support and the best suited remedial support that the students need.

Every six to eight weeks, students receive a Periodic Test based on the four or five topics covered thus far. The scores obtained by each student in the worksheets, assignments and the Periodic Tests are collated and contribute toward the Term Report. There are two term examinations — the half-year and the term-end. Thus, twice a year, students sit through formal, summative examinations for each curricular subject. The Term Report is the averaged score of formative and summative assessments.

Data collection

First we lay out the time frame for the study's data collection processes. The academic year in the study schools, by quarter, is:

Q1 → April–May–June
Q2 → July–August–September
Q3 → November–December
Q4 → January–February

Classroom observations took place from April 2017 through September 2018. During this time, Academic Relationship Managers (ARMs) observed 557 classroom lessons from 74 schools. The ARMs both train and support the schools in the implementation of MyPedia components and in developing teachers' pedagogical skills and instructional practices. All ARMs have a teaching background (most are former teachers) and are full-time Pearson employees. Each ARM may support 15–20 schools, with a minimum of nine visits in a school year (three for training purposes, six for classroom observations and resolving any emerging issues).

During these school visits, the ARMs tried to observe a range of grade levels, subject areas and teachers. Because the ARMs are responsible for supporting the entire school and try to cover as much of each school as possible, rarely was a teacher observed twice in the same school year. Implementation, skill, confidence, and expected student impact were rated against specific benchmarks on as many as 20 individual criteria (which are described more fully below in the Measures section).

In addition to the observations, test scores were collected in May–June 2017 (n = 9,844) and in May–July 2018 (n = 12,537) on the MyPedia Skills Mapping Tests (SMT), which are also described more fully in the Measures section. Students in 2nd through 5th grade were tested in 2017, and these same students were tested again in 2018, when they were in 3rd through 6th grade. Students took the SMT appropriate for their current grade level in each of the three subject areas: English, math, and science. The SMTs were scaled using item response theory.

To assess the relationship between classroom observation ratings and test scores, it was necessary to have both 2017 and 2018 SMT scores for a student, as well as observation ratings for that student's classroom/grade level on the subject area tested (that is, English, math or science; see Tables F1 through F4). Ultimately, 3,596 SMT and classroom rating pairings from 22 schools allowed for this relationship to be tested.

Teachers' and parents' opinions about MyPedia were collected each year with an independent survey conducted by an independent company in India.

Classroom observation tool

The research team systematically collected information to describe MyPedia classroom implementation using the classroom observation tool. The observation form was put together by the research team specifically to measure the implementation of MyPedia and estimate its expected impact on students. The form included input on measurement criteria to cover all the elements that the product development, product improvement, and professional development teams felt were essential to proper implementation of the program. Descriptors of proper implementation also accompanied each criterion. Because the form was developed to be closely aligned with MyPedia classroom instruction, it is expected to be valid in assessing its implementation and the expected student impact following from that implementation. However, because a core component of the MyPedia program is the professional development aimed at improving teaching quality, assessing implementation entails capturing essential aspects of good MyPedia teaching practice, such as effective questioning of students, providing feedback during teaching, and caring for students.

Pearson

Further, the observers are ARMs, who are extensively trained and experienced in MyPedia classroom practice, so they are very familiar with how to evaluate each criterion in the context of a given lesson. The observers undergo 8–10 days of training to prepare them for their role training and supporting MyPedia teachers. They also provide professional development to these teachers and maintain a relationship with them. Not counting MyPedia product training, the observers are scheduled to be in each school five days each school year. The ARMs can gauge, for example, a teacher's confidence or content knowledge by both observing them and talking with them about the lesson. The repeated visits also allow the observers to become more familiar with teachers' progress, concerns, strengths, and areas for improvement. Observers were trained specifically to use the observation tool during a full two day seminar. They also received a day of follow up training and ongoing support from their professional development team when questions or issues arose.

Operationally, each observation was completed by a single rater, so it is not possible to calculate inter-rater reliability. Also, the same schools were observed by a single observer. The observers reported the information back to the research team via an online form.

This tool is comprised of three parts:

- the School Rating System (SRS)
- the Student Impact Expectation (SIE) rating
- the Teacher Confidence rating

These components combine to define and measure effective classroom practice by providing a picture of program usage, pedagogical skills, teacher confidence, and expectations of learner impact. See Appendix A for a detailed description of the rating criteria for the classroom ratings.

School Rating System

The School Rating System (SRS) gives a rating to each observed lesson on each of 20 separate criteria against specific benchmarks designed to indicate that the instruction does not meet (1), meets (2), or exceeds (3) expectations. Some examples of the criteria include the teacher's understanding of content, teaching methodology/pedagogy, effective questioning and feedback during teaching, connecting to previous knowledge, quality of classwork/homework assignments, and use of ancillary/reference materials and in-class digital applications. Appendix A details the SRS criteria and the specific descriptions for each level of rating.

The descriptor for a rating of 2 or “meets expectations” can be viewed as doing the minimum prescribed by the program for proper implementation. As this implementation model was designed in concert with the instructional materials and professional development plan, measuring adherence to the implementation model amounts to measuring the construct of MyPedia teaching quality. Further, demonstrating a change in MyPedia teaching quality over time may be viewed as a measured change in classroom practice.

For example, a mean score of 2.0 or above across the criteria would indicate that the average teacher in the sample is “meeting expectations”, and that this might be the expectation for a similar population of teachers. If the average moves from 1.5 in the beginning of the school year to 2.5 at the close of the school year, this would indicate that the MyPedia teaching quality has changed for the better.

Student Impact Expectation rating

The purpose of the Student Impact Expectation (SIE) rating is to predict student impact from implementation. These ratings were begun late in the second quarter of 2017 and were completed as normal practice for site observations in the third quarter. The criteria from the SRS framework informs the SIE rating. The same observer completes both the SIE ratings and the SRS rating.

The construct (that is, impact on the learner) and the scale (that is, expectation categories) are the same across all the criteria. The score for each criterion and the overall score mean the same thing, which is that it speaks to the expected impact on the learner. We would expect a higher score to indicate a greater impact on the students.

The SIE ratings take the following form:

Stem ⇒ This teacher's _____ was observed at a level where you expect it will:

- Level 1 — Not have a positive impact on the learners
- Level 2 — Have a positive impact on the learners
- Level 3 — Have a greater than expected impact on the learners

Pearson

Similarly to the SRS, a mean SIE of 2.0 would indicate that the average teacher in the sample is thought to be having a positive impact on the learners and that this might be the expectation for a similar population of teachers. The SIE criteria are:

1. Classroom management
2. Teacher understanding of subject
3. Use of teacher lesson plan
4. Use of MyPedia textbooks and workbooks
5. Use of teaching methods and techniques
6. Monitor student progress
7. Use of quality of assignments
8. Use of connecting to previous knowledge
9. Use of manipulative ancillary supplementals
10. Use of MyPedia digital features in classroom
11. Use of involving students in learning

Teacher Confidence rating

The third component of the classroom observation tool is the Teacher Confidence rating. This part assesses teachers' confidence level on usage of six critical components and pedagogical skills. Here ratings are given on a scale ranging from 0 through 10:

0: Not at all
1–3: Very little confidence
4–7: Moderate confidence
8–10: Very confident

Ultimately we would like to see teachers at a confidence rating of 7 or above, indicating teachers are comfortable with the demands of MyPedia and are making MyPedia their own.

The six components are confidence:

1. in using the questions to develop higher order thinking skills
2. in instruction methods and techniques
3. in checking for understanding
4. when using checkpoints
5. when connecting to previous knowledge
6. when involving students in learning

Pearson

Classroom observation tool measurement error and statistics

The composite SRS, SIE, and Confidence ratings demonstrated little measurement error and correlated well with each other. The individual criteria ratings correlated well with their composite rating.

- Coefficient alphas (all criteria rated): SRS = 0.94, SIE = 0.90, Confidence = 0.98
- Criteria to total correlations: SRS = 0.45–0.77, SIE = 0.54–0.77, Confidence = 0.88–0.95
- Pearson product moment correlation:
- SRS–SIE = 0.834
- SRS–Confidence = 0.646
- SIE–Confidence = 0.667

An exploratory factor analysis also shows the SIE ratings measure a single construct, presumably expected student impact. The first factor explained 52% of the variation in ratings for the listwise sample (n = 124) and from 54.1% to 55.6% for five multiple imputation samples (n = 454). Only the listwise sample had a second factor Eigenvalue over 1.0 (that is, 1.07).

Table 6: Percent of variance across SIE principal components

Criterion	Listwise mean (SD)	Pooled mean (SD)	Component	Percent of variance					
				Listwise	5 imputed samples				
1	1.95 (0.539)	1.88 (0.601)	1	52.39	54.99	55.62	55.13	55.26	54.10
2	2.10 (0.379)	2.08 (0.492)	2	9.70	8.10	7.24	7.07	7.45	7.44
3	2.00 (0.526)	1.97 (0.583)	3	6.34	5.62	5.79	5.94	6.23	6.19
4	2.02 (0.477)	1.96 (0.534)	4	6.17	5.33	5.13	5.22	5.12	5.96
5	2.07 (0.463)	2.03 (0.527)	5	5.51	4.82	4.82	4.75	5.02	5.11
6	1.98 (0.371)	1.92 (0.493)	6	4.32	4.13	4.24	4.27	4.31	4.32
7	2.05 (0.379)	1.97 (0.500)	7	4.26	4.03	4.05	4.04	3.98	3.91
8	2.02 (0.517)	2.00 (0.559)	8	3.58	3.69	3.56	3.80	3.46	3.61
9	1.87 (0.493)	1.86 (0.538)	9	3.12	3.50	3.46	3.58	3.38	3.49
10	1.77 (0.572)	1.79 (0.542)	10	2.61	3.17	3.21	3.16	3.00	3.14
11	2.06 (0515)	1.98 (0.564)	11	2.00	2.61	2.90	3.04	2.80	2.74

The high correlation between SRS and SIE would indicate that about two thirds of the variation in these ratings is shared or measuring the same thing. That does, of course, also mean that one third is measuring something different. As is described more in the sections below, the SIE attempts to take the next step from the SRS criteria, and asks the same observer to then gauge the expected learner impact. In addition, both the SRS and SIE ratings are statistically significant predictors of student achievement, after adjusting for prior achievement (a key finding of this study).

See Appendix B for a breakdown of the frequencies of the individual rating criteria. In looking at the frequencies, it is important to keep in mind that not all criteria would be observed in each visit. Also, more SIE and Confidence criteria will be unobserved because these rating systems were added later in the process or the observer may not have felt confident in assigning a valid rating in a situation. The frequencies do indicate, however, that the observers were mindful of accurately applying ratings to what they experienced. For example, they did not just give all teachers high or average ratings; they used all rating categories.

The research team was not able to collect inter-rater reliability or rater agreement information on the classroom observation tool. For this category of reliability evidence, we are referring to how two different ARMs would rate the same classroom instruction on the same occasion. If the classroom observation tool is part of a standardized measurement process it should not matter, in a meaningful way, which ARM observes which school or classroom. Though this information would undoubtedly be valuable, it was very difficult to collect in the field during the data collection process. The ARMs support a system of schools, and altering the observation schedule to have multiple ARMs revisit the same schools and observe the same classrooms in short order may mean less support to another school and more disruption to the schools that are revisited.

Teacher and parent surveys

Teachers' and parents' opinions about MyPedia were collected each year with an independent survey. The surveys were conducted by a company in India well known for conducting customer satisfaction surveys. A representative sample (by location) of parents were interviewed by phone if their student's school had used MyPedia during the current school year. Respondents were interviewed over a month's time. Table 7 shows the number of schools, teachers, and parents responding to the surveys, as well as those representing the 74 schools from this study. The results presented for these surveys include all responses, not just those for the schools in this study. The response rate for the 2017 parent survey was reported by the survey company to be 77%. Face-to-face interviews were sought with teachers using MyPedia: the response rate was reported as 91%. In these surveys, the research team included questions regarding MyPedia as an integrated, multi-subject, school wide solution. Summaries of the survey questions may be found listed in Appendix C.

It should be noted here that in 2017, 4% (17/385) of the parents surveyed did not respond to the question asking about the overall quality of MyPedia. This disclosure is of interest because response to this question, in part, forms the basis for one of the efficacy statements.

Pearson

Table 7: Breakdown of satisfaction survey respondents

Survey participants	2017 study	2017 total	2018 study	2018 total
Schools	15	45	40	79
Teachers/Principals	50	181	80	165
Parents	212	385	108	179

MyPedia Skills Mapping Tests

The Skills Mapping Tests (SMTs) are offered as part of the MyPedia curriculum. They are short student achievement benchmarking tests in the subject areas of English, math, and science. The tests cover the topics from the previous school year's syllabi. There are 30 questions in each subject area test that vary in their level of cognitive demand (that is, Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating). The SMTs were administered at the beginning of each school year, May–June 2017 (3rd–5th grade) and again to the same students in May–July 2018 (4th–6th grade). In each year, students take the SMT content aligned for their current grade level. Table D1 shows the sample sizes for each type of assessment (grade level 2017, grade level 2018), subject (English, math, science), and students' current grade levels.

In an attempt to ensure the SMT scores were on a similar scale, and that the achievement information was maximized and measurement error was minimized, scaled scores were calculated using Item Response Theory (IRT). Prior to fitting IRT models for each grade level within each subject, two criteria were applied to ensure that the best questions were contributing to students' standardized score estimates. The first selection criterion involved removing questions that were either too difficult or too easy. A question was deemed too difficult if fewer than 5% of students got the question correct; a question was deemed too easy if more than 95% of students got the question correct. The second selection criterion involved removing questions that exhibited a weak relationship to students' overall performance; this was measured using the point-biserial correlation coefficient. Questions with a point-biserial correlation coefficient less than 0.30 were removed. These criteria were applied to each assessment by grade, subject, and year.

Pearson

Figure D1 shows the remaining questions for each SMT by grade (along the horizontal access), subject (by row on the right hand side), and year (by column). If a black dot is present for a given question (shown along the vertical axis on the left hand side), the question was retained for subsequent analyses. For instance, all 30 questions were retained for the 2nd grade English test in 2017. For the same year and grade, one question was removed from the math (observable 5 removed) and science (observable 9 removed) tests, leaving 29 questions for each of these tests.

After applying the selection criteria for questions, the two-parameter logistic (2-PL) IRT model was applied to each grade, subject, and year combination. This type of IRT model allows each question to have its own value of difficulty and discrimination (that is, strength of relationship to students' standardized scores). Each student's standardized score was also captured and available for further analyses. All models were fitted using the mirt package in R 3.5.0.

Figure D2 shows the strength of the relationship between students' performance on tests across 2017 and 2018 for each grade level (looking within a given column) and subject (looking within a given row) combination. Correlation coefficients were computed using the IRT estimated scale scores (blue bars within panels) and number correct total scores (green bars within panels). In nearly all cases, students' 2017 and 2018 performance were positively correlated (that is, high scores are associated with high scores) and moderately strong. To our surprise, however, the strength of the correlations were quite similar for the scale scores and total scores. Since IRT models incorporate the degree of precision from questions into the scale score, it was expected that the scale scores would be a more reliable estimate of students' performance than the more coarse estimate using the total score.

The internal test reliability (that is, the coefficient alpha) for the 2017 SMTs ranges from 0.73 to 0.90 and from 0.71 to 0.92 for 2018 SMTs (see Table D2). These estimates include all questions scored as correct or incorrect, without deleting weak questions. Table D3 shows the median standard errors for the IRT estimated scale scores across grade levels and years. These standard errors tend to increase with the grade level, or scaled test scores are less reliable as grade level goes from 2nd to 6th. Nevertheless, the internal reliability for the scaled scores ranges from 0.79 to 0.92, and is suitable for our analyses. The internal test reliability for the scaled scores is estimated as one minus the squared median standard error.

Pearson

Data analysis

Essentially, what we are doing, to address research question 4 is estimating and statistically testing a semi-partial correlation between student level test scores and classroom level observation ratings. More specifically, a multi-level fixed effects model not a traditional Hierarchical Linear Model, but rather a linear model with factors measured at levels different from the outcome variable was used to estimate and test the magnitude of the relationship between MyPedia program implementation (that is, SRS and SIE ratings) and student achievement. This relationship was tested separately for those observation ratings collected in the 2017–18 school year and those collected in the 2018–19 school year. Splitting the analyses this way allows the relationship to be tested across two time frames. This has a practical significance: the test scores are aligned to content taught in the 2017–18 school year, though the tests were administered in the first quarter of the 2018–19 school year.

The SPSS v25 Generalized Estimating Equations software was used for analyses. To assess the relationship between classroom observation ratings and test scores, it was necessary to have both 2017 and 2018 SMT scores for a student on each of the three subject areas (English, math, and science), as well as observation ratings for that student's classroom on the subject area tested (see Tables F1 through F4).

Schools were set as the independent unit or subject of analysis using an empirical standard error formulation with a naive or independent model based estimator (that is, a sandwich estimator). Since the standard errors are estimated using an empirical estimator sandwiched between two naive, or model-based, estimators, there are no variance parameters to estimate other than the usual error or residual variance. This method, though usually conservative, is very general, consistent, and robust to model misspecification. One need only correctly specify the independent units, in this case the schools. The standard error estimates remain consistent even though incomplete (time points) data arising from different nested levels of measurement is combined into one statistical model. The specific formulation used for the standard errors adjusts for the number of parameters estimated in the statistical model so they are also unbiased in finite samples.

2017 SMT scores were entered into the model as a covariate to adjust the relationship for prior student achievement. Grade level and subject area (with interaction) were entered into the model to equate 2018 SMT scores, already on a standardized scale. Lastly, mean SRS or SIE (across observations observed) for each case was entered into the model to estimate the correlation between SRS/SIE rating and 2018 SMT performance. The case for SRS/SIE is mean rating across quarters observed for each school–subject–grade combination. We periodically refer to this school–subject–grade combination as a classroom, a lesson or an observation.

Pearson

The structural portion of the statistical model may be defined as:

2018 SMT scaled score (s,i,g,c) = intercept (not random) +
2017 SMT scaled score (s,i,g,c) * B1 (correlation coefficient) +
average observation rating (r,y-s,g,c) * B2 (semi-partial relationship) +
content area (c) * **B3** + grade level (g) * **B4** + content by grade (g,c) * **B5** +
residual (s,i,g,c).

For the sake of estimating the fixed parameters, the residuals are assumed to be independently identically normally distributed (so residuals ~ iid N(0,V)).

The design is unbalanced and the factors are fixed and are defined and indexed as follows:

- schools are independent units or subjects (indexed as S)
- teachers' classroom lesson observation ratings (continuous)
- SRS or SIE (indexed as R)
- averaged across 2017–18 or first two quarters of the 2018–19 school year (indexed as Y)
- type of rating and school year are constants in each model as the analyses are split for rating type by school year (see Table DA1)
- observation ratings are specific to lessons at a grade level in a subject area (indexed as G,C)

- students (indexed as I)
- repeated testing each year on three subject/content areas (indexed as C)
- in a grade level each school year (indexed G)
- repeated tests across years
- 2017 SMT → prior achievement (continuous)
- 2018 SMT → outcome measure of interest (continuous)

- content area, grade level and their interaction terms are coded orthogonally

When SMT scores for one of the years were missing, multiple imputation (MI) was used (m = 10) to estimate the missing test scores. The results of the analyses are pooled across the 10 complete datasets in an attempt to give the most accurate results. MI was chosen to impute missing test scores because it is widely seen as a state of the art technique for that purpose. MI is recognized by the What Works Clearinghouse (What Works Clearinghouse Standards Handbook Version 4.0).

Pearson

The imputation model included:

- grade level and subject area, to distinguish specific SMTs
- school, to account for school differences influencing achievement
- 2017 average SRS and SIE
- 2017 and 2018 SMT score

Note that only missing test scores were imputed, not SRS or SIE ratings. The research team felt imputing observation ratings would be less productive as classrooms were not observed consistently and there are no other mutually exclusive endogenous measures of MyPedia implementation to include in the model (recall that observers rate both SRS and SIE on the same criteria).

The imputation was performed across the entire sample of test scores after IRT scaling. The full imputed sample (n = 16,275) now includes students with test scores for both years (note that this is not the analytic sample). Twenty-three percent of the 2018 test scores were imputed (missing = 3,738), while 39.5% of the 2017 test scores were imputed (missing = 6,431). Appendix G shows simple descriptive statistics (count, mean, standard deviation, minimum, maximum) in contrast for the original sample, the 10 samples of imputed values, and the 10 fully imputed samples.
Table 8 shows the counts for the listwise and pooled analytic samples.

As with the wider statistical analyses, the multiple imputation was completed using SPSS v25. The fully conditional iterative Markov chain Monte Carlo (MCMC) method for non-monotone missing patterns was used. Imputed test scores were restricted to fall within the range -4 to 4 as the scores were scaled, using IRT, to fit a standard normal distribution. All imputed values fell within this range. Interactions between the factors were not included in the imputation model.

For each variable, at each iteration, the procedure fits a univariate model using all other variables as predictors. Ten new datasets are created as conditional values are “imputed“ into the missing cell, assuming the model residuals are normally distributed (that is, imputed(i) = bX + r(i)). (In theory, these values are drawn from a joint distribution. In practice, the procedure amounts to repeated regression imputation.) The iterations continue, for a maximum of 10, until “convergence” is achieved — that is, until the estimates of the model parameters change very little.

For the imputed test scores to be unbiased estimates for the missing scores, the missing scores must be missing at random (MAR). Theoretically this means that the cause for the missing scores has nothing to do with the missing scores themselves, but all to do with the data that is observed. In practice, this means that after adjusting for the factors included in the imputation model, it is reasonable to assume that the systematic causes of missing scores are accounted for. In our case we could feel confident in the MAR assumption if we can account for achievement related factors for those students with missing scores, as achievement gives rise to the test scores.

Pearson

In this imputation model, we have the best predictor for the missing test score — another student test score in the same subject — along with reliable measures of the classroom environment. If test scores tend to be missing for lower or higher achieving students, this situation is covered by the imputation model. If scores tend to be missing for certain grade levels or subject areas, this situation is also covered. Lastly, if scores tend to be missing for classes experiencing higher or lower MyPedia quality instruction, that situation is covered. In this model, with the strong relevant achievement related predictors, the research team feels confident that the systemic causes of missing test scores are accounted for.

Table 8: Counts for schools and cases in analytic samples

Predictor	Year	Imputation	Number of schools	Number of cases
School Rating System	2017	Listwise	19	1,307
		Pooled	19	3,304
	2018	Listwise	21	2,949
		Pooled	21	7,395
Student Impact Evaluation	2017	Listwise	13	680
		Pooled	13	1,724
	2018	Listwise	21	2,949
		Pooled	21	7,395

Note: Cases constitute instances where there is a classroom observation for the same subject area and grade level tested.

Pearson

Results

The participating schools (n = 74) used the MyPedia program across all elementary grades and course subject areas as their primary curriculum, student instruction and progress monitoring plan, and professional development system. The myriad of classroom observations (557 observations, 161 visits, 74 schools), spread across the 2017–18 school year and the first half of the 2018–19 school year, indicate that teachers are employing the MyPedia instructional model with fidelity and confidence.

How are MyPedia core components implemented in classrooms by teachers and students?
When averaged across observed classrooms and all criteria, teaching and learning classroom (SRS) ratings increased steadily across all four quarters in 2017–18 and into quarter one of 2018–19, and moved from below to above the "meets standards" cutoff line in quarter three (see Figure E1). The increase in the average rating was statistically significant from quarter to quarter, leveling off in quarter 2 of 2018. The average rating was statistically greater than the "meets standards" cutoff from quarter 4 of 2017 until the last observation period in quarter 2 of 2018 (see Table E1 and Table E2).

Teacher confidence in instructing with MyPedia (see Figure E2) was also observed to increase steadily across all four quarters in 2017–18 and into quarter one of 2018–19 before leveling off, landing confidence squarely in the high–moderate range (that is, 6.5 on the 10 point scale, where the moderate range is 4–7). The major increase from quarter 1–2 to quarter 3 in 2017 was statistically significant.

Have classroom practices changed while implementing MyPedia?
The expected impact (SIE) from crucial instruction related criteria also rose each quarter in 2017–18 and into quarter one of 2018–19, finally breaching the cutoff line for expecting a positive learner impact in quarter 4 of 2017–18. The increase was statistically significant from quarter to quarter, leveling off in quarter 2 of 2018. The average rating was statistically greater than the "expected positive impact" cutoff in quarters 1 and 2 of 2018.

Along with the average SRS and SIE ratings increasing each quarter from 2017 to 2018, so did the proportion of teachers observed to be meeting standards (i.e., average at or above 2.0 across all criteria for both SRS and SIE). The percentage of teachers whose average MyPedia teaching quality rating (SRS) indicated they were "meeting standards" increased from 50% in 2017 to 67% in 2018, maxing out at 77% in quarter 1 (see Table E3). Likewise, the percentage of teachers whose average SIE rating indicated they were likely to have a positive impact on learners increased from 36% in 2017 to 72% in 2018 (see Table E4). Lastly, 18% of teachers were rated "very confident" in 2017, whereas 27% were very confident in 2018 (see Table E5).

Pearson

What are teachers' and parents' opinions of MyPedia?
After some time to become familiar with MyPedia, a representative sample of both teachers and parents were surveyed as to their opinions about the program. The results presented for these surveys include all responses, not just those for the schools in this study. See Table 7 for a breakdown of the respondents. Teachers and parents had high opinions of MyPedia in 2017 and again in 2018. When asked how they would rate the overall quality of MyPedia as a solution for their school, 95% of teachers (2017 n = 181, 2018 n = 165) rated MyPedia as either "good", "very good", or "excellent", and 93% of parents (2017 n = 385, 2018 n = 179) rated MyPedia as either "good", "very good", or "excellent".

When delving into the reasons teachers gave for their overall opinion, comments regarding students' responses to the program and its materials were most positive. In addition, 94% of the 2017 teachers and 90% of the 2018 teachers agreed MyPedia contributes to students' positive attitudes towards learning and 77% of comments regarding improvements to MyPedia (made 2016–2017) were positive.

Lastly, combining the 2017 and 2018 teacher responses to 13 survey questions regarding the perceived value of MyPedia as a solution, the average rating was 1.34 on a scale where a -2 indicates a response of "strongly disagree"/"not at all likely" and +2 indicates a response of "strongly agree"/"very good or excellent"/"very likely or extremely likely". This result demonstrates the overall positive attitude teachers had toward the program.

Does student academic improvement vary with classroom implementation?
Given that the data indicates a change in classroom practice for the MyPedia teachers, we now look to see if there is evidence that the change predicts student achievement. Appendix F (see pooled multiple imputation results) shows the results for the statistical modeling and the Data analysis section details the process.

Averaged across the students from the 19 schools with 2017 and 2018 test scores and corresponding 2017–18 teacher observations, a one point increase in myPedia teaching quality rating (SRS) is related to a 0.44 standard deviation increase (that is, 17 percentile points) in student test scores. Similarly, a one point increase in expected teacher impact rating (SIE) is related to a 0.71 standard deviation increase (that is, 26 percentile points) in student test scores (in 13 schools). No statistically significant relationship was seen between 2018–19 classroom observation ratings and student test scores. This is not surprising as the test is aligned to content covered in the previous grade.

It should be kept in mind that these relationships have been adjusted for the students' achievement measured in 2017. This helps mitigate any bump in the relationship from better students being clustered with certain teachers. Also, the ratings associated with each student's test scores are averaged across the school year for all available observations. This means that the relationship does not track the changes across time. Lastly, it follows from how the rating system was designed that a one point increase in average rating could be affected by modestly moving from meeting the lowest standards across the 20 criteria to meeting the minimally required standards.

Pearson

Table 9: 2017 SRS and SIE listwise and multiple imputation pooled semi-partial correlations

Predictor	Imputation	Number of schools	Number of cases	Effect size	Std. error	P-value
School Rating System	Listwise	19	1307	0.514	0.1744	0.003
	Pooled	19	3304	0.435	0.1386	0.002
Student Impact Evaluation	Listwise	13	680	0.792	0.3174	0.013
	Pooled	13	1724	0.708	0.2271	0.002

Table 10: 2018 SRS and SIE listwise and multiple imputation pooled semi-partial correlations

Predictor	Imputation	Number of schools	Number of cases	Effect size	Std. error	P-value
School Rating System	Listwise	21	2949	0.056	0.1748	0.750
	Pooled	21	7395	0.161	0.1594	0.313
Student Impact Evaluation	Listwise	21	2949	0.065	0.1350	0.629
	Pooled	21	7395	0.191	0.1403	0.175

Pearson

Efficacy statements

In the context of this study, we can make the following correlational statements about the efficacy of MyPedia.

Higher observed MyPedia teaching quality and expected teacher impact ratings are associated with better students' summative scores.

- A one point increase in the MyPedia teaching quality rating is related to a 0.44 standard deviation increase (i.e., 17 percentile points) in students' summative test scores.
- A one point increase in MyPedia teacher impact rating is related to a 0.71 standard deviation increase (i.e., 26 percentile points) in students' summative test scores.

This finding addresses research question 4, Does student academic improvement vary with classroom implementation? In the same context, we can also make several descriptive statements about the efficacy of MyPedia. These statements are as follows.

Findings related to research question 1

Overall, teachers and parents had high opinions of MyPedia in 2017 and 2018.

- When asked how they would rate the overall quality of MyPedia as a solution for their school, 95% of teachers (2017 = 169/181, 2018 = 158/165) and 93% of parents (2017 = 335/368, 2018 = 173/179) rated MyPedia as either Good, Very Good or Excellent.
- 94% (170/181) of the 2017 and 90% (149/165) of the 2018 teachers agreed MyPedia contributes to students' positive attitude toward learning.

Findings related to research question 2

Averaged across all observed classrooms, overall MyPedia teaching quality ratings increased each quarter from 2017 to 2018 moving from "does not meet standards" to "meets standards."

- The percentage of teachers whose average MyPedia teaching quality rating indicated they were "meeting standards" increased from 50% in quarters 1 and 2 of 2017 to 67% in quarter 2 of 2018

MyPedia teacher confidence rating increased each quarter between 2017 and 2018.

- 18% of teachers were rated "very confident" in quarters 1 and 2 of 2017, whereas 27% were "very confident" in quarter 2 of 2018.

Findings related to research question 3

Averaged MyPedia teacher impact ratings increased each quarter from 2017 to 2018.

The percentage of teachers whose average MyPedia teacher impact rating indicated they were "likely to have a positive impact on learners" increased from 36% in quarters 1 and 2 of 2017 to 72% in quarter 2 of 2018.

Pearson

Discussion

Limitations of the study

The bulk of analyses in this report depend on expert ratings of classroom practice. As such, the veracity of the results is limited by the accuracy of these ratings. The results do, however, depend on a large sample of raters (60) across many school visits with observed lessons (557 observations, 161 visits). The expert raters are also the Academic Relationship Managers (ARMs) that maintain a relationship with the schools and provide professional development, so an accurate assessment of classroom implementation is their day to day responsibility. The ARMs are also extensively trained, supported, and supervised. For these reasons the research team feels the ARMs are the best people to provide accurate ratings.

The composite SRS and SIE ratings demonstrated little measurement error and correlated well with each other, and the individual criteria ratings correlated well with their composite rating. These ratings also predict student achievement. The research team was not, however, able to collect information to test rater agreement. This was because of limitations of how the ARMs support the MyPedia schools in the field. The absence of an estimation of the homogeneity among raters is a limitation of the study. It should be noted here that the extent to which the ratings are not consistent across raters, and the measurement error that inconsistency introduces into the results, would be expected to negatively bias the relationship between implementation/MyPedia teaching quality and student achievement. That is, if the rater agreement is low, the estimates of the relationship reported in the efficacy statements will be conservative.

Another limitation is that we could not observe classrooms from every school during every quarter across the school year. In fact, 56 percent (28/50) of the original schools randomly selected to participate in the study were not included in the final data analysis for the relationship between implementation and student performance. This attrition of schools is because of three factors: schools not continuing in the study, missing test data, or no available observation data. These schools did not have both classroom ratings and two years of student test data. All available complete cases, and those that could be imputed, were included in the analyses. Results were also provided for both complete cases (that is, listwise deletion) and the imputed sample.

The attrition, though large, should not be a serious threat to the (internal) validity of the findings because it does not change the relationship as long as there still remained adequate variation in the classroom ratings and student test scores to estimate the relationship. That is, so long as there is enough information throughout the sample space to estimate the parameter with the required precision, it does not matter whom you sample. This is not a situation of differential attrition between study groups in a study design that attempts to make causal efficacy statements; they are simply correlational.

Pearson

In addition, we can say that the analytic sample (that group of 22 schools) had similar total and 1st through 5th grade enrollment as the original 50 school sample, and for the schools that had such data, the ratio of 1st through 5th grade students to trained teachers was very similar. The original and analytic samples had the same percentage of CBSE certified schools. Schools tended to have a single computer lab, with only one school (a member of the group of 22) having no lab, and three schools having more than one. Lastly, the yearly fees are known for 14 of the schools in the group of 22. The average fee for attending one of these schools per school year is US$442.

The original sample had 50% of schools located in Tier 3 minor cities and district towns. The analytic sample had more Tier 2 schools, or schools from small cities, and equally fewer Tier 1 (largest cities) and Tier 3 locations. This shift away from Tier 3 in the analytic sample did not seem to mean higher fees, larger enrollments, more trained teachers, nor better access to educational technology. In fact, the locations, fee structure, enrollments, teacher to student ratios, and technology availability are indicative of the wider class of aspiring private schools MyPedia is designed for. See Tables 1–3 for more details.

Lastly, we address two limitations of the statistical models. We were not able to include school specific factors in the statistical models and thus remove the effects of such factors from the estimates of the relationships between implementation/MyPedia teaching quality and student achievement. Stricter controls, including those at the school level, are in mind for further research plans.

Missing student test scores are estimated and imputed to create multiple full analytic samples. These samples are then used in the statistical models and the results are pooled. These procedures are expected to produce unbiased results when the scores are missing at random (that is, the cause for the missing scores has nothing to do with the missing scores themselves). The research team feels confident that these concerns have been addressed in the imputation model to the extent possible. Specifically, the imputation model includes strong relevant achievement related predictors. We can not, however, be certain that the imputation model fully addresses all causes of systematic bias (that is, reasons for missing scores) and thus this remains as a caution when considering the results.

In conjunction with the missing scores in the analytic sample, scores are missing because schools did not fully complete the study tasks. We can not say the schools remaining in the analytic sample are a random sample of schools, and thus cannot say for certain whether the results generalize to the wider population of such schools. We can only say that the remaining schools appear to be indicative of these schools in several relevant ways.

Pearson

Generalizability of the findings

Because the schools participating in the study are schools that adopted MyPedia and the type of schools the program was designed for, the findings should generalize to that population of schools: private schools in India serving middle class families. These schools are mostly located in small cities or district towns, have moderate enrollment and large student-to-teacher ratios, charge modest fees, and have limited but available technology.

Conclusion

The large number of classroom observations showed the percentage of teachers observed to be meeting standards and teaching with quality rose each quarter from 2017 to 2018. When coupled with student test scores, higher MyPedia teaching quality ratings were associated with better student performance. With MyPedia, there may be a real chance to offer middle income Indian families the winning combination of more successful teachers and students. Research is ongoing to replicate these results and offer further insights.

Further research

The results of this study would indicate that the schools, teachers, and learners MyPedia is designed for, both enjoy MyPedia and are able to use the program effectively. Further research is planned to replicate the results with stricter observation and achievement testing protocols. It is hoped that these processes will allow student achievement gain to be reliably measured and tied to classroom practice. It is also the hope of the research team that higher order thinking skills may be measured for the same purpose.

Pearson

References

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Reinhart, and Winston.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5-25.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*, 61-100.

Parruck, R.C., Ghosh, A., Sheppard, S. British Council. (December 2014). *Indian School Education System* [PDF File]. Retrieved from https://www.britishcouncil.in/sites/default/files/indian_school_education_system_-_an_overview_1.pdf

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: Origins and effectiveness*. Colorado Springs, Co: BSCS.

Fonseca, B. A., & Chi, M. T. H. (2011). Instruction based on self-explanation. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of Research on Learning and Instruction* (pp. 296-321). New York, NY: Routledge.

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review, 26*, 9-25.

Hidi, S. (2006). Interest: A unique motivational variable. *Educational Research Review, 1*, 69-82.

Jacobs, H. H. (Ed.) (1989). *Interdisciplinary curriculum: Design and implementation*. Alexandria, VA: Association for Supervision and Curriculum Development.

Kant, A., Kapoor, V., Nagaich, R. (2018, January 23). India's Burgeoning Middle Class. *LiveMint*. Retrieved from https://www.livemint.com/Opinion/TvcFydQcN6KEFkvdW7BprM/Indias-burgeoning-middle-class.html

Krapp, A., Hidi, S., & Renninger, K. A. (1992). Interest, learning, and development. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 3-26). New York: Erlbaum.

Mayer, R. E. (1979). Twenty years of research on advance organizers: Assimilation theory is still the best predictor of results. *Instructional Science, 8*(2), 133-167.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist*, *27*(1), 91-109.

Roschelle, J. (1995). Learning in interactive environments: Prior knowledge and new experience. In J. H. Falk & L. D. Dierking (Eds.), *Public institutions for personal learning: Establishing a Research Agenda* (pp. 37-51). Washington, D.C.: American Association of Museums.

Saha, D. (2017, April 17). In 5 Years, Private Schools Gain 17 Mn Students, Govt Schools Lose 13 Mn. *India Spend*. Retrieved from http://archive.indiaspend.com/cover-story/in-5-years-private-schools-gain-17-mn-students-govt-schools-lose-13-mn-26146

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. P. Mestre (Ed.), *Transfer of Learning From a Modern Multidisciplinary Perspective* (pp. 1-51). Greenwich, CT: Information Age Publishing.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions revisited: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115-163.

Vars, G. F. (1997). Effects of integrative curriculum and instruction. In J. L. Irvin (Ed.), *What current research says to the middle level practitioner* (pp. 179-186). Columbus, OH: National Middle School Association.

van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science, 328*, 453-456.

What Works Clearinghouse Standards Handbook Version 4.0, p39.

William, D. (2011). What is assessment for learning?. *Studies in Educational Evaluation*, *37*(1), 3-14.

Pearson

Appendix A. Classroom instruction rating system

Table A1: School Rating System

Domain 3 criteria informing SRS	Descriptors
All criteria: Generalized ratings	Level 1 — Meets lowest standards of the criteria Level 2 — Meets middle standards of the criteria Level 3 — Meets or exceeds highest standards of the criteria
Criteria 3.1: Classroom management	Level 1 — Learners sit in rows facing the blackboard and teacher instructs the class from a fixed position. Grouping was not done for the class. Level 2 — Learners sit in the group formed by the teacher. Attention to group of active learners. Level 3 — Learners sit and collaborate in groups as per the needs of the classroom activity. Attention to whole class.
Criteria 3.2: Understanding of students	Level 1 — Calls out students by name. Has a general idea of students' learning levels. Level 2 — Calls out students by name. Is aware of learning needs of learners. Level 3 — Calls out students by name. Addresses individual learning needs, learning styles, and strengths of learners.
Criteria 3.3: Understanding of subject	Level 1 — Demonstrates low level of proficiency in concerned subject, evident as difficulty in teaching certain concepts or inability to explain. Level 2 — Demonstrates satisfactory understanding for most of the concepts. Level 3 — Demonstrates mastery level of knowledge and understanding in the concerned subject.
Criteria 3.4: Understanding of teaching methodology/pedagogy	Level 1 — Lack of necessary pedagogical skills is evident. Level 2 — Demonstrates limited pedagogical skills. Uses conventional lecture method to explain most of the concepts. Level 3 — Demonstrates mastery of pedagogy. Uses suitable teaching method and technique for various concepts (may go beyond MyPedia lesson plans).
Criteria 3.5: Teacher's confidence	Level 1 — Lack of confidence is evident during interaction in the classroom. Level 2 — Demonstrates a satisfactory level of confidence during interaction with learners. Level 3 — High level of confidence is evident throughout the teaching period.

Criteria 3.6: Use of MyPedia teaching/lesson plan	Level 1 — Teaches lesson directly from the textbook. Occasionally refers to the teaching plan. Level 2 — Follows most of the components of the teaching plan, like teaching and assessment strategies and recommended teaching–learning material. Level 3 — Consults lesson plan throughout the period. Has own lesson plan with added value. Connects learning with intermediate context and environment.
Criteria 3.7: Use of MyPedia textbooks	Level 1 — Uses textbooks for chalk and talk or chapter reading method; skips suggested activities. Level 2 — Follows activities suggested in the textbook, but not in an effective manner that promotes learning. Level 3 — Delivers textbook content as per the lesson plan, or even better than that.
Criteria 3.8: Use of MyPedia workbooks	Level 1 — Teacher solves the workbook tasks or questions, or writes on board and learners copy. Level 2 — Teacher discusses worksheet and students solve. Teacher does not consider individual differences. Level 3 — Teacher assigns worksheets as per individual differences and provides support when needed.
Criteria 3.9: Questioning during teaching	Level 1 — Rarely asks questions. If asked, repeats sentences from the book, or chapter end. Level 2 — Asks few explaining questions to 2–3 students, after completing the topic/lesson. Level 3 — Asks questions to encourage discussion and thinking, whenever possible during teaching.
Criteria 3.10: Feedback during teaching	Level 1 — Rarely provides any feedback. Level 2 — Students receive very general feedback like right/wrong, good/bad. Level 3 — Students receive specific feedback during learning, like what they should do and how.
Criteria 3.11: Connecting to previous knowledge	Level 1 — Makes very limited effort to relate to what learners know or have experienced. Level 2 — Asks questions to gauge what they know, but fails to relate to classroom activities. Students are not given opportunities to share their experiences.

	Level 3 — Students are given opportunities to share what they know and have experienced. Class builds on this.
Criteria 3.12: Quality of classwork/homework assignment	Level 1 — Asks students to do some marked bookend questions. Level 2 — Discusses bookend questions and then asks students to write them in notebook. Level 3 — Assigns higher order thinking questions/tasks. Homework/assignments encourage creativity and innovation in the students.
Criteria 3.13: Use of TLM	Level 1 — Mostly textbook and blackboard. Level 2 — Reference books or materials, charts, maps, models, digital learning kits, local resources; uses science, mathematics, and language kits/laboratories as and when appropriate. Level 3 — Integrates the use of TLM, local community resources, ICT support material, laboratories, library, etc. with the lessons appropriately.
Criteria 3.14: Classroom assessments	Level 1 — Uses checkpoints as question–answer (teacher is the one who asks). Level 2 — Uses checkpoints as tool for frequent check of understanding. Level 3 — Checkpoints are used as a tool to check level of learning and adjust teaching to promote learning.
Criteria 3.15: Use of assessments/sharing feedback/reporting	Level 1 — Generates basic report cards and shares with parents. Level 2 — Generates value added, skill-based report cards and shares with students and parents. Level 3 — Level 2 + plans for remedial teaching based on Level 2 and shares actionable reports with students and parents.
Criteria 3.16: Use of in class digital	Level 1 — Displays digital materials occasionally, as per convenience (as a separate activity). Level 2 — Displays digital materials after completing the content as per textbook (as separate period). Level 3 — Integrates digital features as per lesson plan. Facilitates the digital features well — stops at regular intervals and asks questions on prediction/understanding.
Criteria 3.17: Effort to ensure learners' punctuality	Level 1 — Teachers mark/record attendance of learners regularly, identify learners who are frequently absent or not punctual. Not concerned about punctuality of work. Level 2 — Demonstrates concern about students' physical presence/absence in the classroom, as well as punctuality of work.

	Level 3 — Motivates students to be regular and punctual in classroom and school.
Criteria 3.18: Involving students in learning	Level 1 — Learners listen quietly to teachers without much interaction (teacher talk 75% and more). Level 2 — A few learners (usually rank holders) actively participate in the class discussion (teacher talk 50–75%). Level 3 — All learners participate actively in discussion and interact with teachers and peers (teacher talk < 50%).
Criteria 3.19: Participation in school activities	Level 1 — Few students (usually same) participate in mandated school functions and co-scholastic activities. Level 2 — School organizes a variety of co-scholastic activities and cultural programmes. Teachers motivate learners to actively participate in these activities and a large number of students participate. Level 3 — School identifies talent of learners in different co-scholastic areas and provides them suitable experience and opportunities. All learners take interest and participate enthusiastically in various school functions and co-scholastic activities.
Criteria 3.20: Maintenance of students' progress	Level 1 — School maintains records of learners' progress data as per norms and mandate. Level 2 — School continuously tracks individual learners' progress against curricular expectations (scholastic and co-scholastic); creates a cumulative database and updates it periodically. Level 3 — School tracks individual learners' progress, keeping in mind learners' differential pace of learning; analyzes the cumulative database to identify progress patterns and trends for classes and groups of learners. Shares with stakeholders.

Pearson

Table A2: Student Impact Expectation (SIE) rating system

Domain 3 criteria informing SIE	Ratings
This teacher's ***classroom management*** was observed at a level where you expect it will:	Level 1 — Not have a positive impact on the learners Level 2 — Have a positive impact on the learners Level 3 — Have a greater than expected impact on the learners
This teacher's ***understanding of subject*** was observed at a level where you expect it will:	
This teacher's ***use of teacher/lesson plan*** was observed at a level where you expect it will:	
This teacher's ***use of MyPedia Textbooks and Workbooks*** was observed at a level where you expect it will:	
This teacher's ***use of teaching methods and techniques*** was observed at a level where you expect it will:	
This teacher's use of 3.10 (feedback during teaching), 3.14 (classroom assessments), 3.15 (use of assessments/sharing feedback/reporting), and 3.20 (maintenance of students' progress) was observed at a level where you expect it will:	
This teacher's ***use of quality assignments*** was observed at a level where you expect it will:	
This teacher's ***use of connecting to previous knowledge*** was observed at a level where you expect it will:	
This teacher's ***use of TLM (manipulative/ancillary/supplemental)*** was observed at a level where you expect it will:	
This teacher's ***use of digital features of MyPedia in classroom*** was observed at a level where you expect it will:	
This teacher's ***use of involving students in learning*** was observed at a level where you expect it will:	

3. The purpose of the SIE ratings is to better correlate/predict student impact from implementation. The criteria from the SRS Framework inform the SIE rating. The construct (i.e., impact on the learner) and the scale (i.e., expectation categories) are the same across all the criteria. The score for each criterion and the overall score mean the same thing, which is that it speaks to the expected impact on the learner. We would expect a higher score to indicate a greater impact on the student.

Pearson

Table A3: Teacher Confidence rating system

Domain 3 criteria informing Teacher Confidence	Ratings
Please indicate the teacher's response as to their confidence ***in using the questions to develop higher order thinking skills***.	Scored on a scale of 0–10 0: Not at all 1–3: Very little confidence 4–7: Moderate confidence 8–10: Very confident
Please indicate teacher confidence on ***instruction methods and techniques***.	
Please indicate teacher confidence in ***checking for understanding***.	
Please indicate teacher confidence when ***using checkpoints***.	
Please indicate teacher confidence when ***connecting to previous knowledge***.	
Please indicate teacher confidence when ***involving students in learning***.	

Pearson

Appendix B. Classroom rating statistics

Table B1: School Rating System frequencies

Criteria	Rating 0	Rating 1	Rating 2	Rating 3	N/A or not used	Blank
	Frequency (Percent)	Frequency (Percent)	Frequency (Percent)	Frequency (Percent)	Frequency (Percent)	Frequency (Percent)
Classroom management	2 (0.4)	272 (50.3)	175 (32.3)	92 (17.0)	13 (2.4)	3 (0.5)
Understanding of students	1 (0.2)	132 (24.1)	304 (55.5)	111 (20.3)	8 (1.5)	1 (0.2)
Understanding of subject	1 (0.2)	36 (6.5)	336 (61.0)	178 (32.3)	6 (1.0)	-
Understanding of teaching methodology pedagogy	6 (1.1)	70 (12.7)	341 (61.8)	135 (24.5)	4 (0.7)	1 (0.2)
Teacher confidence	3 (0.5)	47 (8.5)	328 (59.1)	177 (31.9)	2 (0.4)	-
MyPedia teaching lesson plan	2 (0.4)	136 (25.1)	291 (53.8)	112 (20.7)	15 (2.7)	1 (0.2)
MyPedia textbooks	-	96 (18.0)	262 (49.1)	176 (33.3)	20 (3.6)	3 (0.5)
MyPedia workbooks	-	88 (20.2)	236 (54.3)	111 (25.5)	116 (20.9)	6 (1.1)
Questioning during teaching	-	67 (12.3)	246 (45.1)	232 (42.6)	12 (2.2)	-
Feedback during teaching	-	76 (14.3)	296 (55.8)	158 (29.8)	27 (4.8)	-
Connecting to previous knowledge	-	109 (20.3)	246 (45.7)	183 (34.0)	18 (3.3)	1 (0.2)

Quality of classwork homework assignment	-	84 (18.1)	292 (63.1)	87 (18.8)	88 (15.8)	6 (1.1)
Use of TLM	12 (2.4)	222 (45.3)	215 (43.9)	41 (8.4)	63 (11.3)	4 (0.7)
Classroom assessments	2 (0.4)	126 (25.9)	275 (56.6)	83 (17.1)	61 (11.0)	10 (1.8)
Assessments sharing feedback reporting	-	142 (50.0)	120 (42.3)	22 (7.7)	260 (46.7)	13 (2.3)
In-class digital	10 (3.8)	94 (35.5)	114 (43.0)	47 (17.7)	245 (44.0)	47 (8.4)
Effort to ensure learner punctuality	-	80 (24.5)	177 (54.3)	69 (21.2)	195 (35.0)	36 (6.5)
Involving students in learning	-	99 (18.5)	250 (46.7)	186 (34.8)	19 (3.4)	3 (0.5)
Participation in school activities	-	50 (14.3)	233 (66.6)	67 (19.1)	167 (30.0)	40 (7.2)
Maintenance of students' progress	1 (0.3)	135 (42.2)	148 (46.3)	36 (11.3)	221 (39.7)	16 (2.9)

Pearson

Table B2: Student Impact Expectation rating system frequencies

Criteria	Rating 1	Rating 2	Rating 3	N/A or not used	Blank
	Frequency (Percent)	Frequency (Percent)	Frequency (Percent)	Frequency (Percent)	Frequency (Percent)
Classroom management	111 (24.8)	279 (62.3)	58 (12.9)	4 (0.7)	105 (18.9)
Teacher understanding of subject	38 (8.4)	339 (75.2)	74 (16.4)	1 (0.2)	105 (18.9)
Use of teacher lesson plan	82 (18.6)	290 (65.8)	69 (15.6)	10 (1.8)	106 (19.0)
Use of MyPedia textbooks and workbooks	61 (16.5)	260 (70.5)	48 (13.0)	7 (1.3)	181 (32.5)
Use of teaching methods and techniques	54 (12.1)	322 (72.4)	69 (15.5)	6 (1.1)	106 (19.0)
Monitor student progress	57 (18.8)	223 (73.4)	24 (7.9)	134 (24.1)	119 (21.4)
Use of quality of assignments	53 (14.3)	276 (74.6)	41 (11.1)	29 (5.2)	158 (28.4)
Use of connecting to previous knowledge	67 (15.6)	294 (68.4)	69 (16.0)	12 (2.2)	115 (20.6)
Use of manipulative ancillary supplementals	88 (23.9)	248 (67.4)	32 (8.7)	57 (10.2)	132 (23.7)
Use of MyPedia digital features in classroom	67 (33.0)	118 (58.1)	18 (8.9)	120 (21.5)	234 (42.0)
Use of involving students in learning	73 (16.9)	292 (67.7)	66 (15.3)	6 (1.1)	120 (21.5)

Pearson

Table B3: Teacher Confidence rating system frequencies

	Rating 1–3	Rating 4–7	Rating 8–10	N/A or not used	Blank
Criteria	Frequency (Percent)	Frequency (Percent)	Frequency (Percent)	Frequency (Percent)	Frequency (Percent)
Using the questions to develop HOTSs	73 (17.4)	245 (58.5)	101 (24.1)	22 (3.9)	116 (20.8)
Instruction methods and techniques	51 (11.4)	265 (59.3)	131 (29.3)	3 (0.5)	107 (19.2)
Checking for understanding	54 (12.2)	259 (58.6)	129 (29.2)	3 (0.5)	112 (20.1)
Using checkpoints	51 (12.4)	236 (57.6)	123 (30.0)	12 (2.2)	135 (24.2)
Connecting to previous knowledge	60 (13.8)	255 (58.8)	119 (27.4)	4 (0.7)	119 (27.4)
Involving students in learning	50 (11.5)	254 (58.3)	132 (30.3)	2 (0.4)	119 (21.4)

Pearson

Appendix C. Teacher and parent survey summaries

2017 and 2018 parent survey questions summary

- Identifier information
- Overall quality of MyPedia
  - Rating and reason for rating
- Rating of MyPedia value
- Describe the price charged for MyPedia
- Do you agree with the following statements about MyPedia
- MyPedia is a one-point innovative learning solution
- MyPedia is at an appropriate level for your child
- MyPedia promotes high level of thinking among your child
- MyPedia helps your child enjoy learning
- MyPedia promotes independent learning habits among your child
- MyPedia helps improve children's attitude towards learning
- MyPedia improves child's engagement with learning (keeps them interested in learning)
- MyPedia enhances ability to relate with real-life scenarios
- How likely are you to suggest the school to continue MyPedia for your child in the future classes
- How likely are you to recommend MyPedia to other parents/guardians
- Clarity of text printed in both the textbook and the workbook
- Textbook
- Durability and attractiveness
- Level of engagement of your child with both the textbook and the workbook
- Adequacy of the content in terms of its length and depth in explaining concepts
- Relatedness of textbook content with the digital content (Home App)
- Practice exercises in the workbook
- Variety of and quality of
- Overall quality of MyPedia textbooks and workbooks
- Digital content
- Quality and attractiveness
- Ease of understanding the digital content
- Appropriateness as per your child's age
- Level of engagement of your child
- Ease of understanding the language
- Variety (animations, games, interactive, photographs, drawings, etc.)
- Overall quality
- Home App
- Overall quality
- 9 additional sub-questions

- Question papers
- Ease, length, and alignment with the content taught
- Quality of information in report cards
- Assessments
- Ease and overall quality
- Parent orientation session
- Overall quality of parent orientation session
- 5 additional sub-questions
- How likely are you to recommend MyPedia to others
- Aspects liked about MyPedia
- Suggestions for MyPedia
- Complaints/complaint handling (2018 only)
- Overall quality
- 5 additional sub-questions
- All things considered, how do you rate MyPedia (2018 only)
- How strongly do you prefer MyPedia rather than any other learning solution (2018 only)

Pearson

2017 and 2018 principal/teacher survey questions summary

- Identifier information
- Overall quality of MyPedia
- Rating and reason for rating
- Rating of MyPedia value
- Describe the price charged for MyPedia
- Do you agree with the following statements about MyPedia
  - Is a market leader among companies providing learning solutions
  - Is a learning solution that provides good value for money
  - Is an innovative solution
  - Is easy to understand and use in classroom
  - Helps in interdisciplinary learning across different subjects
  - Promotes high level of thinking among students
  - Is from a company where people are accessible
  - Has updated content as per latest syllabus or curriculum
  - Is flexible to adjust to school's requirements
  - Is a solution that provides learning oriented products and services
  - Contributes towards students' positive attitude towards learning
  - Helps reduce teachers' time and efforts in preparatory and administrative tasks
- Please state your level of agreement/disagreement with the following statements on MyPedia
  - You feel very loyal to using MyPedia
  - It would matter a lot to you if MyPedia is not available and you have to opt for a solution from another company
- Please state the likelihood of the following statements regarding MyPedia
  - How likely are you to continue using MyPedia in the future
  - How likely are you to recommend MyPedia to your peers in other schools
- Curriculum
  - Overall quality
  - 10 additional sub-questions (2 questions are in 2017 only)
- Print content/material
  - Overall quality
  - 11 additional sub-questions
- Digital content
  - Overall quality
  - 8 additional sub-questions
- Home App
  - Overall quality
  - 6 additional sub-questions
- Assessments
  - Overall quality
  - 13 additional sub-questions (4 questions are in 2017 only)

- Training programs
  - Overall quality
  - 9 additional sub-questions
- Delivery support
  - Overall quality
  - 9 additional sub-questions (2 questions are in 2018 only)
- Sales manager
  - Please let me know which of the following company personnel have you ever interacted with regarding MyPedia
  - Overall quality
  - 8 additional sub-questions
- Academic Relationship Manager (ARM)
  - Please let me know which of the following company personnel have you ever interacted with regarding MyPedia
  - Overall quality
  - 8 additional sub-questions (1 question is in 2017 only)
- Complaint handling
  - Have you had any occasion to complain to the Customer Care Helpline of MyPedia in the past 6 months
  - Overall quality
  - 6 additional sub-questions (1 question is in 2018 only)
- How likely are you to recommend MyPedia to others on a scale of 0 to 10
- All things considered, how do you rate MyPedia (2018 only)
- Other learning solutions
  - 9 sub-questions regarding learning solutions
- Improvements with MyPedia (please provide 2 things that worked well for you in 2018)
- Suggestions with MyPedia
- Would you give Kantar IMRB permission to contact you again in case we need to understand your feedback in further detail
- Would you give Kantar IMRB permission to identify your responses with your name

Pearson

Appendix D. Skills Mapping Test score scaling tables

(Back)

Table D1: Sample sizes for each assessment type, subject, and student grade level

Assessment	Subject	2nd grade	3rd grade	4th grade	5th grade	6th grade
Grade level 2017	English	1,103	874	1,035	301	
Grade level 2017	Mathematics	1,110	885	1,084	272	
Grade level 2017	Science	1,079	808	991	302	
Grade level 2018	English		1,300	1,255	1,232	395
Grade level 2018	Mathematics		1,294	1,255	1,230	395
Grade level 2018	Science		1,301	1,259	1,230	394

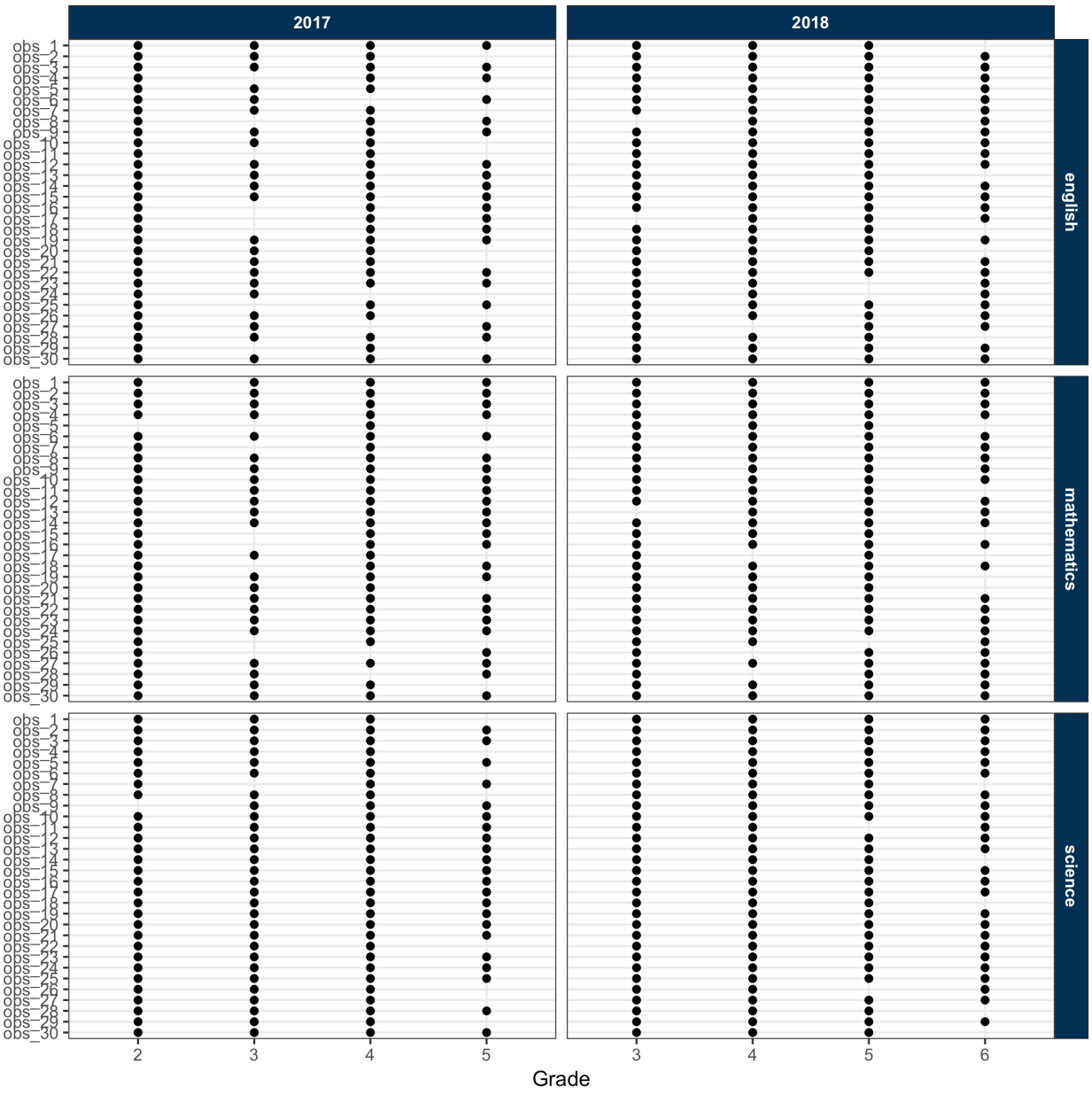


Figure D1: Visual of SMT questions retained after applying selection criteria

Note: Black dots indicate that the item was retained for subsequent analyses.

Figure D2: Comparison of correlation strength by different types of ability estimates

Pearson

Table D2: Estimated internal reliability coefficients (alpha) for questions scored correct or incorrect

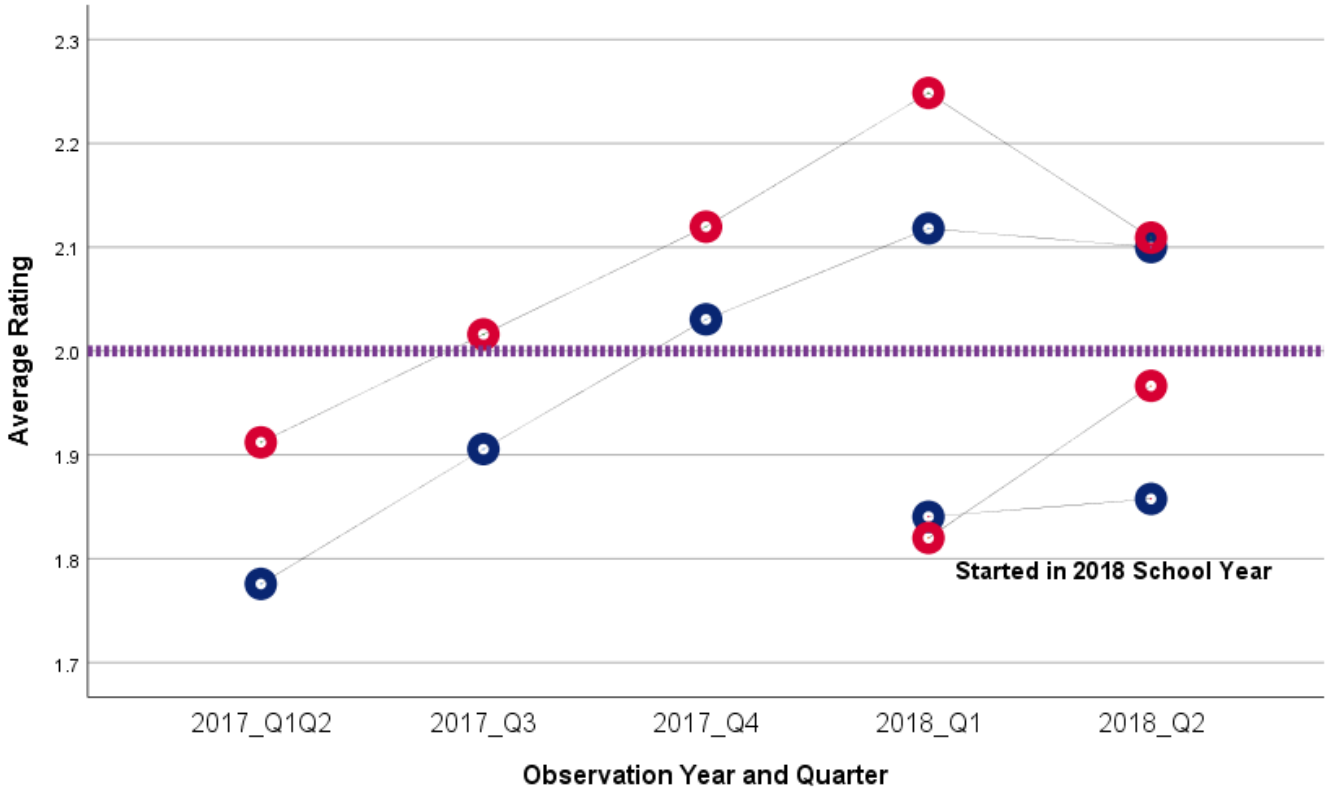
Content	Year	2nd grade	3rd grade	4th grade	5th grade	6th grade
English	2017	0.904	0.749	0.839	0.728	-
	2018	-	0.856	0.865	0.802	0.717
Math	2017	0.891	0.786	0.813	0.803	-
	2018	-	0.886	0.856	0.814	0.708
Science	2017	0.887	0.869	0.810	0.760	-
	2018	-	0.917	0.869	0.795	0.761

Table D3: Median standard errors for scaled scores

Content	Year	2nd grade	3rd grade	4th grade	5th grade	6th grade
English	2017	0.279	0.392	0.353	0.382	-
	2018	-	0.337	0.330	0.409	0.463
Math	2017	0.315	0.370	0.373	0.322	-
	2018	-	0.294	0.314	0.393	0.449
Science	2017	0.341	0.343	0.400	0.370	-
	2018	-	0.275	0.321	0.403	0.447

Pearson

Appendix E. Classroom observation ratings figures



red = SRS, _blue_ = SIE

Figure E1: Average MyPedia classroom teaching and learning observation ratings

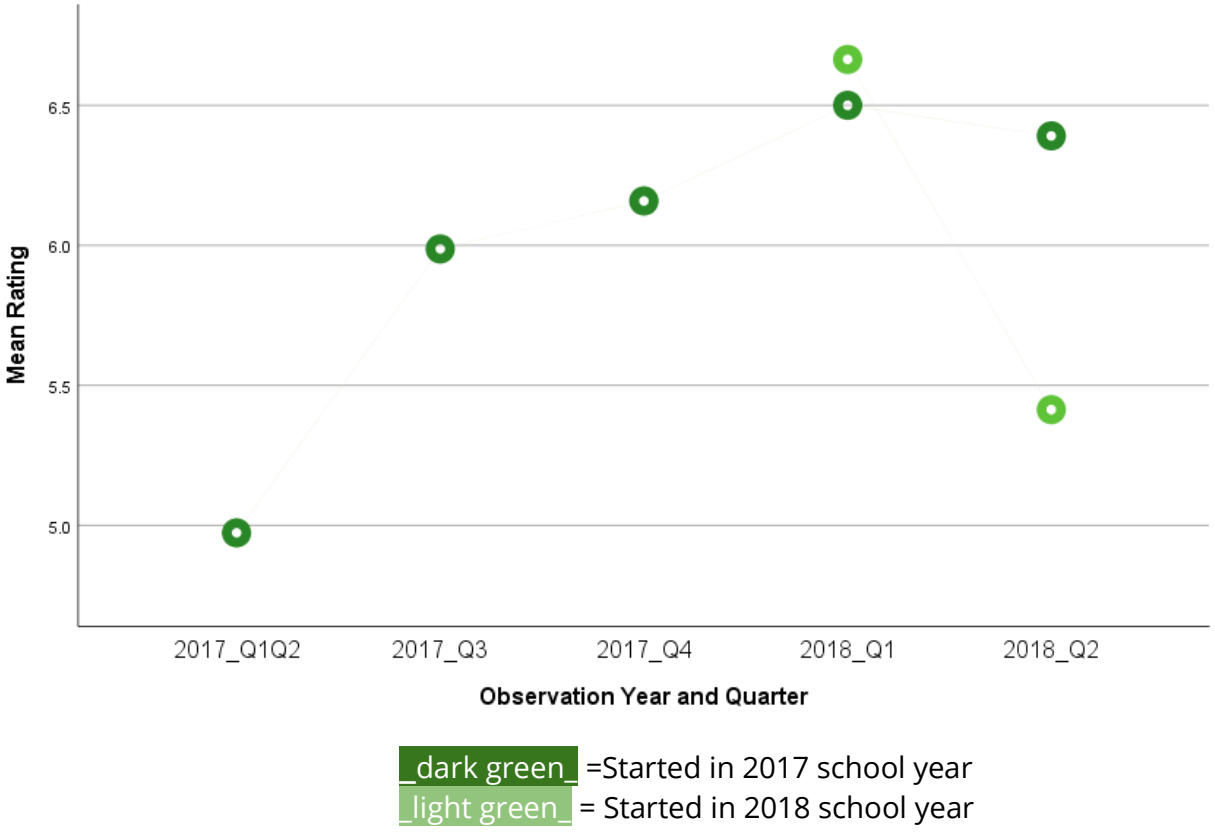


Figure E2: Average MyPedia teacher confidence observation ratings

Pearson

Table E1: Classroom lesson rating criteria from "meets standards" or "expected impact"

Rating type	Year, quarter	Number of criteria	Mean	Std. deviation	Std. error mean
School Rating System (includes 20 criteria)	2017 Q1–2	2,003	-0.09	0.73	0.02
	2017 Q3	1,305	0.02	0.67	0.02
	2017 Q4	643	0.12	0.68	0.03
	2018 Q1	515	0.25	0.65	0.03
	2018 Q2	1960	0.11	0.68	0.02
Student Impact Evaluation (includes 11 criteria)	2017 Q1–2	107	-0.22	0.59	0.06
	2017 Q3	667	-0.09	0.48	0.02
	2017 Q4	328	0.03	0.55	0.03
	2018 Q1	305	0.12	0.53	0.03
	2018 Q2	1120	0.10	0.58	0.02

Note: Only includes observations for schools that started using MyPedia in 2017. All criteria are not always rated for each classroom lesson observation. For the mean, a positive value indicates the mean is above 2.0 on the 3 point scale, or the "meets standards" mark.

Pearson

Table E2: Classroom lesson rating criteria from "meets standards" or "expected impact" T-test tables

Rating type	Year, quarter	t	df	P-value	Mean difference	95% confidence interval of the difference	
						Lower	Upper
School Rating System	2017 Q1–2	-5.388	2,002	0.000	-0.09	-0.12	-0.06
	2017 Q3	0.867	1,304	0.386	0.02	-0.02	0.05
	2017 Q4	4.474	642	0.000	0.12	0.07	0.17
	2018 Q1	8.657	514	0.000	0.25	0.19	0.30
	2018 Q2	7.074	1,959	0.000	0.11	0.08	0.14
Student Impact Evaluation	2017 Q1–2	-3.948	106	0.000	-0.22	-0.34	-0.11
	2017 Q3	-5.056	666	0.000	-0.09	-0.13	-0.06
	2017 Q4	1.010	327	0.313	0.03	-0.03	0.09
	2018 Q1	3.927	304	0.000	0.12	0.06	0.18
	2018 Q2	5.798	1,119	0.000	0.10	0.07	0.13

Note: Only includes observations for schools that started using MyPedia in 2017.

Pearson

Table E3: Classroom lesson observations rated on average meeting standards for MyPedia teaching quality (SRS)

Year, quarter		Meets standards		Lessons observed
		No	Yes	
2017 Q1–2	Count	53 (50.5)	52 (49.5)	105
2017 Q3	Count	32 (43.2)	42 (56.8)	74
2017 Q4	Count	14 (37.8)	23 (62.2)	37
2018 Q1	Count	8 (23.5)	26 (76.5)	34
2018 Q2	Count	40 (33.3)	80 (66.7)	120
Total	Count	147 (39.7)	223 (60.3)	370

Note: Only includes observations for schools that started using MyPedia in 2017.

Pearson

Table E4: Classroom lesson observations rated on average meeting standards for Student Impact Expectation (SIE)

Year, quarter		Positive impact		Lessons observed
		No	Yes	
2017 Q1–2	Count	7 (63.6)	4 (36.4)	11
2017 Q3	Count	27 (39.1)	42 (60.9)	69
2017 Q4	Count	10 (30.3)	23 (69.7)	33
2018 Q1	Count	11 (32.4)	23 (67.6)	34
2018 Q2	Count	34 (28.3)	86 (71.7)	120
Total	Count	89 (33.3)	178 (66.7)	267

Note: Only includes observations for schools that started using MyPedia in 2017.

Pearson

Table E5: Classroom lesson observations teachers rated very confident

Year, quarter		Confident		Lessons observed
		No	Yes	
2017 Q1–2	Count	9 (81.8)	2 (18.2)	11
2017 Q3	Count	60 (87.0)	9 (13.0)	69
2017 Q4	Count	27 (81.8)	6 (18.2)	33
2018 Q1	Count	29 (85.3)	5 (14.7)	34
2018 Q2	Count	88 (73.3)	32 (26.7)	120
Total	Count	213 (79.8)	54 (20.2)	267

Note: Only includes observations for schools that started using MyPedia in 2017.

Pearson

Appendix F. Student impact model tables

Table F1: School Rating System observations by quarter in 2017 school year

School ID	2017 SMT month tested	Content and grade level observed by quarter		
		2017 Q1–2	2017 Q3	2017 Q4
NOHAKW	May–June	Eng4, Math5, Sci2	Sci5	-
NOHATN	May–June	Eng2	-	-
NOUPKO	May–June	Sci4	-	-
NOUPSA	May–June	Math2, Sci4	Eng2	-
SOAPRL	May–June	-	Eng3, Math3	-
SOAPSM	May–June	-	Eng3	-
SOKABS	May–June	-	Math3, Sci3	-
SOKACH	May–June	-	-	-
SOKASM	May–June	Math2	-	-
SOKASP	May–June	Eng4	-	Eng2, Sci5
SOKAVI	May–June	Eng5, Math3	Math4	Eng5
SOTAAV	May–June	-	Math4, Math5	Sci4
SOTABD	May–June	Eng3, Math3, Sci3	-	-
SOTASG	May–June	-	-	-
SOTEKP	May–June	-	Sci4	Eng2, Sci4
SOTERA	May–June	Sci2	-	-
SOTESS	May–June	-	-	-

SOTETH	May–June	-	Math3	-
SOTETJ	May–June	Eng3	-	-
WEGUDI	May–June	-	-	Sci5
WEGUSU	May–June	-	-	Eng5, Sci4
WEMAPA	May–June	-	-	Eng2, Math2, Sci2

Notes:
Eng = English, Sci = Science
NOHATN — Math3 was observed in Q1/Q2, but not SMT tested
NOUPSA — Sci5 was observed in Q3, but not SMT tested
SOKACH — Sci5 was observed in Q1/Q2, but not SMT tested
SOTASG — Math5 and Sci5 were observed in Q1/Q2, but not SMT tested
SOTERA — Eng5 was observed in Q1/Q2, but not SMT tested
SOTETH — Math5 and Sci5 were observed, but not SMT tested

Pearson

Table F2: School Rating System observations by quarter in 2018 school year

School ID	2018 SMT month tested	Content and grade level observed by quarter	
		2018 Q1	2018 Q2
NOHAKW	May	Eng5, Math5, Sci3	Eng3, Math6, Sci5
NOHATN	May	Eng3	Sci3
NOUPKO	May	-	Eng4, Eng5, Math4, Math5, Sci4, Sci5
NOUPSA	April	Eng5, Math3, Sci4	Eng4, Math3, Math4, Math5, Sci5
SOAPRL	July	-	Eng4, Math3, Sci3, Sci4, Sci5
SOAPSM	July	-	Eng4, Eng5, Math4, Sci5
SOKABS	June	-	Eng6, Math3, Math4, Math6, Sci4, Sci5
SOKACH	June	-	Eng3, Math4, Sci5
SOKASM	June	Eng3, Math4, Math5, Sci3, Sci4	Eng4, Math4, Sci5
SOKASP	June	Math3, Math4	Eng6, Math3, Math5
SOKAVI	June	Eng3, Math4, Sci5	Eng3, Math3, Math6, Sci4
SOTAAV	June	-	Eng5, Math3, Sci4
SOTABD	June	Eng5, Math5, Sci5	Eng3, Math3, Math5, Sci3
SOTASG	June	Eng3, Eng4, Math5, Sci3	-
SOTEKP	June	-	Eng3, Eng4, Sci4
SOTERA	July	-	Math3, Math5, Sci4
SOTESS	June	-	Eng5, Sci3
SOTETH	June	-	Eng4, Sci3
SOTETJ	June	-	Math4
WEGUDI	June	-	-
WEGUSU	June	-	Eng3, Math4, Math5
WEMAPA	June	-	Eng3, Math3

Notes: Eng = English, Sci = Science
NOHATN — Sci4 was observed in Q2, but not SMT tested

Pearson

Table F3: 2017 Student Impact Expectation observations by quarter in 2017 school year

School ID	2017 SMT month tested	Content and grade level observed by quarter		
		2017 Q1–2	2017 Q3	2017 Q4
NOHAKW	May–June	-	Sci5	-
NOHATN	May–June	-	-	-
NOUPKO	May–June	Sci4	-	-
NOUPSA	May–June	-	Eng2	-
SOAPRL	May–June	-	Eng3, Math3	-
SOAPSM	May–June	-	Eng3	-
SOKABS	May–June	-	Math4, Sci3	-
SOKACH	May–June	-	-	-
SOKASM	May–June	-	-	-
SOKASP	May–June	-	-	Eng2, Sci5
SOKAVI	May–June	-	Math4	Eng5
SOTAAV	May–June	-	Math4, Math5	Sci4
SOTABD	May–June	-	-	-
SOTASG	May–June	-	-	-
SOTEKP	May–June	-	Sci4	Eng2, Sci4
SOTERA	May–June	-	-	-
SOTESS	May–June	-	-	-
SOTETH	May–June	-	Math3	-
SOTETJ	May–June	-	-	-
WEGUDI	May–June	-	-	Sci5
WEGUSU	May–June	-	-	Eng5, Sci4
WEMAPA	May–June	-	-	Eng2, Math2, Sci2

Notes: Eng = English, Sci = Science
NOUPSA — Sci5 was observed in Q3, but not SMT tested

Pearson

Table F4: Student Impact Expectation observations by quarter in 2018 school year

School ID	2018 SMT month tested	Content and grade level observed by quarter	
		2018 Q1	2018 Q2
NOHAKW	May	Eng5, Math5, Sci3	Eng3, Math6, Sci5
NOHATN	May	Eng3	Sci3
NOUPKO	May	-	Eng4, Eng5, Math4, Math5, Sci4, Sci5
NOUPSA	April	Eng5, Math3, Sci4	Eng4, Math3, Math4, Math5, Sci5
SOAPRL	July	-	Eng4, Math3, Sci3, Sci4, Sci5
SOAPSM	July	-	Eng4, Eng5, Math4, Sci5
SOKABS	June	-	Eng6, Math3, Math4, Math6, Sci4, Sci5
SOKACH	June	-	Eng3, Math4, Sci5
SOKASM	June	Eng3, Math4, Math5, Sci3, Sci4	Eng4, Math4, Sci5
SOKASP	June	Math3, Math4	Eng6, Math3, Math5
SOKAVI	June	Eng3, Math4, Sci5	Eng3, Math3, Math6, Sci4
SOTAAV	June	-	Eng5, Math3, Sci4
SOTABD	June	Eng5, Math5, Sci5	Eng3, Math3, Math5, Sci3
SOTASG	June	Eng3, Eng4, Math5, Sci3	-
SOTEKP	June	-	Eng3, Eng4, Sci4
SOTERA	July	-	Math3, Math5, Sci4
SOTESS	June	-	Eng5, Sci3
SOTETH	June	-	Eng4, Sci3
SOTETJ	June	-	Math4
WEGUDI	June	-	-
WEGUSU	June	-	Eng3, Math4, Math5
WEMAPA	June	-	Eng3, Math3

Notes: Eng = English, Sci = Science
NOHATN — Sci4 was observed in Q2, but not SMT tested

Pearson

Table F5: 2017 School Rating System model continuous variable information

Analytic sample	Measurement	N	Mean	Std. deviation
Listwise	SMT scaled score 2018	1,307	0.17	0.95
	SMT scaled score 2017	1,307	0.12	0.95
	School Rating System mean	1,307	2.05	0.37
Pooled (10 imputations)	SMT scaled score 2018	3,304	0.07	
	SMT scaled score 2017	3,304	0.07	
	School Rating System mean	3,304	2.01	

Pearson

Table F6: School Rating System categorical variable information

Analytic sample	Factor	N	Percent
Listwise	English	418	32.0
	Mathematics	398	30.5
	Science	491	37.6
	Total	1,307	100.0
	GRADE 2017 = 2	576	44.1
	GRADE 2017 = 3	184	14.1
	GRADE 2017 = 4	423	32.4
	GRADE 2017 = 5	124	9.5
	Total	1,307	100.0
Pooled (10 imputations)	English	1,034	31.3
	Mathematics	1,016	30.8
	Science	1,254	38.0
	Total	3,304	100.0
	GRADE 2017 = 2	1,424	43.1
	GRADE 2017 = 3	452	13.7
	GRADE 2017 = 4	943	28.5
	GRADE 2017 = 5	485	14.7
	Total	3,304	100.0

Table F7: 2017 School Rating System model correlated data summary

Analytic sample	Number of levels subject effect SchoolID	Number of levels within-subject effect Case_ID	Number of measurements per subject minimum	Number of measurements per subject maximum
Listwise	19	1,307	2	273
Pooled (10 imputations)	19	3,304	22	624

Pearson

Table F8: 2017 School Rating System model listwise parameter estimates

Model effect	B	Std. error	95% Wald confidence interval		Hypothesis test		
			Lower	Upper	Wald Chi-square	df	P-value
Intercept	-1.198	0.3622	-1.908	-0.488	10.934	1	0.001
School Rating System mean	0.514	0.1744	0.172	0.856	8.683	1	0.003
SMT scaled score 2017	0.286	0.0614	0.166	0.407	21.759	1	0.000
CONTENT = English	0.111	0.2311	-0.342	0.564	0.230	1	0.632
CONTENT = Mathematics	0.352	0.1880	-0.017	0.720	3.498	1	0.061
CONTENT = Science	0						
GRADE 2017 = 2	0.247	0.1995	-0.144	0.638	1.530	1	0.216
GRADE 2017 = 3	-0.062	0.2137	-0.481	0.356	0.085	1	0.770
GRADE 2017 = 4	0.354	0.2060	-0.050	0.758	2.952	1	0.086
GRADE 2017 = 5	0						
CONTENT = English * GRADE 2017 = 2	-0.006	0.3061	-0.605	0.594	0.000	1	0.985
CONTENT = English * GRADE 2017 = 3	0.348	0.2515	-0.145	0.841	1.913	1	0.167
CONTENT = English * GRADE 2017 = 4	-0.094	0.2395	-0.563	0.376	0.153	1	0.695
CONTENT = English * GRADE 2017 = 5	0						
CONTENT = Mathematics * GRADE 2017 = 2	-0.152	0.4007	-0.938	0.633	0.144	1	0.704
CONTENT = Mathematics * GRADE 2017 = 3	-0.141	0.1704	-0.475	0.193	0.682	1	0.409
CONTENT = Mathematics * GRADE 2017 = 4	-0.858	0.3820	-1.607	-0.110	5.051	1	0.025
CONTENT = Mathematics * GRADE 2017 = 5	0						
CONTENT = Science * GRADE 2017 = 2	0						
CONTENT = Science * GRADE 2017 = 3	0						
CONTENT = Science * GRADE 2017 = 4	0						
CONTENT = Science * GRADE 2017 = 5	0						
Scale	0.701						

Pearson

Table F9: 2017 School Rating System model pooled parameter estimates

Model effect	B	Std. error	95% Wald confidence interval		P-value	Fraction missing info	Relative increase variance	Relative efficiency
			Lower	Upper				
Intercept	-1.105	0.2926	-1.679	-0.532	0.000	0.043	0.044	0.996
School Rating System mean	0.435	0.1386	0.164	0.707	0.002	0.028	0.029	0.997
SMT scaled score 2017	0.284	0.0407	0.204	0.363	0.000	0.122	0.136	0.988
CONTENT = English	0.204	0.1374	-0.066	0.474	0.139	0.182	0.215	0.982
CONTENT = Mathematics	0.420	0.1444	0.135	0.704	0.004	0.191	0.226	0.981
CONTENT = Science	0							
GRADE 2017 = 2	0.293	0.1114	0.074	0.512	0.009	0.128	0.143	0.987
GRADE 2017 = 3	-0.157	0.1645	-0.479	0.166	0.341	0.089	0.095	0.991
GRADE 2017 = 4	0.309	0.1481	0.018	0.599	0.037	0.078	0.083	0.992
GRADE 2017 = 5	0							
CONTENT = English * GRADE 2017 = 2	-0.114	0.1992	-0.505	0.277	0.567	0.118	0.131	0.988
CONTENT = English * GRADE 2017 = 3	0.271	0.1787	-0.081	0.622	0.131	0.157	0.181	0.985
CONTENT = English * GRADE 2017 = 4	-0.163	0.1848	-0.526	0.200	0.378	0.142	0.161	0.986
CONTENT = English * GRADE 2017 = 5	0							
CONTENT = Mathematics * GRADE 2017 = 2	-0.191	0.2991	-0.778	0.395	0.523	0.062	0.065	0.994
CONTENT = Mathematics * GRADE 2017 = 3	-0.170	0.2013	-0.565	0.225	0.400	0.099	0.108	0.990
CONTENT = Mathematics * GRADE 2017 = 4	-0.904	0.3496	-1.590	-0.219	0.010	0.042	0.044	0.996
CONTENT = Mathematics * GRADE 2017 = 5	0							
CONTENT = Science * GRADE 2017 = 2	0							
CONTENT = Science * GRADE 2017 = 3	0							

CONTENT = Science * GRADE 2017 = 4	0	
CONTENT = Science * GRADE 2017 = 5	0	
Scale	0.684	0.0088

Table F10: Student Impact Expectation continuous variable information

Analytic sample	Variable type	Measurement	N	Minimum	Maximum	Mean	Std. deviation
Listwise	Dependent variable	SMT scaled score 2018	680	-2.69	2.55	-0.01	0.89
	Covariate	SMT scaled score 2017	680	-2.35	2.32	-0.02	0.90
		SIE mean	680	1.25	2.30	1.92	0.35
Pooled (10 imputations)	Dependent variable	SMT scaled score 2018	1724			-0.15	
	Covariate	SMT scaled score 2017	1724			-0.08	
		SIE mean	1724			1.87	

Pearson

Table F11: Student Impact Expectation categorical variable information

Analytic sample	Factor	N	Percent
Listwise	English	213	31.3
	Mathematics	155	22.8
	Science	312	45.9
	Total	680	100.0
	GRADE 2017 = 2	125	18.4
	GRADE 2017 = 3	87	12.8
	GRADE 2017 = 4	346	50.9
	GRADE 2017 = 5	122	17.9
	Total	680	100.0
Pooled (10 imputations)	English	584	33.9
	Mathematics	377	21.9
	Science	763	44.3
	Total	1724	100.0
	GRADE 2017 = 2	323	18.7
	GRADE 2017 = 3	278	16.1
	GRADE 2017 = 4	733	42.5
	GRADE 2017 = 5	390	22.6
	Total	1724	100.0

Pearson

Table F12: Student Impact Expectation correlated data summary

Analytic sample	Number of levels subject effect (school)	Number of levels within-subject effect Case_ID	Number of measurements per subject minimum	Number of measurements per subject maximum
Listwise	13	680	2	135
Pooled (10 imputations)	13	1724	22	357

Pearson

Table F13: Student Impact Expectation listwise parameter estimates

Model effect	B	Std. error	95% Wald confidence interval		Hypothesis test		
			Lower	Upper	Wald Chi-square	df	P-value
Intercept	-1.685	0.5540	-2.770	-0.599	9.247	1	0.002
Student Impact Evaluation mean	0.792	0.3174	0.170	1.414	6.230	1	0.013
SMT scaled score 2017	0.338	0.1000	0.142	0.534	11.438	1	0.001
CONTENT = English	0.087	0.2195	-0.343	0.517	0.158	1	0.691
CONTENT = Mathematics	0.212	0.2023	-0.185	0.608	1.095	1	0.295
CONTENT = Science	0						
GRADE 2017 = 2	0.327	0.2811	-0.224	0.878	1.354	1	0.245
GRADE 2017 = 3	0.002	0.1441	-0.281	0.284	0.000	1	0.991
GRADE 2017 = 4	0.176	0.1906	-0.198	0.549	0.849	1	0.357
GRADE 2017 = 5	0						
CONTENT = English * GRADE 2017 = 2	0						
CONTENT = English * GRADE 2017 = 3	0.558	0.1484	0.267	0.848	14.116	1	0.000
CONTENT = English * GRADE 2017 = 5	0						
CONTENT = Mathematics * GRADE 2017 = 3	-0.196	0.2707	-0.726	0.335	0.522	1	0.470
CONTENT = Mathematics * GRADE 2017 = 4	-0.677	0.2516	-1.170	-0.184	7.238	1	0.007
CONTENT = Mathematics * GRADE 2017 = 5	0						
CONTENT = Science * GRADE 2017 = 3	0						
CONTENT = Science * GRADE 2017 = 4	0						
CONTENT = Science * GRADE 2017 = 5	0						
Scale	0.612						

Pearson

Table F14: Student Impact Expectation pooled parameter estimates

Model effect	B	Std. error	95% Wald confidence interval		P-value	Fraction missing info	Relative increase variance	Relative efficiency
			Lower	Upper				
Intercept	-1.603	0.4411	-2.468	-0.737	0.000	0.113	0.124	0.989
Student Impact Evaluation mean	0.708	0.2271	0.262	1.154	0.002	0.106	0.116	0.989
SMT scaled score 2017	0.293	0.0635	0.168	0.417	0.000	0.060	0.063	0.994
CONTENT = English	0.170	0.1135	-0.055	0.395	0.136	0.306	0.414	0.970
CONTENT = Mathematics	0.131	0.1158	-0.105	0.368	0.265	0.570	1.185	0.946
CONTENT = Science	0							
GRADE 2017 = 2	0.354	0.2085	-0.056	0.763	0.090	0.135	0.152	0.987
GRADE 2017 = 3	-0.066	0.1355	-0.333	0.202	0.629	0.224	0.275	0.978
GRADE 2017 = 4	0.170	0.1116	-0.049	0.389	0.129	0.137	0.154	0.987
GRADE 2017 = 5	0							
CONTENT = English * GRADE 2017 = 2	0							
CONTENT = English * GRADE 2017 = 3	0.368	0.1573	0.056	0.681	0.021	0.325	0.452	0.968
CONTENT = English * GRADE 2017 = 5	0							
CONTENT = Mathematics * GRADE 2017 = 3	-0.031	0.3081	-0.635	0.574	0.921	0.111	0.122	0.989
CONTENT = Mathematics * GRADE 2017 = 4	-0.567	0.2547	-1.067	-0.067	0.026	0.105	0.115	0.990
CONTENT = Mathematics * GRADE 2017 = 5	0							
CONTENT = Science * GRADE 2017 = 3	0							
CONTENT = Science * GRADE 2017 = 4	0							
CONTENT = Science * GRADE 2017 = 5	0							
Scale	0.604	0.0132						

Pearson

Appendix G. Multiple imputation statistics

Table G1: 2017 Skills Mapping Test statistics

Data	Imputation	N	Mean	Std. deviation	Minimum	Maximum
Original data		9,844	-0.000	0.921	-2.850	2.519
Imputed values	1	6,431	0.009	0.940	-3.326	3.194
	2	6,431	0.004	0.947	-3.337	3.089
	3	6,431	0.019	0.938	-3.832	3.091
	4	6,431	0.007	0.917	-3.582	3.008
	5	6,431	-0.002	0.930	-3.334	3.245
	6	6,431	0.003	0.933	-3.796	3.818
	7	6,431	-0.004	0.951	-3.951	3.299
	8	6,431	-0.008	0.921	-3.484	3.057
	9	6,431	0.001	0.941	-3.658	3.301
	10	6,431	0.021	0.931	-3.220	3.641
Complete data after imputation	1	16,275	0.003	0.929	-3.326	3.194
	2	16,275	0.002	0.931	-3.337	3.089
	3	16,275	0.008	0.928	-3.832	3.091
	4	16,275	0.003	0.919	-3.582	3.008
	5	16,275	-0.001	0.925	-3.334	3.245
	6	16,275	0.001	0.926	-3.796	3.818
	7	16,275	-0.002	0.933	-3.951	3.299
	8	16,275	-0.003	0.921	-3.484	3.057
	9	16,275	0.001	0.929	-3.658	3.301
	10	16,275	0.008	0.925	-3.219	3.641

Pearson

Table G2: 2018 Skills Mapping Test statistics

Data	Imputation	N	Mean	Std. deviation	Minimum	Maximum
Original data		12,537	-0.000	0.923	-2.804	2.838
Imputed values	1	3,738	-0.080	0.923	-3.299	3.425
	2	3,738	-0.069	0.908	-3.060	3.184
	3	3,738	-0.071	0.906	-3.573	3.329
	4	3,738	-0.063	0.919	-3.270	3.219
	5	3,738	-0.042	0.894	-3.097	3.150
	6	3,735	-0.028	0.901	-2.828	3.445
	7	3,738	-0.064	0.893	-3.019	3.312
	8	3,738	-0.048	0.909	-2.812	3.511
	9	3,738	-0.065	0.912	-3.158	3.295
	10	3,738	-0.045	0.893	-3.243	3.130
Complete Data After Imputation	1	16,275	-0.018	0.923	-3.299	3.425
	2	16,275	-0.016	0.920	-3.060	3.184
	3	16,275	-0.016	0.919	-3.573	3.329
	4	16,275	-0.015	0.922	-3.270	3.219
	5	16,275	-0.010	0.916	-3.097	3.150
	6	16,275	-0.007	0.918	-2.828	3.445
	7	16,275	-0.015	0.916	-3.019	3.312
	8	16,275	-0.011	0.920	-2.812	3.511
	9	16,275	-0.015	0.921	-3.158	3.295
	10	16,275	-0.011	0.916	-3.243	3.130