CCWD
CENTER FOR CHILD WELL-BEING
& DEVELOPMENT

University of Zurich[UZH]

Proxies for Artificial Intelligence

Guilherme Lichand

University of Zurich

supported by

unicef Schweiz Suisse Svizzera

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Direktion für Entwicklung
und Zusammenarbeit DEZA

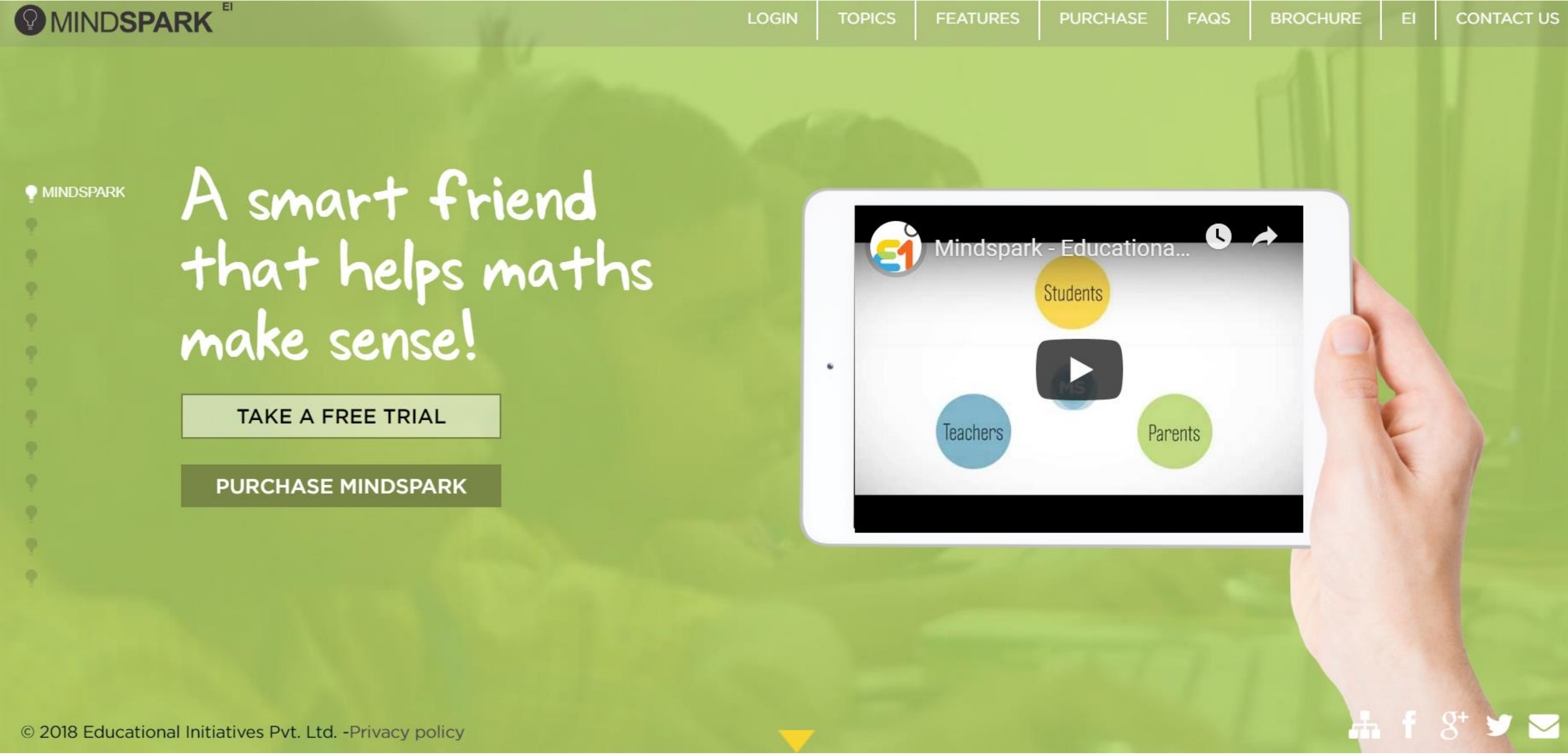
A fundamental mismatch

Muralidharan, Singh and Ganimian (2018):

“A leading candidate explanation for this low productivity is that existing patterns of education spending and instruction may not alleviate a key binding constraint to learning, which is the mismatch between the level of classroom instruction and student learning levels.”

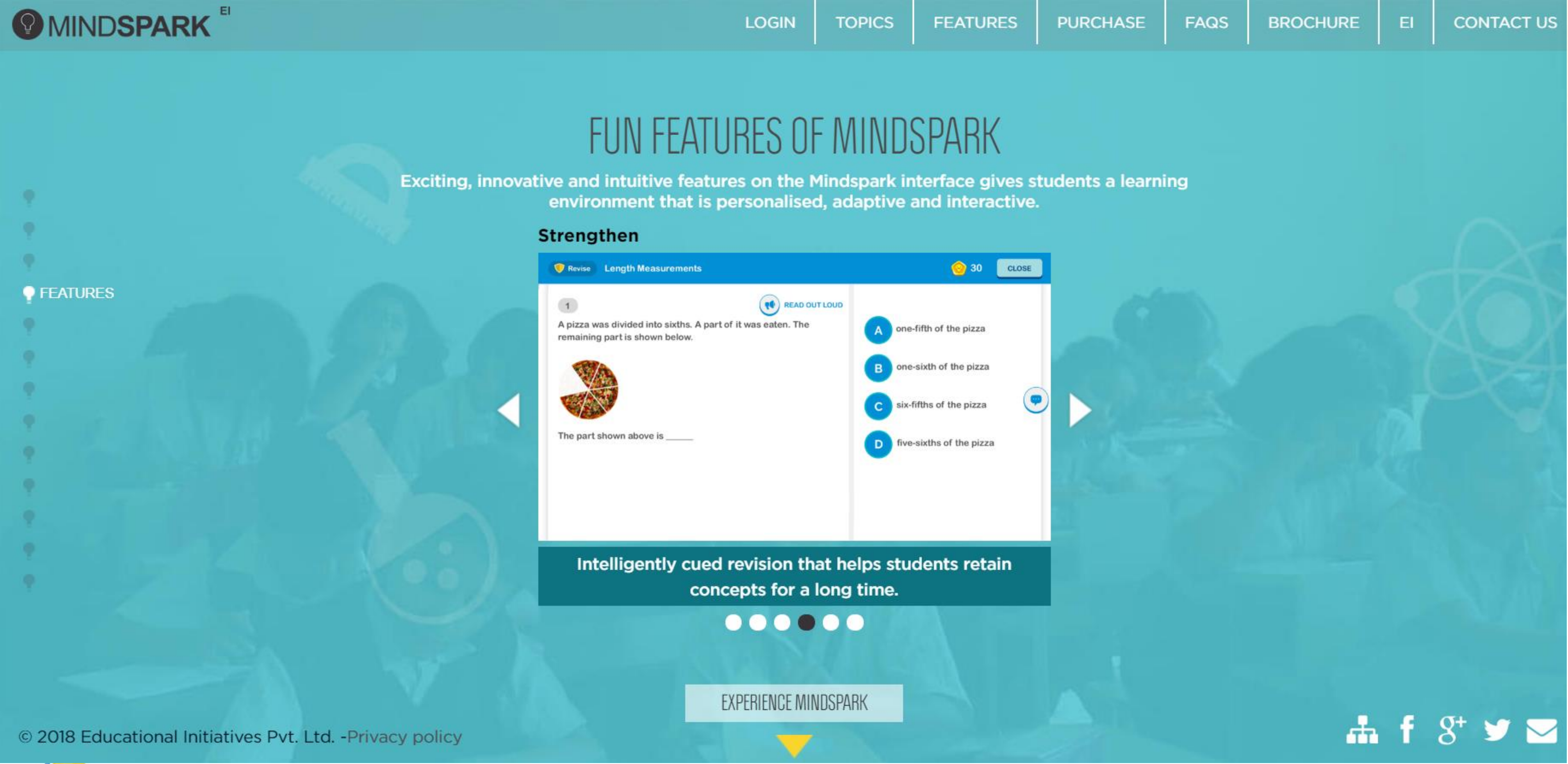
What to do about it?

- Teaching at the Right Level
  - Proven impacts
  - Challenges: hard to scale (Barnerjee et al., 2016)
- Adaptive learning platforms
  - Limited evidence in developing countries



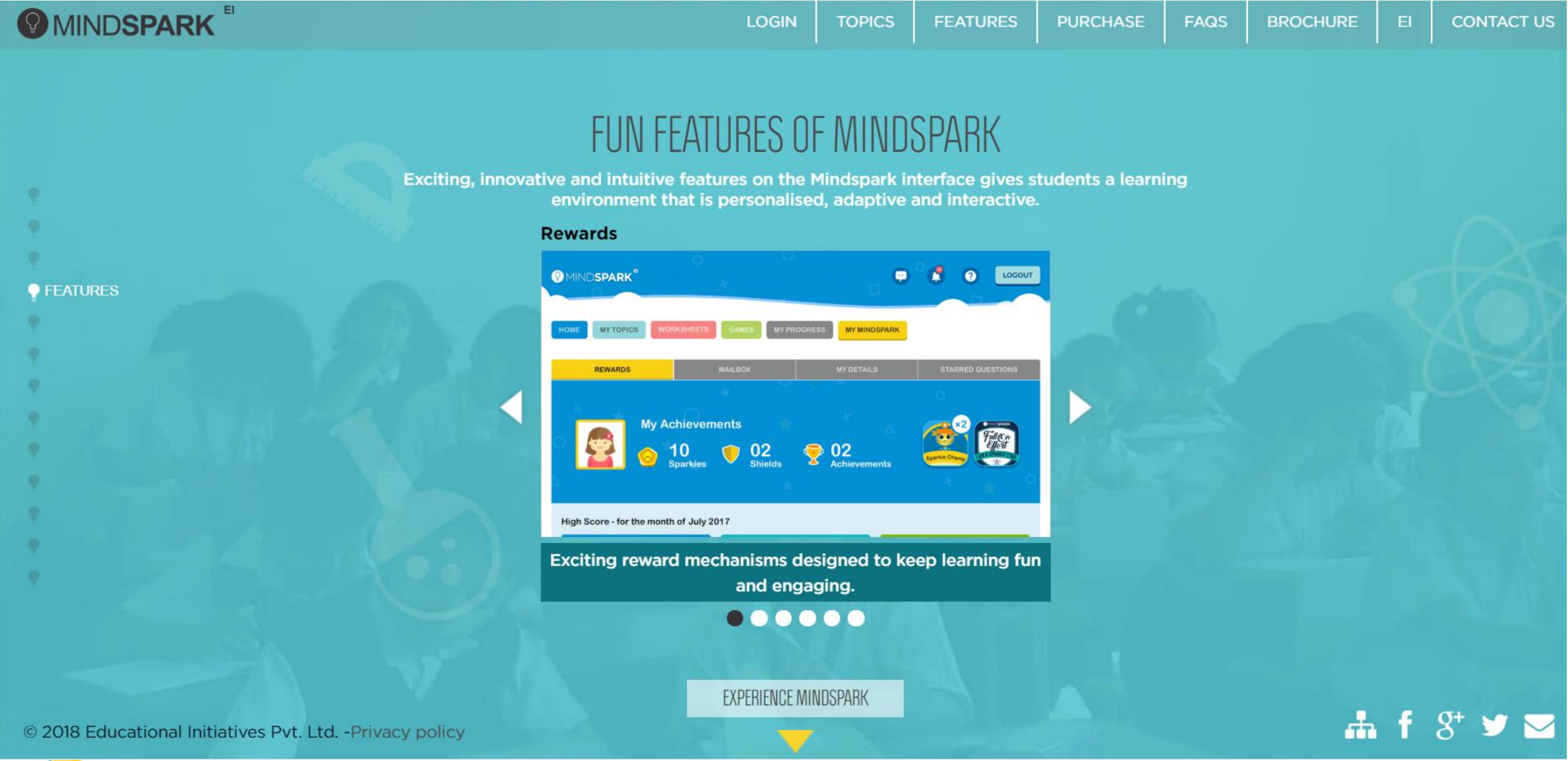
CCWD
CENTER FOR CHILD WELL-BEING & DEVELOPMENT

University of Zurich[UZH]



CCWD
CENTER FOR CHILD WELL-BEING & DEVELOPMENT

University of Zurich[UZH]



CCWD CENTER FOR CHILD WELL-BEING & DEVELOPMENT

University of Zurich UZH

Mindspark

Muralidharan, Singh and Ganimian (2018): Mindspark – tablet-based learning in India, with real-time fine tuning of content for customized learning

- Huge effects on Math test scores: equivalent to 1-year improvement

Customizing works!

Why not do it for all educational programs?

Fundamental constraint

In poor countries, typically a missing data problem:

1. Learning outcomes typically low-frequency
2. Lack of devices, connectivity and teacher training challenge the scale up of approaches such as that of Mindspark
3. Even in Mindspark, measurement is not separate from intervention, limiting the policy space that can leverage on such high-frequency measurement

This project…

Is about using mobile surveys to generate:

- ➢ High-frequency proxies for low-frequency learning outcomes
- ➢ With measurement separate from interventions

Research questions

1. Which mobile survey questions generate high-frequency proxies most predictive of low-frequency learning outcomes?
2. Can fine-tuning educational interventions based on those high-frequency proxies improve learning?

Issues

- For what kind of programs does it make sense?
  - Particularly for teacher training programs, since we can piggyback on publich health's *syndemic approach* – surveying students and parents (beyond teachers) to predict "learning epidemics" at the classroom-level
  - Such predictions tend to be much more precise than individual-level

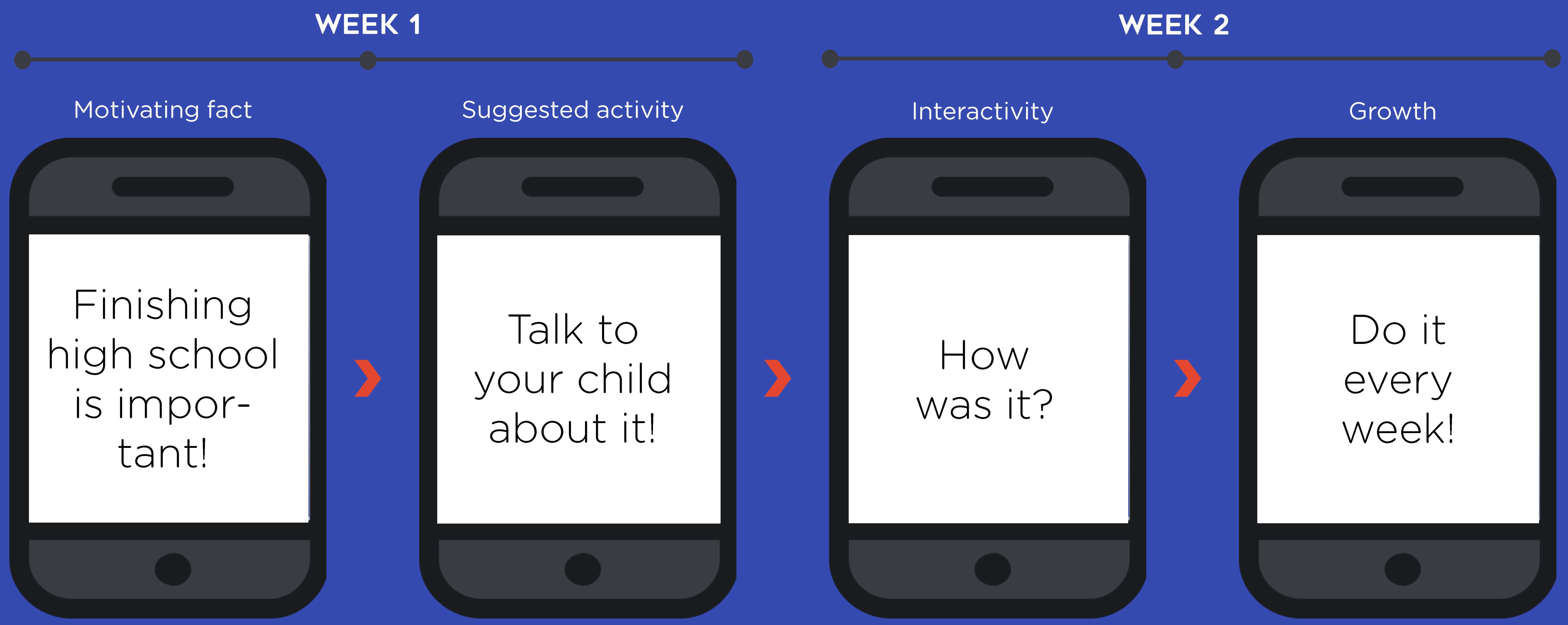
- Dynamic effects: what is the optimal frequency of fine-tuning?
  - In principle, as high as the frequency of proxies
  - In practice, depends on the program

Literature

- Blumenstock et al. (2018): mobile usage patterns to predict income shocks in poor countries in real time
  - Income vs Relevant metrics for child development

- Pillsbury et al. (2017): SMS reporting of symptoms to predict disease outbreaks
  - Syndromic approach for Health vs. Education

Example: Eduq+ (Bettinger et al., 2018)



Treatment Variations

Nudging: parental engagement via SMS (Bettinger et al., 2018)

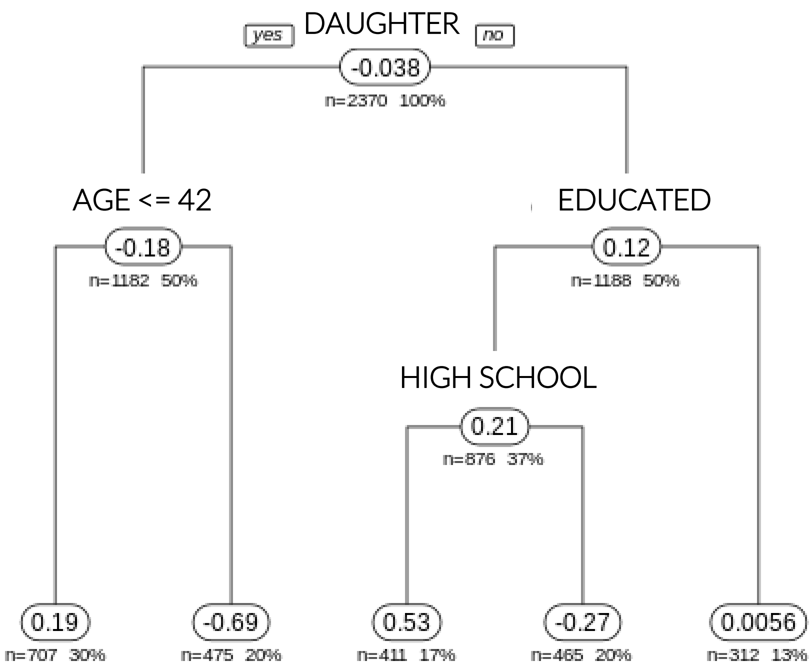
24 Variations:

- Number of messages per week (1, 2 or 3)
- Time of delivery (afternoon or evening)
- Consistency (same time or rotating)
- Interactivity (with or without)

Causal Random Forrest

Nudging: parental engagement via SMS (Bettinger et al., 2018)

> Causal tree for variation #1:

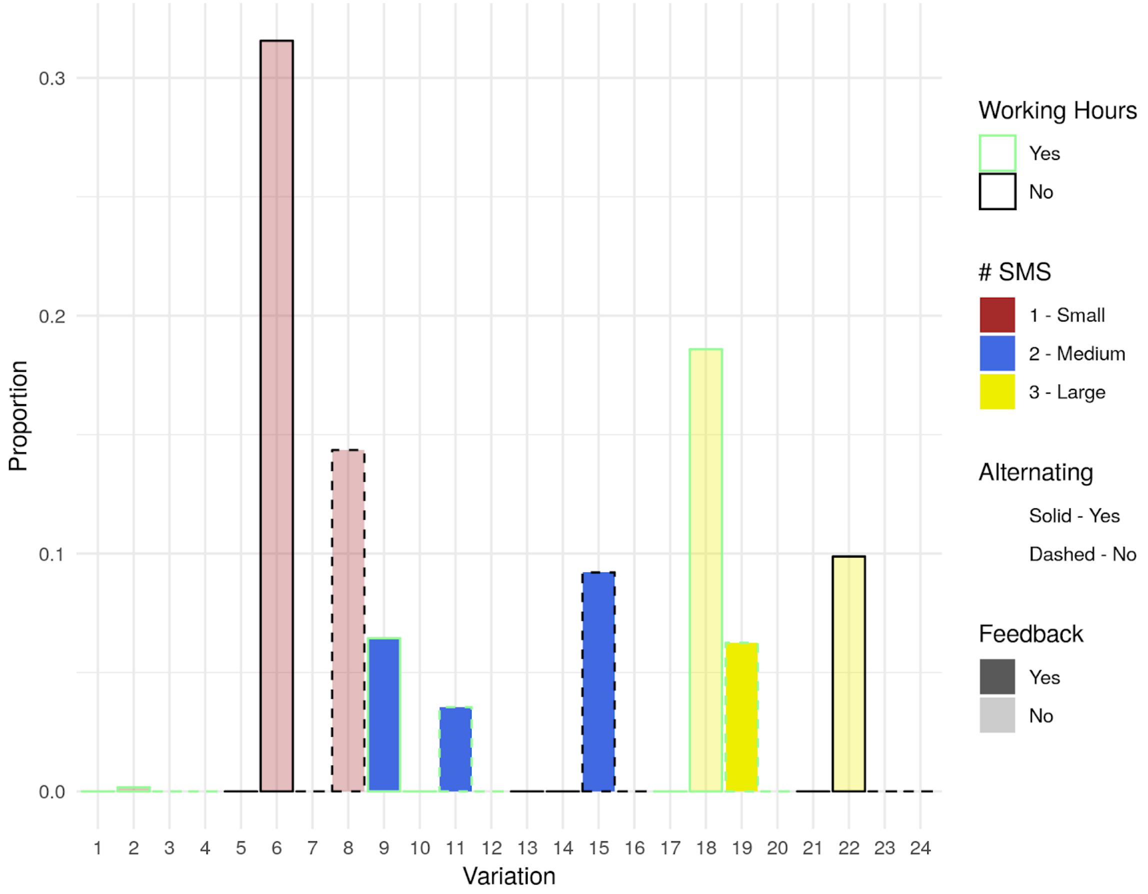


CCWD CENTER FOR CHILD WELL-BEING & DEVELOPMENT

University of Zurich UZH

% for whom variation is the most impactful

Nudging: parental engagement via SMS (Bettinger et al., 2018)

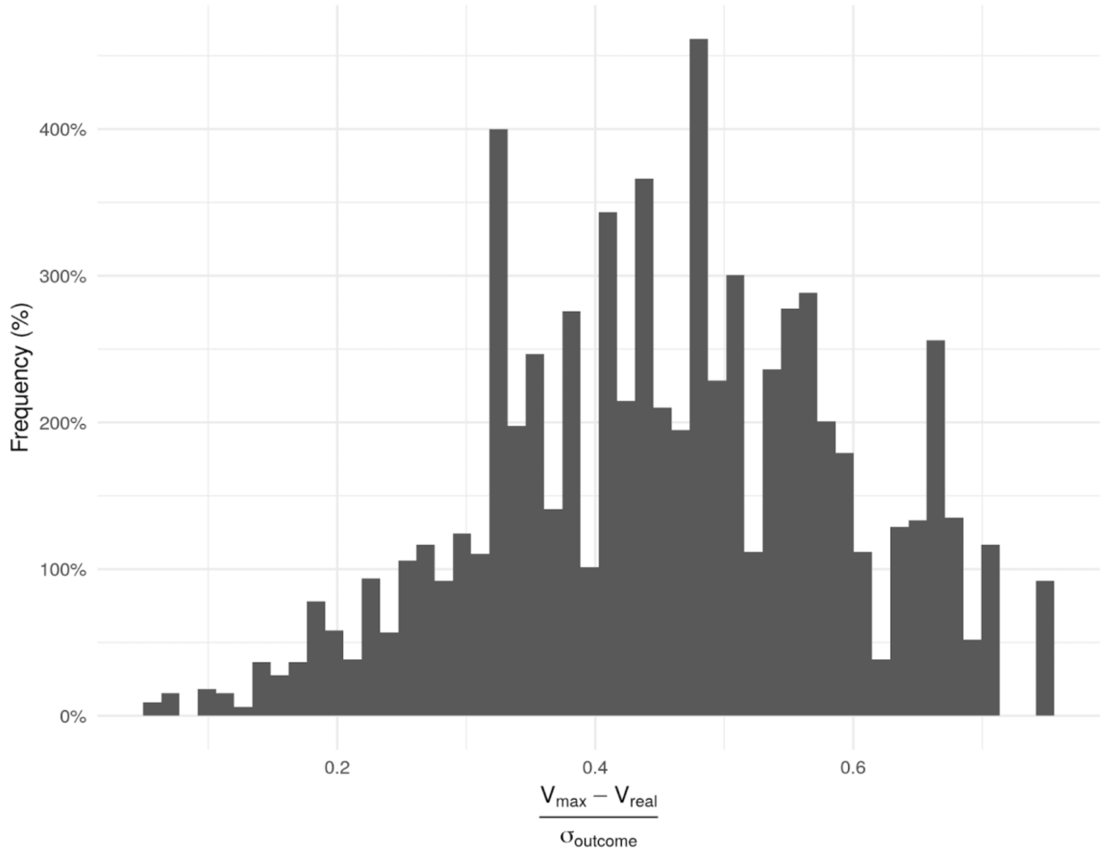


CCWD CENTER FOR CHILD WELL-BEING & DEVELOPMENT

University of Zurich[UZH]

Learning gains from oracle assignment

Nudging: parental engagement via SMS (Bettinger et al., 2018)



CCWD
CENTER FOR CHILD WELL-BEING & DEVELOPMENT

University of Zurich[UZH]

Input generation

Nudging: parental engagement via SMS (Bettinger et al., 2018)

- Bi-weekly SMS survey: “Did you do the activity? YES or NO?”
- 10-15% weekly response rate.

Feature Selection

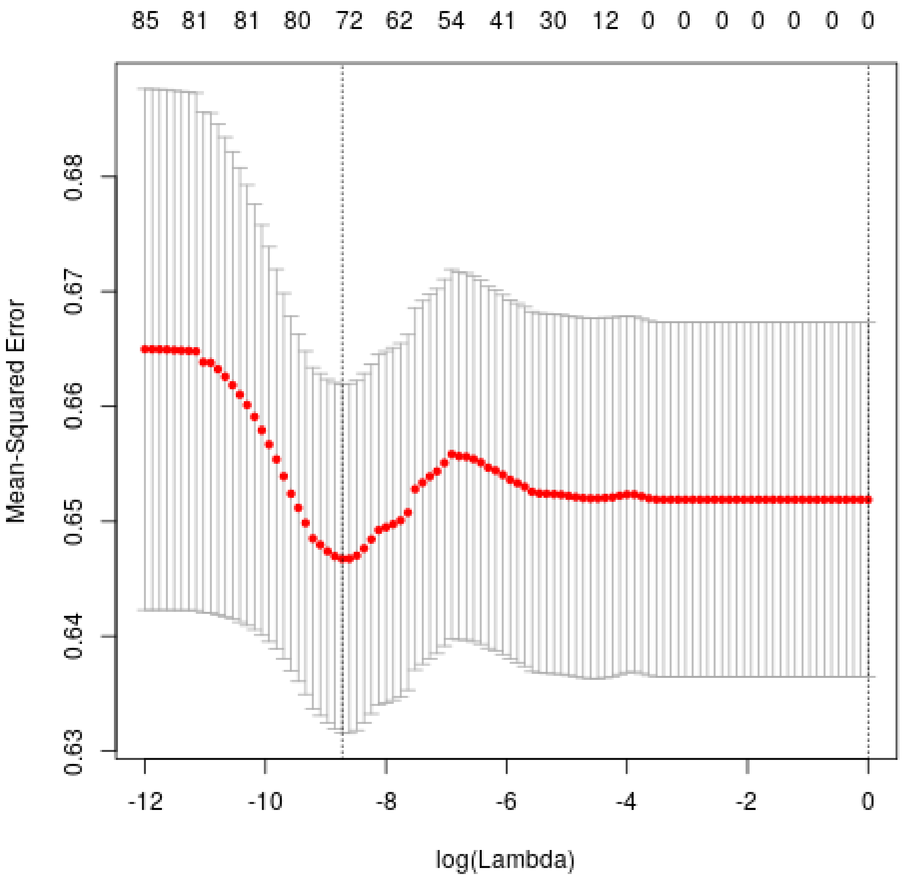
Nudging: parental engagement via SMS (Bettinger et al., 2018)

“Did you do the activity?”

What features of users’ responses predict increase in Match achievement?

LASSO picks:

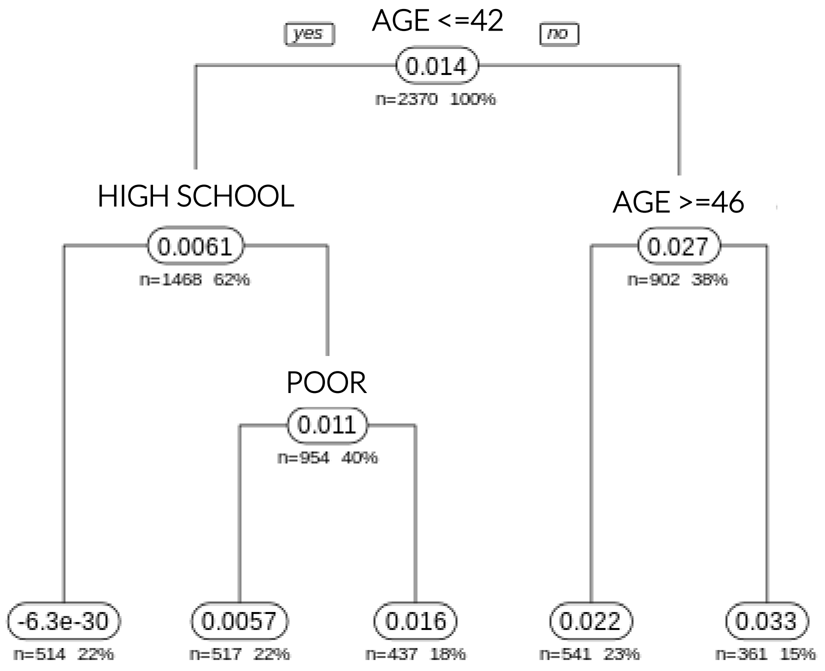
```
[1] "nchar_mean"      "nword_mean"      "long_mean"       "lower_mean"      "ans.cool_mean"
[7] "ans.thank1_mean" "ans.sure_mean"   "ans.truth_mean" "ans.affection_mean" "ans.love_mean"   "ans.love1_mean"
[13] "ans.friends_mean"   "ans.feel_mean" "ans.math_mean"   "ans.activity_mean"  "ans.dream_mean" "ans.dream1_mean"
[19] "ans.yes_mean"    "ans.cancel_mean" "nchar_max"       "nword_max"       "long_max"        "lower_max"
[25] "ans.boy_max"     "ans.cool_max"    "ans.thank_max"   "ans.thank1_max" "ans.sure_max"    "ans.truth_max"
[31] "ans.affection_max"  "ans.love_max" "ans.love1_max"   "ans.friends_max" "ans.feel_max"    "ans.no_max"
[37] "ans.math_max"    "ans.activity_max"   "ans.dream_max" "ans.dream1_max" "ans.yes_max"     "ans.cancel_max"
[43] "nword_min"       "long_min"        "lower_min"       "ans.boy_min"     "ans.thank_min"   "ans.thank1_min"
[49] "ans.love_min"    "ans.love1_min"   "ans.dream1_min" "ans.yes_min"     "ans.cancel_min" "nchar_var"
[55] "nword_var"       "long_var"        "lower_var"       "ans.boy_var"     "ans.cool_var"    "ans.thank_var"
[61] "ans.thank1_var" "ans.sure_var"    "ans.truth_var"   "ans.affection_var"  "ans.love_var" "ans.love1_var"
[67] "ans.feel_var"    "ans.no_var"      "ans.math_var"    "ans.dream_var"   "ans.dream1_var" "ans.yes_var"
[73] "week_mean"       "week_max"        "week_min"
```



Causal Proxy Random Forrest

Nudging: parental engagement via SMS (Bettinger et al., 2018)

> Causal proxy tree for variation #1:

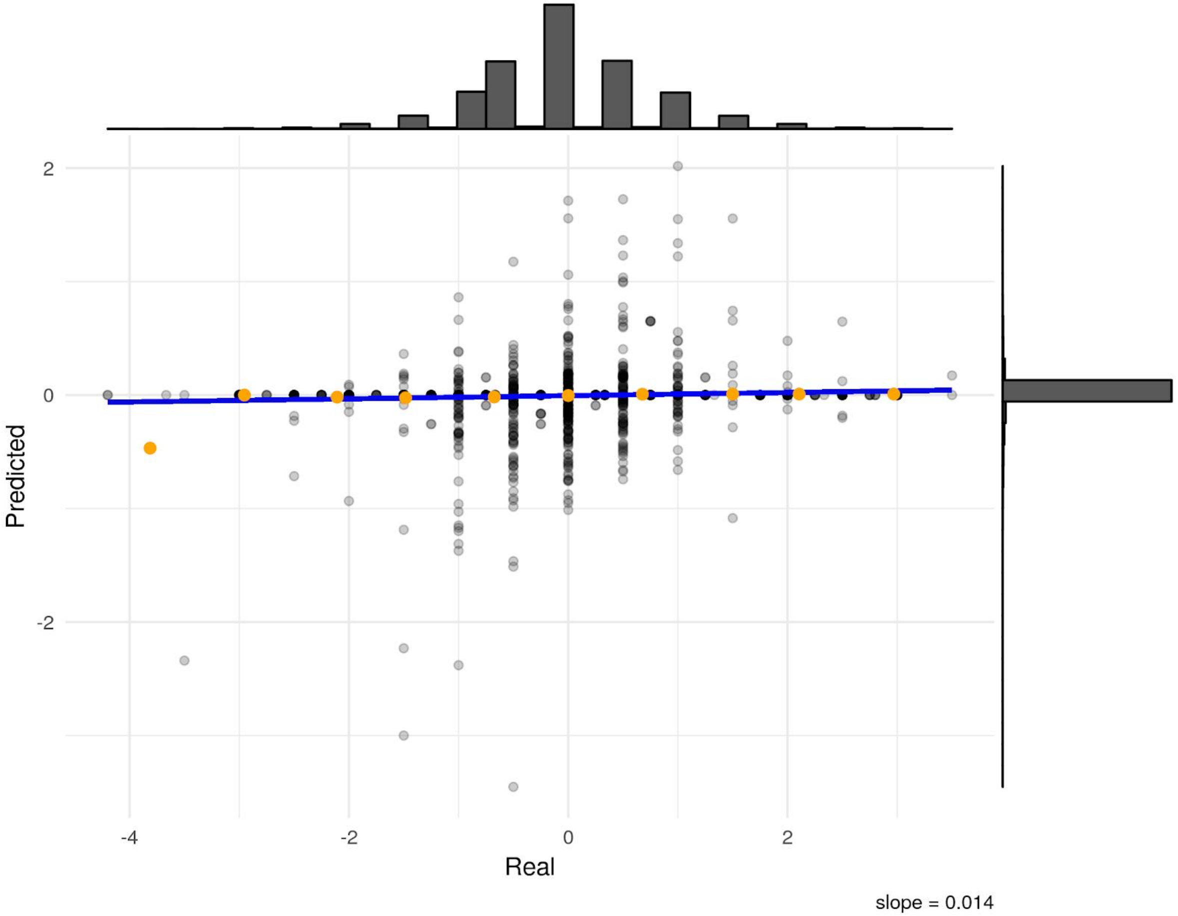


How does the proxy perform?

Nudging: parental engagement via SMS (Bettinger et al., 2018)

“Did you do the activity?”

A poor ‘thermometer’…



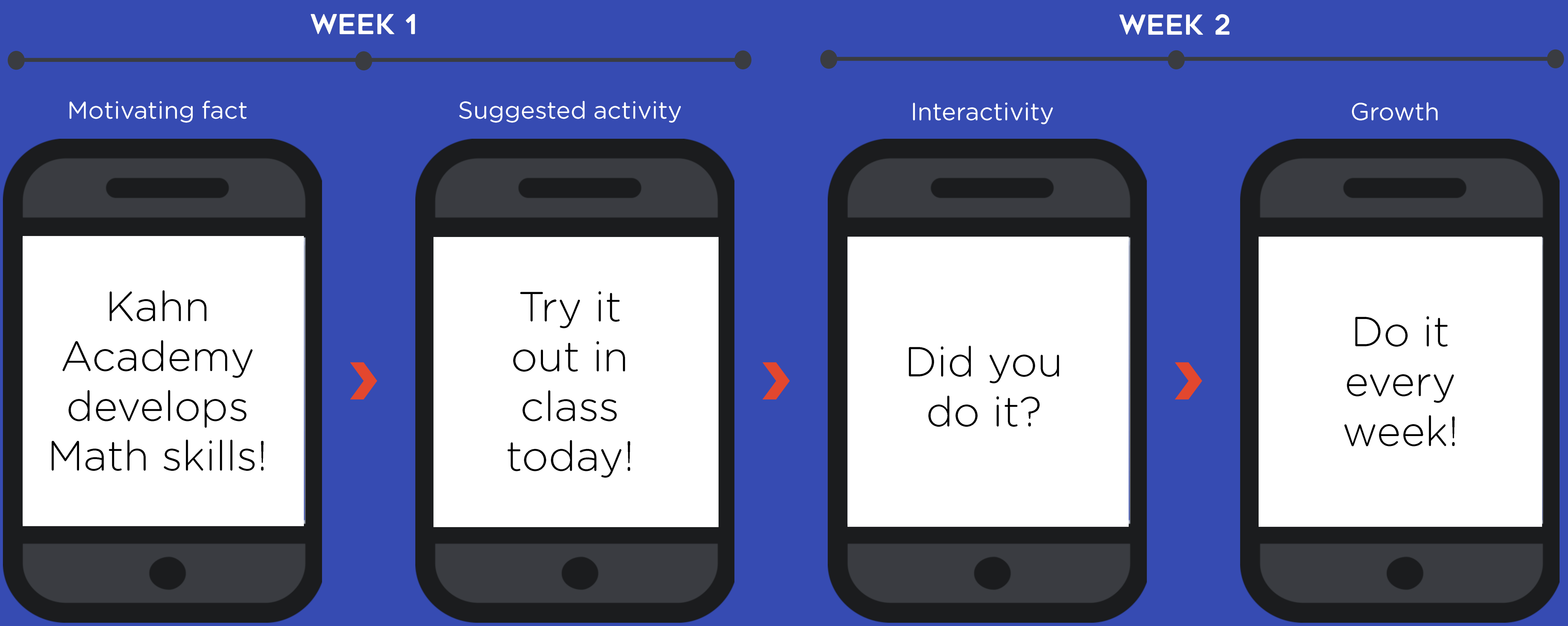
CCWD
CENTER FOR CHILD WELL-BEING & DEVELOPMENT

University of Zurich[UZH]

Coming next…

- **Eduq+ Teachers:** SMS nudges to 20,000 teachers in Brazil

Eduq+ Teachers



Empirical strategy

1. Evaluates multiple strategies to build high-frequency proxies (‘thermometers’) via SMS surveys with teachers, parents and students, identifying which strategy is most predictive of impacts on low-frequency learning outcomes.
2. Develops and tests AI4D based on high-frequency proxies from SMS surveys in Brazil, for the case of communicating with teachers (96 variations) with the goal of improving students’ learning.

Roadmap

1. **Input generation**
2. Treatment variations: randomizing communication
3. Feature selection
4. Causal random forests
5. Randomizer and fine-tuning assignment
6. Evaluating AI4D

Testing different ‘thermometers’

Examples of strategies for capturing teachers’ inputs:

1. Was the activity useful? (YES or NO)
2. How hard is it to get rid of old habits? (1 to 10)
3. How open are you to best practices?
4. Did you rethink what you do in the classroom?
5. Do you feel like sharing best practices with other teachers?
6. How engaged are parents/students?
7. Has X happened? [X = each step in Eduq+’s Theory of Change]
8. How confident do you feel about teaching your classes?
9. Do you feel that your work is high-quality/making an impact?
10. Predict your students’ average Math standardized test score

Testing different ‘thermometers’

Optimal thermometer may differ across stakeholders:

- Teachers
- Parents
- Students

CCWD
CENTER FOR CHILD WELL-BEING & DEVELOPMENT

University of Zurich[UZH]

Roadmap

1. Input generation
2. Treatment variations: randomizing communication
3. Feature selection
4. Causal random forests
5. Randomizer and fine-tuning assignment
6. Evaluating AI4D

Variations

Motivating Fact			
Baseline version	**Shocking version**	**Negative framing**	**Colloquial Tone**
5,000	5,000	5,000	5,000
Suggested Activity			
Baseline version	**Relies on other individuals**	**Draws on authority for credibility**	**Draws on emotional appeal**
5,000	5,000	5,000	5,000
Growth message			
Baseline version		**More demanding**	
10,000		10,000	
Cross-Cutting			
Baseline version	**Abbreviated Whenever possible**	**Exaggerated use of punctuation marks (???, !!!)**	
6,666	6,667	6,667	

Sample distribution under *Business as Usual*

Roadmap

1. Input generation
2. Treatment variations: randomizing communication
3. Feature selection
4. Causal random forests
5. Randomizer and fine-tuning assignment
6. Evaluating AI4D

CCWD
CENTER FOR CHILD WELL-BEING & DEVELOPMENT

University of Zurich[UZH]

Roadmap

1. Input generation
2. Treatment variations: randomizing communication
3. Feature selection
4. Causal random forests
5. Randomizer and fine-tuning assignment
6. **Evaluating AI4D**

CCWD CENTER FOR CHILD WELL-BEING & DEVELOPMENT

University of Zurich[UZH]

Sample design

Business as Usual	AI4D (proxy-based)
10,000	10,000

Final remarks

- Lots of work ahead…
- Your inputs are greatly appreciated!