**Indonesia Aspirations and Career Choices**

**Pre-Analysis Plan**

**Drafted: April 2019**

# Introduction

The aim of this pre-analysis plan is to define the methodology that we will follow to conduct our analysis. The first section of the plan will provide an overview of the study, including a description of the interventions, the sampling process, data sources and sample size calculations. The second section will describe the hypotheses that we will be testing and how we are defining the variables that will test the hypotheses. The final section will discuss the methodology and specifications to be used in the analysis.

# Overview of the study

This study is a large-scale randomized controlled trial that aims to improve the academic achievement and aspirations of 9th grade students in Indonesia. Indonesian students perform worse on international achievement tests than their counterparts in most other countries (PISA, 2015). Moreover, the Indonesian labor market is characterized by low female labor force participation and occupational sorting along gender lines. This study tests whether two interventions delivered under the Semua Bisa Sukses program (SBS, or “All Can Succeed”) to boost students’ socioemotional skills can improve their school outcomes and raise their aspirations for their future education and career.

## Description of the interventions

The first intervention consists of a 6-week socioemotional skills curriculum based in the concept of growth mindset. Growth mindset refers to the perception that intelligence is malleable and that it is possible to improve one’s ability through practice and effort. The curriculum is rolled out in one 45-minute session per week for six weeks. The sessions include comics that introduce the topic of the week, guided classroom discussions, student reflection, small group discussions, and other interactive activities.

Provided in combination with the student curriculum, the second intervention includes tools and activities that teachers can integrate in their daily interactions with students. The goal of these tools and activities is to shift the classroom environment to be conducive to adopting a growth mindset and to reinforce the principles of the training daily. The teachers receive tip sheets with ways to make their feedback to students promote growth mindset and materials for two activities: “I see SBS behavior” and “Sharing SBS experiences.” In "I see SBS behavior," teachers and students together will look for examples of students demonstrating the behaviors taught in the SBS program. Every morning, teachers and students will celebrate the examples of students who were observed demonstrating the promoted SBS behavior the day before. For the "Sharing SBS experiences," students and teachers will anonymously write about their experiences with the behaviors promoted in the SBS program and post them on a community sharing board. Each week, the prompts for these activities will be aligned with the concepts being taught in the student curriculum sessions. The schools also receive instructions for creating a Tim Perubahan, or change team, comprised of the headmaster, guidance counselor, and some teachers and tasked with organizing a training workshop for teachers, supporting teachers in implementing the program, and organizing school-wide celebrations of classrooms that are particularly active.

## Description of the sampling process

To evaluate the impacts of the program and compare the two packages of interventions, the team uses RCT methodology. Due to both concerns over spillovers and implementation constraints, the randomization is carried out at the school level. The impact evaluation takes place in 2,097 schools in 60 districts in Java and Sumatera: 699 schools are randomly selected to receive a 6-week program teaching growth-mindset and related concepts (treatment arm 1); 699 schools are randomly selected to receive additional teaching tools and activities for classroom teachers in addition to the 6-week program offered in the first treatment arm (treatment arm 2); 699 schools are randomly selected to serve as a control group for the impact evaluation and will not receive the interventions.

Only public (Negeri) middle schools (Sekolah Menengah Pertama – SMP) are targeted—Islamic schools (Madrasah Tsanawiyah – MTs) and technical or remedial schools (SMPT) are excluded from the sample. In addition, the sample is restricted to the islands of Java and Sumatera in order to facilitate logistics and contain costs; however, the sample is fairly representative of public middle schools (SMP Negeri) in Java and Sumatera.[[1]](#footnote-2)

To conduct the selection, the team obtained a list of all SMP schools from the Ministry of Education and Culture and applied the eligibility criteria (public school located in Java or Sumatera). Then, the team used the technique of probability proportional to size sampling to select the 2097 sample schools, restricting the sample to only 60 districts in Java and Sumatera, again for logistics and cost issues. With this technique, every school in Java and Sumatera has the same opportunity to be selected into the sample.[[2]](#footnote-3) In the first step, we randomly select the districts, and districts with a larger number of schools have a greater probability of being selected. In the second step, we randomly select the same number of schools from each selected district, so schools within small selected districts have a greater probability of being selected than schools within large selected districts.

After selecting the 2100 schools, we randomly allocated treatment assignment, stratifying by district.

## Description of the data sources

The impact evaluation uses a combination of survey and administrative data. The survey data is measured at baseline and endline and is focused on measuring student aspirations and socio-emotional skills using both self-reported and behavioral measures. The baseline survey was collected in February and March 2018, when the students were in 8th grade. The endline survey was collected in February and March 2019, when the students are nearing the end of their 9th grade year.[[3]](#footnote-4) The timing of the surveys takes into account the national exam schedule. National exams take place in May. In April, teachers are busy preparing students for exams, so taking class time for surveys would be more challenging.

The baseline and end-line survey take place in all 2,097 sample schools. In schools with multiple classrooms of 8th graders, one classroom was randomly selected to be interviewed at baseline, and these same students were interviewed at endline. The number of students interviewed per school varies depending on the size of the selected classroom; however, the average class size is 27 students. The student interviews are self-administered during school using paper-based surveys and take approximately 40 minutes to complete. A trained enumerator reads the questions out loud to the class and can assist students if they have questions or are unsure how to mark their responses on the forms. In addition, 4 students (2 boys and 2 girls) were randomly selected at baseline to complete the PERC, a task-based assessment of persistence, effort, resilience and challenge seeking, which is self-administered on tablets under the guidance of a trained enumerator. These same students were invited to complete the PERC at endline as well.

In addition to the student interviews, the guidance counselor and two teachers, including the homeroom teacher and the math teacher were interviewed in each school. They are asked about their mindset, expectations of students, familiarity with the topics taught in the program, and some basic demographic information. The endline surveys also include questions on treatment compliance and fidelity of implementation. This data will enable the team to test for channels through which the program influences students as well as how teacher or counselor characteristics can moderate the effectiveness of the program. The teachers and guidance counselors are interviewed using paper-based, face-to-face interviewing techniques.

In addition to survey data, administrative data will be collected from Ministry of Education and Culture (MOEC), including national test score data. The team is also seeking administrative data on students’ drop-out, graduation, and track choices for after SMP from either MOEC or the schools.

Note: In the list of variables below, each variable’s source is indicated with A for administrative data and S for survey data.

# Hypotheses and variable definitions

The question numbers used below is taken from the baseline survey, unless otherwise indicated. Unless otherwise indicated, the survey questions are referring to the student questionnaire. Questions from the teacher questionnaire are indicated with TS for teacher survey before the question number, and questions from the counselor questionnaire are indicated with CS for counselor survey before the question number.

## Groups of hypotheses:

**A: Impact of the interventions on primary outcomes**

Hypothesis group A looks at the impacts of the program on the main outcomes that the program was trying to influence.

**B: Impact of the interventions on secondary outcomes**

Hypothesis group B explores the impacts of the program on outcomes that the program was not necessarily targeting but that may have changed due to the program.

**C: Process and Mechanisms**

In hypothesis group C, we will test the theory of change that links the program participation to final outcomes. This hypothesis group explores how and why any observed effects were produced.

**D: Heterogeneity of impacts**

Hypothesis group D examines whether the program impacts different types of beneficiaries in a different way than others. Notably, this section will examine gender disaggregated effects, which is of primary interest for this evaluation.

## Group A: Impact of the interventions on primary outcomes

### Hypothesis A1: The interventions improve students’ academic performance

We will test the results on each of the indicators individually and on a standardized z-score indicator of all the following indicators:

* National test scores, Bahasa Indonesia (A): The raw value of the students’ national test score will be used.
* National test scores, Math (A): The raw value of the students’ national test score will be used.
* National test scores, Science (A): The raw value of the students’ national test score will be used.
* National test scores, English (A): The raw value of the students’ national test score will be used.
* National test scores, average of four core subjects (A): The average value of students’ scores in Bahasa Indonesia, Math, Science and English on the national exams will be used.
* School test scores, Bahasa Indonesia (S): The raw value of the students’ final semester school grade will be collected from the schools.
* School test scores, English (S): The raw value of the students’ final semester school grade will be collected from the schools.
* School test scores, Math (S): The raw value of the students’ final semester school grade will be collected from the schools.
* School test scores, Natural Sciences (S): The raw value of the students’ final semester school grade will be collected from the schools.
* School test scores, average of four core subjects (S): The average value of the students’ final semester school grades in Bahasa Indonesia, English, Math and Natural Sciences.
* School test scores, average of all subjects (S): The average value of the students’ final semester school grades in Civic Education, Religion, Bahasa Indonesia, English, Math, Natural Sciences, and Social Sciences.

### Hypothesis A2: The interventions improve students’ mindsets, perceptions and socioemotional skills

We will begin by exploring measurement models to determine whether the scales below can be aggregated into one or more composite scores. We will retain the measurement model that best balances representing the data and the theory supporting these scales. Then, we will create standardized composite scores based on the higher-order factors of the best measurement model.

Scales:

* Growth mindset
* Effort beliefs
* Self-esteem
* Sense of belonging
* Perseverance
* Learning orientation
* Performance avoidance orientation
* Learning goal
* Performance avoidance goal

We will also test the effect of the program on students’ mastery behavior, assessed by the PERC, a task-based assessment administrated to a random subsample of the surveyed students and giving four task-based measures: persistence, effort investment, resilience, and challenge seeking. We will create a composite score of students’ mastery behavior based on their scores in this four aspects in the task test. Given a composite is created from students’ performance in the task, students’ responses to the PERC will not be used in the measurement exploration of the survey scales.

After we create composite scores, we will test the impacts of the program individually on each composite.

## Hypothesis A3: The interventions will decrease drop-out rates

* Planned school enrollment (S): Dummy taking the value of 1 if the student intends to enroll in any type of school during the next school year.
* Continued school enrollment (A): Dummy taking the value of 1 if the student is enrolled in any type of school for the next school year.
  + At the time of writing the pre-analysis plan, we are unsure of the quality of administrative data that we will get. Analysis of this outcome is contingent upon receiving reliable data. If partial data is received, we will test whether there is balance across treatment and control groups and only conduct analysis if there is balance.

## Group B: Impact of the interventions on secondary outcomes

### Hypothesis B1: The interventions change students’ learning and studying behaviors

We will test the results on each of the indicators individually and on a standardized z-score indicator of all of the following indicators:

* Has a family member help them with homework or studying (S): isq01
* Time spent studying in a typical week (S): SD21a, noted in minutes
* Number of days missed in previous month (S): SD03
* Uses study strategies presented in the training (s): ISQ02 An item was created to capture each behavior taught in the lessons with the goal of testing whether students in the treatment conditions did each behavior more often than the control condition. This was not intended as tapping a single construct, but rather as an index of frequency of performing intervention-related behaviors. Thus, a cumulative score will be calculated by summing up the responses across the items. If the data are skewed, we will transform the data using logs.

### Hypothesis B2: The interventions change teachers’ mindsets, teaching practices, and perceptions of students

We will begin by exploring measurement models to determine whether the scales below can be aggregated into one or more composite scores. We will retain the measurement model that best balances representing the data and the theory supporting these scales. Then, we will create standardized composite scores based on the higher-order factors of the best measurement model.

Scales:

* Growth mindset
* Attitudes toward failure
* Perception of student’s ability to learn
* Perception of students’ ability to regulate behavior
* Percentage of students who are able to receive a passing grade
* Percentage of students expected to receive a passing grade
* Percentage of students expected to continue studying

### Hypothesis B3: The interventions increase students’ aspirations and expectations for educational attainment and future occupation

We measure aspirations as students’ wishes for their future outcomes in a hypothetical scenario in which they had no constraints. Expectations are measured as students’ expected future outcomes given their current situation. We will test the results on each indicator individually and on three z-scores that capture the overall aspirations, expectations, and aspirations-expectation gap, for both educational attainment and future occupation.

* A z-score of aspirations of educational attainment and future occupation (S): Continuous variable based on the following variables:
  + Aspirations for educational attainment in years (S): Continuous variable taking the following values:
    1. 9 if AP05 = 4 “SMP / MTs” and 13 “don’t want to continue beyond my current level” [The equivalent of general junior secondary]
    2. 12 if AP05 = 5 “SMA/MA” and 6 “SMK” [Senior secondary, general and vocational]
    3. 13 if AP05 = 7 “Diploma 1” [1 year of vocational post-secondary]
    4. 14 if AP05 = 8 “Diploma 2” [2 years of vocational post-secondary]
    5. 15 if AP05 = 9 “Diploma 3” [3 years of vocational post-secondary]
    6. 16 if AP05 = 10 “S1” [Bachelor’s degree]
    7. 18 if AP05 = 11 “S2” [Master’s degree]
    8. 21 if AP05 = 12 “S3” [Doctorate’s degree]
  + Aspirations for an education level that is viewed as prestigious: The prestige of the education level will be determined using the district-level average of control students’ answers to AP19.[[4]](#footnote-5) This variable will show the matched prestige for the education level indicated in AP05.
  + Aspirations for a future occupation requiring a higher education level (S): Continuous variable indicating the median years of education that workers in the aspired to occupation have. To calculate this, we will use the most recently available national labor force survey to calculate the median education level of workers in each 3-digit occupation code. We will then match this statistic with the aspired occupation (AP13; occupation code at the 3-digit level of the ISCO-08 classification) to have an indicator at the student level.
  + Aspirations for a future occupation with higher earnings (S): Continuous variable indicating the median yearly income that workers in the aspired to occupation have. To calculate this, we will use the most recently available national labor force survey to calculate the median yearly income of workers in each 3-digit occupation code. We will then match this statistic with the aspired occupation (AP13; occupation code at the 3-digit level of the ISCO-08 classification) to have an indicator at the student level.
  + Aspirations of a future occupation with higher prestige: Each occupation will be linked to a field of study included in endline question AP20. The district-level average of control students’ answers to that field of study in endline AP20 will be used to measure that occupation’s prestige.[[5]](#footnote-6) This variable will show the matched prestige for the occupation given in AP13.
* A z-score of expectations of educational attainment and future occupation (S): Continuous variable based on the following variables:
  + Expectations for educational attainment in years (S): Continuous variable taking the following values:
    1. 9 years if AP08 = 1 “None,” 4 “SMP / MTs,” and 13 “don’t want to continue beyond my current level” [the equivalent of junior secondary, general]
    2. 12 if AP08 = 5 “SMA/MA” and 6 “SMK” [senior secondary, general and vocational]
    3. 13 if AP05 = 7 “Diploma 1” [1 year of vocational post-secondary]
    4. 14 if AP05 = 8 “Diploma 2” [2 years of vocational post-secondary]
    5. 15 if AP05 = 9 “Diploma 3” [3 years of vocational post-secondary]
    6. 16 if AP08 = 10 “S1” [Bachelor’s degree]
    7. 18 if AP08 = 11 “S2” [Master’s degree]
    8. 21 if AP08 = 12 “S3” [Doctorate’s degree]
  + Expectations of an education level that is viewed as prestigious: The prestige of the education level will be determined using the district-level average of control students’ answers to AP19. This variable will show the matched prestige for the education level indicated in AP08.
  + Expectations of a future occupation requiring a higher education level (S): Continuous variable indicating the median years of education that workers in the expected occupation have. To calculate this, we will use the most recently available national labor force survey to calculate the median education level of workers in each 3-digit occupation code. We will then match this statistic with the expected occupation (AP18; occupation code at the 3-digit level of the ISCO-08 classification) to have an indicator at the student level.
  + Expectations of a future occupation with higher earnings (S): Continuous variable indicating the median yearly income that workers in the expected occupation have. To calculate this, we will use the most recently available national labor force survey to calculate the median yearly income of workers in each 3-digit occupation code. We will then match this statistic with the expected occupation (AP18; occupation code at the 3-digit level of the ISCO-08 classification) to have an indicator at the student level.
  + Expectations of a future occupation with higher prestige (S): Each occupation will be linked to a field of study included in endline question AP20. The district-level average of control students’ answers to that field of study in endline AP20 will be used to measure that occupation’s prestige.[[6]](#footnote-7) This variable will show the matched prestige for the occupation given in AP18.
* A z-score of the gap between aspirations and expectations for education and future occupations (S):[[7]](#footnote-8) Continuous variable based on the following variables:
  + Students have lower expectations of their future education than their desired level with no constraints (S): Dummy variable taking the value of 1 if AP08 is for a degree that requires fewer years of education than the degree indicated in AP05
  + Gap in years between aspirations and expectations in educational attainment (S): Continuous variable equal to the difference between AP05 (aspired) and AP08 (expected).[[8]](#footnote-9) The number of years of education for AP05 and AP08 will be defined as described above.
  + Difference in median education levels between aspired to and expected occupations (S): As above, the median education levels for each occupation will be calculated using the most recently available national labor force survey. This indicator will be a continuous variable showing the difference between the median education level of the aspired to occupation and the median education level of the expected occupation.
  + Students expect to be in a career with lower median education levels than their desired occupation (S): Dummy variable taking the value of 1 if the matched median education level of their expected occupation (AP18) is lower than the matched median education level of their desired occupation (AP13)
  + Difference in median yearly income between aspired to and expected occupations (S): As above, the median yearly income levels for each occupation will be calculated using the most recently available national labor force survey. This indicator will be a continuous variable showing the difference between the median yearly income level of the aspired to occupation (AP13) and the median yearly income level of the expected occupation (AP18).
  + Students expect to be in a career with lower earnings than their desired occupation (S): Dummy variable taking the value of 1 if the matched yearly income level of their expected occupation (AP18) is lower than the matched yearly income level of their desired occupation (AP13)
  + Not expecting to get a degree in preferred subject (S): Dummy taking the value of 1 if AP10 is “No” or “Don’t know” and 0 if “yes.”

### Hypothesis B4: The interventions change students’ career choices

We will test the results on each indicator individually and on 4 z-scores.

* A z-score indicating increased aspirations and expectations to study in a technical track, based on the following variables:
  + Actual enrollment in vocational schools (A): Dummy value taking the value of 1 if the student is enrolled in SMK (Sekolah Menengah Atas; Vocational junior secondary, in English) for the next school year and 0 otherwise.
  + Planned enrollment in vocational schools (S): Dummy value taking the value of 1 if the student is planning on enrolling in SMK for the next school year and 0 otherwise.
  + Aspires to a vocational education track (S): Dummy taking the value of 1 if AP05 = 6,7,8,9 and 0 otherwise.
  + Expects a vocational education track (S): Dummy taking the value of 1 if AP08 = 6,7,8,9 and 0 otherwise.
* A z-score indicating a change in aspired and expected field of study:
  + Change in likelihood of getting a degree[[9]](#footnote-10) in math, physical or natural sciences
  + Change in the likelihood of getting a degree in medicine or psychology
  + Change in the likelihood of getting a degree in business, economics or management
  + Change in the likelihood of getting a degree in law
  + Change in likelihood of getting an education
  + Change in the likelihood of getting a degree in computer science, technology, or engineering
  + Change in the likelihood of getting a degree in political science or related
  + Change in the likelihood of getting a degree in religion, language or cultural studies
  + Change in the likelihood of getting a degree in agribusiness or agro-technology
  + Change in the likelihood of getting a degree in communication, arts, tourism or community welfare
  + Expected high likelihood of pursuing a STEM (science, technology, engineering, and mathematics) field of study: Dummy taking the value of 1 if AP09a=4 (“Likely”) or 5 (“Very Likely”), or AP09b=4 or 5, or AP09f=4 or 5.
* Change in job aspirations
  + Change in aspired occupation (S): Dummy taking the value of 1 if the value of AP13 is different at endline than at baseline. If the answer is “don’t know” in one period and completed in another period, this is considered a change. If missing in either period, the dummy is missing.
  + Change in expected occupation (S): Dummy taking the value of 1 if the value of AP18 is different at endline than at baseline. If the answer is “don’t know” in one period and completed in another period, this is considered a change. If missing in either period, the dummy is missing.
  + Aspired occupation is in STEM (S): AP13= 211, 212, 213, 214, 215, 216, 251, 252, 311, 312, 314, 315, 351, or 352.[[10]](#footnote-11)
  + Expected occupation is in STEM (S): AP18= 211, 212, 213, 214, 215, 216, 251, 252, 311, 312, 314, 315, 351, or 352.
  + Would like to work outside the home if no constraints (S): Endline AP13
  + Expects to work outside the home (S): Endline AP17

## Hypothesis group C: Process and mechanisms of change

In this section, we will focus on how the training could generate changes in primary outcomes, as specified in hypothesis group A.

We will test mediational hypotheses to identify mechanisms of change using the final composites from Hypothesis A2 as the mediators of achievement, (as measured in A1), aspirations, (as measured in B3), and behavior outcomes (PERC, missed days of school). Lastly, we will test teacher behaviors (as measured in Hypothesis B2) as mediators of achievement (as measured in A1) and aspirations (as measured in B3).

## Hypothesis group D: The interventions will have different effects on different types of students

As there is a gender focus of this impact evaluation, we will test the hypothesis D1 on all of the hypotheses in groups A, B and C. Moreover, the sample was constructed in a way to be powered to detect results disaggregated by gender. However, due to the risks of multiple hypotheses testing, for each of the hypotheses D2-D5 mentioned below, we will test the heterogeneous effects of the interventions on the primary outcomes specified in hypothesis group A above. For the characteristics that have statistically significant heterogeneous effects on primary outcomes, we will also test the heterogeneous effects of the training on the secondary outcomes specified in hypothesis group B and the mechanisms specified in hypothesis group C above.

### Hypothesis D1: The program has different impacts on girls than boys.

Heterogeneity based on SD02.

### Hypothesis D2: The program has different impacts on individuals from different socio-economic backgrounds.

Heterogeneity based on an index of assets in the student’s home at baseline. The index will take the first principal component of the following variables of the question SD10 ("Which of the following are in your home?”):

1. A television
2. A car, van or truck
3. A washer
4. A refrigerator or freezer
5. A stove or burner for cooking
6. Cell phones with Internet access (e.g. smartphones)
7. A computer (desktop computer, portable laptop, or notebook)
8. Musical instruments (e.g. guitar, piano, kendang, seruling)

Heterogeneity analysis will be based on this continuous variable. While the continuous variable will be our primary specification, we will test for a non-linear relationship using quartiles, dividing the sample at the 25th percentile or below, between the 25th and 75th percentiles, and at or above the 75th percentile. We will use ANOVA with pair-wise comparisons to conduct the analysis, and if we see the impacts are non-linear we will consider the ANOVA regressions as the primary specification.

### Hypothesis D3: The program has different impacts based on the attitudes of teachers in their school at baseline.

We will use the final composites from Hypothesis B2 and take the average of the baseline teachers composites in the school. Heterogeneity analysis will be based on this school-level continuous variable. Quartiles will be created to examine any effects found, grouping the sample at the 25th percentile or below, between the 25th and 75th percentiles, and at or above the 75th percentile. We will use ANOVA with pair-wise comparisons to conduct the analysis, and if we see the impacts are non-linear we will consider the ANOVA regressions as the primary specification

### Hypothesis D4: The program has different impacts on students in high, medium and low performing schools.

Heterogeneity based on the average school-level 9th grade national test scores from the exams taking in the preceding school year (April-May 2018). Following Yeager et al. 2018,[[11]](#footnote-12) we will divide schools into low performing schools, defined as having average test scores in the bottom 25%[[12]](#footnote-13) of schools in Java and Sumatera; medium-performing schools, defined as having average test scores between the 25th and 75th percentiles[[13]](#footnote-14) of schools in Java and Sumatera; and high-performing schools, coded as being greater than or equal to the 75th percentile of schools in Java and Sumatera

### Hypothesis D5: The program has different impacts based on students’ initial academic performance.

Heterogeneity based on a dummy variable taking the value of 1 if the student’s average classroom grades (average of SD06 a through g) at baseline were greater than or equal to the median for their school.

* We expect that the impact of the intervention will be greater for lower-achieving students, consistent with previous findings (e.g., Yeager et al. 2018[[14]](#footnote-15)).
* However, we will also test baseline achievement as a continuous moderator of the treatment-academic performance relation using linear and quadratic baseline grades that have been standardized at the school level. We will test this alternative model because low achieving in Indonesia might be more equivalent to moderate achieving in the US. Thus, it is plausible that the impact could be strongest with moderate levels of baseline achievement compared to low or high (e.g., an inverted U-shape).

### Hypothesis D6: The program has particularly strong impacts on low achieving students in medium performing schools

We will test the generalizability of the findings from Yeager et al. (2018) in the Indonesia context. Yeager et al. (2018) created three groups based on school-level achievement (bottom 25%, middle 50%, and top 75%) and examined the impact of their growth mindset intervention on low achieving (bottom 50% within each school) students. They hypothesized the impact would be greatest on low achieving students in the middle 50% schools, but found that the effect was strongest for the students in the bottom 25% and middle 50% compared to the top 25% of schools. We hypothesize that with the higher rates of poverty in Indonesia, we will find support for the original hypothesis for lower achieving students in the middle achievement (50%) schools.

### Hypothesis D7: The program has different impacts on students in schools with high and low UN test score variance

Heterogeneity based on the school-level standard deviation (s.d.) levels interacted with school-level means of 9th grade national test scores from the exams taken in the preceding school year (April-May 2018). There is reason to hypothesize we may find differential effects in school environments with low-performing students whom are learning and challenged in classrooms with very high-performing students, and if this translated to higher school-level standard deviation in UN test scores, so long as the socio-emotional learning curriculum is understood as a tool(s) to overcome challenge. We will also analyze heterogeneity of the interaction of high and low means with high and low s.d.’s in the school-level test scores to control for average school performance. We will divide schools in high and low mean and high and low s.d. at the 50% percentile, following the methodology used in the forthcoming paper on the impacts of the predecessor program, Semua Bisa Pintar, which tested the impact of a similar but shortened 2-week curriculum on UN test scores. The analysis looked at the different impact across high mean-high s.d., high mean-low s.d. low mean- high s.d. low mean-low sd schools.

# Methodology

## Identifying the treatment effect

All analysis will be done at the individual-level unless otherwise specified above. For the variables for which we have baseline values, we will calculate the intent to treat effect using the following ANCOVA estimation:

Yist = β0 + β1 T1s + β2 T2s + β3 Yist=0 + β4 Ds +εis

(E1)

Where:

Yist is the outcome variable measured for individual *i* in school *s* at time *t* post-treatment.

T1s and T2s are dummy variables taking the value of one if the school was randomly assigned to treatment 1 or treatment 2 respectively.

β1 and β2 will measure the intent-to-treat effect of being assigned to the treatment 1 and treatment 2 respectively, compared to the control group.

Yist=0 is the baseline value of the outcome variable

Ds is district controls

εis is the error term

For the outcomes for which we do not have baseline values, we will calculate the intent to treat effect using the following regression:

Yist = β0 + β1 T1s + β2 T2s + β3 Ds +εis,

(E2)

The interpretations of the variables are the same as for the previous equation.

For both equations, standard errors will be clustered at the school level. In addition to calculating the intent to treat effect, we will also estimate the treatment on the treated effect by instrumenting the participation in the training program with the random assignment to the treatment group. This estimate will enable us to control for non-compliance with treatment assignment.

We have four sources of compliance data:

1. Data from the field coordination team who called schools during the intervention
2. Photos that schools sent to the field coordination team during the intervention
3. Questions asked of teachers and counselors during the endline survey
4. Notes from the enumerators and supervisors during the endline survey

We are currently triangulating the data from the different sources; however, we believe the survey data to be the most accurate and have the fewest missing values (data type 3). As such, at the time of publishing the pre-analysis plan, the team intends to use the answers from the teacher and counselor surveys as the primary specification of compliance. Three school personnel were interviewed in each school. If there is a discrepancy between the different respondents, we will use the response given by two out of the three respondents.

## Dealing with multiple outcomes

There are two challenges when working with a large number of outcome variables. First, the volume of outcome variables can make it more difficult to interpret results. Second, analysis of multiple outcomes can increase the risk of type 1 error if the significance tests are not adjusted appropriately.

To deal with these risks, we will employ the following strategies:

1. We have clearly identified a set of primary outcomes in hypothesis group A.
2. We will follow the methodology of Kling, Katz and Liebman (2007) to test the significance of families of outcomes in a single aggregate. For each family of outcomes described above, we will:

1) Convert all outcomes so that the sign of all of the variables in a family goes in the same direction

2) Calculate the z-score of each variable by subtracting the control group mean and dividing by the control group standard deviation

3) Take an average of the z-scores in the family

1. We consider not just statistical significance, but also practical significance. Given the benchmarks are based on US studies, and primarily smaller studies with less diversity, we have not set an absolute cutoff for effect sizes. We will translate significant effects to practical significance and reference by referencing comparisons to effect sizes from other programs and interventions focused on increasing learning and achievement outcomes and discussing the policy and development implications of the magnitude of impacts achieved.

## Estimating heterogeneous treatment effects

To test for the heterogeneity of effects, we will interact the variable of interest with the treatment status in equation E2 and with the treatment status and lagged dependent variable in equation E1.

## Addressing missing data and questions with limited variation

### Survey attrition

If Ais represents whether individual i from school s attrite from the study because the individual cannot be found or the individual or school refuses to participate, we will estimate the following equation to test whether survey attrition is related to treatment status:

Ais = β0 + β1 T1s + β2 T2s + β3 Ds +εis

(E3)

where variable definitions are similar to E1 above.

If treatment status does not affect survey attrition at the 5% significance level, then we will not adjust the estimates for attrition.

If treatment status does have a statistically significant effect on survey attrition, we will test the robustness of our results using Lee bounds (Lee 2008).

### Item non response

We will test to see if item non-response is statistically significantly related to treatment status using the same methodology as in the survey attrition section. If the treatment status does not have an effect on item non-response at the 5% level, then no corrections or imputations for the values will be made. If the treatment status does affect item non-response, then we will test the robustness of our results using Lee bounds.

1. Because some districts have less than 35 total eligible schools, the sampling is not perfectly representative; however, it is close to representative. [↑](#footnote-ref-2)
2. See the previous footnote, this section describes the way it works in theory. However, given the limited number of schools in some districts, the proportions are slightly different in our sample. [↑](#footnote-ref-3)
3. At the time of pre-analysis plan registration, endline data collection had been completed; however, the data had not yet been received. Data entry is ongoing, and the authors of the pre-analysis plan have not yet seen even preliminary databases. [↑](#footnote-ref-4)
4. Because this question was not included at baseline and may have been influenced by the intervention, we use the control group average for that district. The average is done at the district level to adjust for regional differences in the prestige of education. [↑](#footnote-ref-5)
5. Because this question was not included at baseline and may have been influenced by the intervention, we use the control group average for that district. The average is done at the district level to adjust for regional differences in the prestige of occupations. [↑](#footnote-ref-6)
6. Because this question was not included at baseline and may have been influenced by the intervention, we use the control group average for that district. The average is done at the district level to adjust for regional differences in the prestige of occupations. [↑](#footnote-ref-7)
7. We expect that several indicators in this index will contain many 0 values, so we will use a zero-inflated model for these indicators. [↑](#footnote-ref-8)
8. Note that in the baseline survey, 4 percent of students expected to complete an educational level that required more years to complete than the level they aspired to. [↑](#footnote-ref-9)
9. This variable and all other similar variables will be constructed in the following way: a dummy taking the value of 1 if the answer to the corresponding 5-level Likert variable of the question AP09 (“How likely would you be to get a degree in each of the following fields of study?”) increases or decreases by at least 2 between baseline and endline. For example, if at baseline the value was 1 (“Not at all likely”) and at endline the value was 3 (“Somewhat likely”), the dummy would be coded 1. Likewise, if the baseline value was 3 and the endline was 1, the dummy would be coded 1. However, if the baseline value was 3 and the endline was 2 (“Not very likely”), the dummy would be coded 0. [↑](#footnote-ref-10)
10. 3-digit ISCO occupation codes classified as STEM fall under to the four 2-digit ISCO occupation categories classified as such: Science and engineering professionals (ISCO 21), Information and communications technology professionals (ISCO 25), Science and engineering associate professionals (ISCO 31), Information and communications technicians (ISCO 35)). [↑](#footnote-ref-11)
11. Yeager, D. S., Hanselman, P., Paunesku, D., Hulleman, C., Dweck, C., Muller, C., … Duckworth, A. (2018, March 1). MANUSCRIPT UNDER REVISION: Where and For Whom Can a Brief, Scalable Mindset Intervention Improve Adolescents’ Educational Trajectories?. <https://doi.org/10.31234/osf.io/md2qa>. [↑](#footnote-ref-12)
12. Coded as being less than or equal to the 25th percentile [↑](#footnote-ref-13)
13. Coded as exclusive of the 25th and 75th percentile [↑](#footnote-ref-14)
14. Yeager, D. S., Hanselman, P., Paunesku, D., Hulleman, C., Dweck, C., Muller, C., … Duckworth, A. (2018, March 1). MANUSCRIPT UNDER REVISION: Where and For Whom Can a Brief, Scalable Mindset Intervention Improve Adolescents’ Educational Trajectories?. <https://doi.org/10.31234/os.f.io/md2qa>. [↑](#footnote-ref-15)