Indoor air quality and student welfare
The Effect of Indoor Air Purifiers in Schools
(Pre-analysis Plan)*

Jacopo Bonan[†] Francesco Granella[‡] Stefania Renna[§] Luis Sarmiento[¶]

August 21, 2023

Bad air quality is a significant problem for children, as it can cause various clinical and subclinical problems, including respiratory infections, asthma, allergies, absenteeism, and cognitive impairment. In this pre-analysis plan, we present the design of a *cluster randomized control trial* on the potential benefits (and cost-effectiveness) of installing air purifiers in schools to reduce children's exposure to poor air quality conditions. We randomly assign 95 classes in three schools to receive or not air purifiers and estimate their effects on indoor air pollution, absenteeism, achievement, cognitive ability, and behavioral outcomes related to mood and aggression. We expect to find a relevant increase in school attendance, learning, cognitive outcomes, and the general well-being of children. The results of this study would allow policymakers to understand the benefits of a scalable defensive strategy to mitigate the exposure of vulnerable groups to a relevant environmental stressor.

1. Introduction

In this study, we analyze the potential benefits and cost-effectiveness of one of these solutions: portable air purifiers in schools. Our main hypothesis is that air purifiers improve indoor air quality in classrooms. This is expected to lead to improvements in learning outcomes and overall well-being. Analyzing the effect in schools is particularly relevant as young children spend most of their awake time in the classroom. To test the effect of purifiers, we propose a cluster *Randomized Control Trial* (RCT), where we randomly assign 95 classes in three schools to treatment and control groups. We stratify classes by school and grade and install 43 consumer-grade air purifiers in the treated classes. Air purifiers contain a certified ultra-low penetration air

*The project received ethical approval from Politecnico di Milano IRB (code 18/2023).

†Politecnico di Milano & RFF-CMCC: European Institute on Economics and the Environment (EIEE), Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC).

‡RFF-CMCC: European Institute on Economics and the Environment (EIEE), Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC).

§Politecnico di Milano & RFF-CMCC: European Institute on Economics and the Environment (EIEE), Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC).

¶RFF-CMCC: European Institute on Economics and the Environment (EIEE), Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC).

filter (ULPA-U15) that removes up to 99.99% of particles larger than 0.026 microns. In addition to the purifiers, we also install low-cost indoor air quality monitors in a subsample of 38 classes. The monitors collect granular data on a series of pollutants and atmospheric conditions.

Highly effective filters such as high-efficiency particulate air (HEPA) and ULPA filters trap airborne pollutants, pathogens, and allergens, such as *fine particulate matter* ($PM_{2.5}$), mold, viruses, bacteria, and pollen. As such, we do not equate the effects of the purifiers with the effects of lower concentrations of air pollutants. Throughout the rest of the study, we use the term *air quality* to encompass the concentration of pollutants and non-pollutant pathogens and allergens in the air.

We examine the effects of purifiers on absenteeism, standardized test results, cognitive abilities, and behavioral outcomes related to mood and aggression by using data from administrative and survey sources. Air pollution values come from indoor air quality monitors installed on a representative sample of classes. Data on students' daily absences come from school records. The Italian National Institute of Evaluation of the Education System (INVALSI) provides data on standardized test scores in language and mathematics. All other variables regarding cognitive tests and behavior come from three questionnaires taken by all students before installing the purifiers, and after four and eight months, that is, at the beginning, middle, and end of the school year.

2. Research Design

2.1. Outcomes and hypothesis

The study's main objective is to examine air purifiers' effects on air pollution, children's absenteeism, cognition, learning, mood, and aggressive behavior. We hypothesize that the improvement in air quality due to the installation of the purifiers would translate into a series of quantifiable benefits for the student population.

Primary outcomes:

First, we examine the effect of purifiers on school absences. Each school provides anonymous administrative records on students' daily absences for one year before and after the intervention. We hypothesize that the installation of air purifiers would translate into a reduction in school absences through the health benefits of improved air quality. Absences are defined with a dummy variable taking the value of one if the student is absent on a given day and zero otherwise.

Hypothesis 1: *Improved indoor air quality reduces absences*

The ancillary assumption is:

Hypothesis 2: *Air purifiers improve indoor air quality*

We evaluate indoor air quality with data from 38 indoor air quality monitors measuring the concentrations of various pollutants at 15-minute intervals. We focus on $PM_{2.5}$ as our main air pollution indicator and use the average daily concentration to estimate the effect of the purifiers.

Then we look at the effect of air purifiers and the expected improvement in air quality on student cognitive abilities and school performance.

Hypothesis 3: *Improved indoor air quality improves students' cognitive ability*

We measure cognitive ability using the Raven Colored Progressive Matrices Test administered to all students at three points in time: at the beginning of the year before purifiers are installed, in the middle of the year, and at the end of the intervention. Every test consists of 18 matrices of progressive difficulty. We sum every correct answer into a score ranging from 0 to 18 and standardize the score to have a mean of zero and a variance of one for each school-grade combination by using the mean of the control group and its standard deviation.

Hypothesis 4: *Improved indoor air quality improves achievement*

We approximate the learning effects of our intervention by looking at the scores of the national standardized test (INVALSI). Second- and fifth-grade students from all Italian schools take the test every spring. We use language, mathematics, and the overall test scores (the sum of the two) and standardize them by grade and subject, using the mean and standard deviation of the control group. The score for each test ranges between 0 and 50. The overall INVALSI score is the sum of language and mathematics scores and ranges from 0 to 100.

Third, we look at students' behavioral responses to improved air quality. In particular, we look at self-reported mood and aggressive behavior.

Hypothesis 5: *Improved indoor air quality improves students' mood and decreases aggressive episodes*

We measure students' mood through a survey question using a Likert scale (very good, good, bad, very bad) and a question inquiring how students felt over the last week (very nice, generally nice, sometimes nice, not very nice). We aggregate them in an index for each student using the strategy suggested in Anderson (2008) and standardize it by survey wave. We then proxy the occurrence of aggressive episodes by generating a dummy variable taking the value of one if students report having argued or quarreled with any classmate over the previous week (sometimes, often, very often) and zero otherwise (never).

Secondary outcomes:

We focus on secondary outcomes to confirm the above hypotheses, perform robustness checks, and shed light on potential mechanisms.

We examine subjective health status to confirm that health is the key channel through which air quality leads to fewer absences. First, we construct a variable, taking the value of one if the student declares that she was sick or very sick (from a four-item Likert scale) during the last week and zero otherwise. This variable is a proxy for the extensive margin of self-reported health.

Second, we construct a variable that sums up the symptoms declared by the students over the past week. In particular, we focus on respiratory system-related symptoms (running nose, closed nose, sneezing, cough, shortness of breath, headaches, tiredness) and sum the symptoms experienced at least once over the previous week from a four-item Likert scale. The resulting variable runs from zero to six and is a proxy for the intensive margin of self-reported health.

Third, we use two questions about health problems that are not (traditionally) related to air quality to perform placebo tests. First, we use a variable to account for the occurrence of accidents or falls. Second, we construct a variable to report stomach pains. In both cases, we obtain dummy variables, taking the value of one if these episodes occurred at least once during the previous week and zero otherwise.

We use secondary measures of student performance to complement Hypothesis 4. For this, we use non-standardized student average scores computed at the end of each semester. In particular, for each topic, a series of grade-specific learning goals are defined. Teachers assess students using a qualitative scale (first acquisition, basic, intermediate, advanced). We focus on two topics: Italian language and mathematics. For each topic, we transform the qualitative assessment into a 1-4 scale, take the average across learning goals, and standardize by grade and semester. Unlike the INVALSI tests, these scores are available for the entire study sample for the pre and post-treatment period.

Regarding air pollution, in addition to $PM_{2.5}$, our monitors collect data on *carbon monoxide*

(CO) and *volatile organic compounds* (VOCs). We complement the $PM_{2.5}$ estimates by running similar models for the effects of purifiers on VOCs and CO.

Another block of secondary outcomes refers to the potential differential reaction of teachers to the presence or absence of air purifiers. Our intervention is purely technological and is not accompanied by any information or awareness campaign. However, we cannot rule out the possibility that some information is provided by teachers and that they behave systematically differently in response to treatment. First, teachers can implement different defensive behaviors in treated and control classes. For example, they could open windows to ventilate class air with different frequencies, depending on perceived levels of outdoor air pollution and the presence of air purifiers. We proxy this behavior using patterns in *carbon dioxide* (CO_2) concentration, which should identify the daily number of opening window episodes. Second, teachers can communicate air quality-related messages to students as a consequence of their change in environmental awareness and perception of the importance of air pollution. We assess students' environmental concerns and awareness through two questions. First, children are asked to assess the importance of general urban problems (lack of green space, traffic, poor littering), among which air pollution is listed. Second, children assess air quality in different locations: in general, in the city, in their classroom, and in the school courtyard. We construct a dummy equal to one if students assess air pollution as a very important problem and zero otherwise. Second, we construct a dummy equal to one if the students report that the air quality in class is excellent and zero otherwise.

Table 1 summarizes the primary and secondary variables, their source, observation unit, and measurement.

Table 1: Outcome variables

Hyp	Variable	Unit of obs.	Type	Source
Panel A: Primary outcomes				
1	Absences	Student-day	Dummy	School data
2	PM2.5	Class-day	Levels	Monitors
3	Cognitive level	Student-round	Std. score	Survey
4	Italian INVALSI test score	Stundent-year	Std. score	INVALSI
4	Math INVALSI test score	Stundent-year	Std. score	INVALSI
5	Mood	Student-round	Anderson index	Survey
5	Aggressive episodes	Student-round	Dummy	Survey
Panel B: Secondary outcomes				
1	Subjective health status (extensive)	Student-round	Dummy	Survey
1	Subjective health status (intensive)	Student-round	Score (0-6)	Survey
1	Subjective health status (placebo)	Student-round	Dummy	Survey
2	CO	Class-day	Levels	Monitors
2	VOC	Class-day	Levels	Monitors
4	Italian grade	Student-semester	Std. score	School data
4	Math grade	Student-semester	Std. score	School data
6	Env. awareness (air pollution problem)	Student-round	Dummy	Survey
6	Env. awareness (air quality)	Student-round	Dummy	Survey
6	N. of times windows are opened	Class-day	Levels	Monitors

2.2. Intervention and randomization

Our intervention involves installing 43 consumer-grade air purifiers in randomly selected classes in three schools. We randomly assign about half of the classes to treatment or control, while stratifying them by school and grade. Randomization is performed by researchers in the presence of each school's director using a random number generator. All purifiers will be installed in early October 2023 while the school is closed, e.g., over the weekend. The intervention is not accompanied by specific information campaigns or communications to teachers or parents. Note that we install the purifiers after administering the first survey. All purifiers operate continuously throughout the study period between October 2023 and June 2024.[1]

The purifiers contain a ULPA-U15 filter that filters up to 99.99% of particles larger than 0.026 microns. The ULPA filtration classification (Ultra Low Particulate Air) is the maximum level of efficiency achievable in mechanical filtration technology (see the Appendix Table ??).[2] The

[1]After discussing with directors, we decided to have them always run, except for the long breaks at Christmas and Easter. This should limit partial compliance and selective issues related to the responsibility to turn them on and off. The research team monitors the functioning of the purifiers through weekly and monthly statistical analyzes of indoor air pollution data in a subsample of classes.

[2]This classification is certified and recognized worldwide, where ULPA is the highest level of the international HEPA classification and is up to 10 to 100 times more efficient than the better-known HEPA filters.

purifiers have a very low energy consumption, similar to an incandescent bulb of 40-60 watts, making them ideal for large-scale deployment. Furthermore, they are relatively silent, with average acoustic pressure levels between 29 and 45 A-weighted decibels. The purifier passes polluted air through the ULPA filter, which is made of layers of ultrafine materials capable of blocking fine particles, and an activated carbon filter that purifies the air before it is returned to the environment. The cost of purifiers equipped with ULPA filters and suitable for the volume of school classes can reach up to 2,500 euros. Other consumer-grade products with lower filtration, such as HEPA, can have a lower price ranging from 500 to 1800 euros.[3]

In addition to the purifiers, we randomly install 38 indoor air quality monitors in a random subsample of classes. We stratify by school, treatment status, and class characteristics such as floor and orientation, that is, whether the class faces the street or the internal courtyard.[4] This should ensure an adequate representation of treatment and control classes facing possible local differences in exposure to outdoor pollution. The monitors measure the concentration of CO_2, VOCs, $PM_{2.5}$, CO, temperature, humidity, and atmospheric pressure at sub-hour intervals. Once connected to electricity and the internet, the monitor saves and stores 15-minute measurements on an online data platform accessible to the researchers.[5] The monitors will be installed in selected classes in October 2023.

2.3. Sample and data

We collaborate with three comprehensive school institutes in Milan.[6] Each school has an average of 32 classes divided into five grades and about 22 students per class. Our study sample includes 95 classes and approximately 2,035 students (see Table 2).

Students' daily absences and end-of-semester grades for each school institute are collected digitally and shared with the researchers after anonymization at the end of the study year. They are available for the academic years 2022-23 and 2023-24.

INVALSI tests are national and are taken every spring by second- and fifth-grade students in primary schools. All Italian students take the tests on two fixed dates within the same week of May; the test is designed to be objective and standardized, allowing comparisons between

[3]The change of filters is the only maintenance required. Depending on the model, it can be done yearly or every two years. The cost of filters can range from 50 to 150 euros.

[4]The monitors are Envira Nanoenvi IAQ.

[5]The indoor air quality monitors have a small LED display color-coded on four levels based on the Indoor Ambient Air Quality Index defined by the manufacturing company. We cover the LED lights with anti-tampering tape to increase the comparability of classes with and without monitors and to avoid stimulating differential behavioral responses.

[6]From now on, for the sake of simplicity, we refer to them as "schools".

Table 2: School descriptives

School	Classes	Students
School 1	35	835
School 2	40	735
School 3	20	400
3	95	2,035

students and schools. After formal authorization from schools, we obtain anonymized data for the Italian language and mathematics tests in 2024 from INVALSI.

Teachers administer paper and pencil surveys to students at three points in time: before the launch of the intervention, that is, at the end of September 2023, in February 2024, and in May 2024. Teachers can flexibly choose the survey administration date within a two-week window, depending on their availability. To facilitate comprehension at all levels, we use capital letters and a visual scale with facial expressions when using Likert scales. First- and second-grade teachers follow a dedicated protocol in which survey questions are projected and read aloud so that students can follow them more easily. The survey includes four blocks and is expected to take about 15 minutes. Appendix Section B shows the survey.

The first block measures cognitive abilities through the Raven Colored Progressive Matrices Test (cards 1-18), a non-verbal intelligence test designed to assess individuals' abstract reasoning and problem-solving abilities (Raven and Court, 1998). The test has been widely used to measure fluid intelligence in various contexts (see Dean et al., 2017; Mani et al., 2013). It consists of 18 visual patterns with one missing piece. The task is to identify the missing element from several options. We draw the suitable version for children between 5 and 12 years of age from the first wave of the Mexican Family Life Survey (Rubalcava and Teruel, 2006).

The second block is dedicated to subjective health status (items 1-3). On the one hand, children are asked to assess their general health status over the last week on a four-item Likert scale (very sick, sick, a bit sick, not sick) (item 1) and, on the other hand, to specify any respiratory and non-respiratory symptoms from within a list (item 2). We draw such questions from a validated survey on acute respiratory diseases designed for children aged 4 to 10 years designed by Schmit et al. (2021), although adapting them to a different temporal span. To check for injury-related illnesses, we add a survey item that asks whether subjects have had injuries or accidents in the past week (item 3).

For the third block, which refers to mood state and behavior (items 4-5), we derive insights from

Sacchi et al. (2023). We ask the children to assess their own and their classmates' moods and feelings during the past week. We also ask the children to report whether they have experienced quarrels and disagreements with classmates during the week before the questionnaire.

Finally, the block on environmental awareness (items 6-9) is inspired by Cori et al. (2020). We collect information on home school routes. In particular, we ask about the typical mode of transport (item 6) and the length of the daily commute from home to school (item 7). Then we ask the students for their subjective assessment of common urban problems, including air pollution (item 8) and how they perceive air quality in the city and at school (item 9).

2.4. Timeline and implementation

Between January and July 2023, we surveyed schools for participation, started the tendering process among various manufacturers of air purifiers and air quality monitors, and began collecting the initial data on school absences. In May 2023, we obtained approval from the Institutional Review Board (IRB) of Politecnico di Milano. In October 2023, we will install air monitors in selected classes. Between the end of September and the beginning of October 2023, we will administer the baseline surveys to students. Following this, the air purifiers company and the research team will install the purifiers in randomly selected classrooms. We will administer midline and end-line surveys in February and May 2024. The schools will provide data on absentees and grades in the summer of 2024. We expect that INVALSI score data will be available to researchers in December 2024. The complete data analysis will be available by February 2025.

Figure 1 presents the timeline of our study.

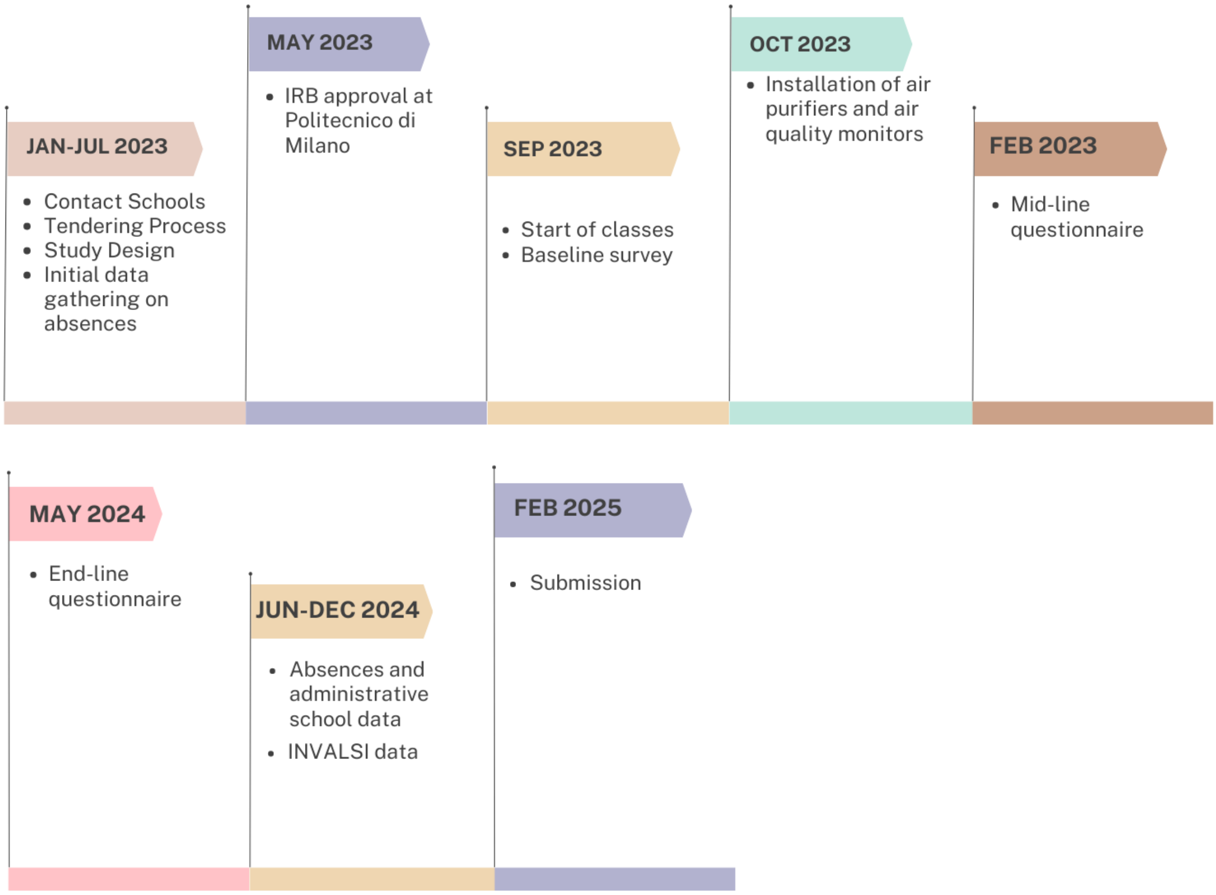


Figure 1: Expected timeline of the intervention

3. Analysis

3.1. Empirical strategy

Our empirical strategy relies on the randomization of air purifiers across classes. Students in treated classes are expected to experience better air quality than those in the control group. This should lead to better outcomes related to air quality exposure. The identifying assumptions are that treated and controlled students are similar in all unobservable and observable characteristics. We also assume that control subjects do not indirectly benefit from treatment or react to the absence of purifiers changing their behavior. We discuss these potential threats to identification in Section **??**. In what follows, we report on the statistical models that we use to test our hypotheses.

Hypothesis 1: Absences We test whether air purifiers affect absences with a linear probability model that compares absence records from students in treated and control classes for the year before and after the installation of the purifiers. Equation 1 presents the econometric specification to examine the effect of purifiers on absences.

$$Absent_{ict} = \beta_1 Post_t + \beta_2 AirPurifier_c * Post_t + \lambda_i + \lambda_t + \varepsilon_{ict} \quad (1)$$

In it, $Absent_{ict}$ is equal to one when the student i in class c on date t is absent and zero otherwise. $AirPurifier_c$ is an indicator variable that takes a value of one when the class c of the student i has an air purifier. $Post_t$ takes the value of one after the air purifiers are installed and zero otherwise. λ_i and λ_t are student and time-fixed effects, allowing us to control for individual time-invariant characteristics and time trends.[7] ε_{ict} is an idiosyncratic error term. To consider the correlation of absences between students in the same class, we cluster standard errors at the treatment level (Abadie et al., 2023), i.e., by class.

Hypothesis 2: Air pollution We test the effects of the purifiers on air pollution with an OLS model that estimates the difference in mean $PM_{2.5}$ for classes with and without the purifiers on the sample of classes equipped with air quality monitors. Differently from school absences, we only estimate the difference in means instead of the difference in the difference because we do not have pre-treatment information on average indoor air pollution. Equation 2 presents the empirical strategy to estimate the effect of purifiers on $PM_{2.5}$. In it, PM_{ct} is the concentration measured in class c on day t as a function of an indicator variable for the treated classes ($AirPurifier_c$) and fixed effects for the date of observation (λ_t).[8]

$$PM_{ct} = \alpha + \beta AirPurifier_c + \lambda_t + \varepsilon_{ct} \quad (2)$$

Hypothesis 3: Cognitive skills We study the effect of the purifier on cognitive abilities by looking at its effects on the Raven test scores with an OLS estimator of the form:

$$Raven_{ict} = \beta_1 Post_t + \beta_2 AirPurifier_c * Post_t + \lambda_s + \lambda_t + \varepsilon_{ict} \quad (3)$$

In it, $Raven_{icst}$ is the normalized score (mean zero and standard deviation one) of the student i, in class c, in strata s, and in the survey round t. λ_s are strata fixed effects, λ_t period fixed effects,

[7] We opt for individual instead of strata fixed effects because fixed effects at the student level allow us to capture relevant between-individual variation in absences. Ultimately, this grants us higher statistical power.

[8] In our power calculations, we also tested for the possibility of strata (school and grade) fixed effects allowing us to account for the randomization method (Bruhn and McKenzie, 2009). The power of our empirical strategy is substantially higher with strata fixed effects (i.e., circa 1%). We opt for a simpler difference in the means model because it is substantially more conservative.

and ε_{ict} the error term. Note that for all survey questions, we build a panel from three waves (baseline, midline, and end-line). The baseline survey is administered before the installation of the air purifiers; as such, $Post_t$ equals zero, while the midline and the end line are taken after the installation ($Post_t = 1$).

Hypothesis 4: Test scores We examine the impact of the intervention on standardized INVALSI scores with a difference in means model by taking the INVALSI scores for a subsample of grades (second and fifth) after the intervention. Equation 4 presents the econometric model to estimate the effect of air purifiers on the INVALSI scores. In it, $INVALSI_{ic}$ is the normalized language, mathematics, and overall test score of the student i in class c. All other variables on the right-hand side are equivalent to the previous models.

$$INVALSI_{ic} = \alpha + \beta AirPurifier_c + \lambda_s + \varepsilon_{ic} \tag{4}$$

Hypothesis 5: Mood and Aggressive Behavior To estimate the effect of the purifier on the mood of children and aggressive behavior, Y_{ict}, we use a model similar to our empirical design to test for the effects on the Raven test (Eq. 3).

$$Y_{ict} = \beta_1 Post_t + \beta_2 AirPurifier_c * Post_t + \lambda_s + \lambda_t + \varepsilon_{ict} \tag{5}$$

For secondary outcomes, we estimate the effect of purifiers on subjective health and environmental concerns with models similar to Equations 3 and 5. For the effect of treatment on other pollutants (CO, VOCs), we use the same model as for $PM_{2.5}$. For non-standard language and mathematics grades, for which four semesters are available (two pre-treatment and two post-treatment), we follow the same model as for school absences. For differential defensive behavior in response to air purifiers, calculated through the estimate of the number of times windows are opened during the day, we use a model similar to Equation 2.

3.2. Power calculations

To estimate the statistical power of our empirical design on air pollution (Eq.2), we use a Monte Carlo simulation procedure described in Appendix A.1. We estimate a minimal detectable effect (*minimal detectable effect* (MDE)) of 8% of the average pretreatment values with 5% significance and 80% power. We expect the effect of our interventions to be considerably greater than this 8% since the purifiers producer reports an expected reduction of between 90% and 99% in $PM_{2.5}$ depending on local conditions. Furthermore, in a similar framework, Gignac et al. (2021) estimates a reduction in $PM_{2.5}$ for Barcelona schools after installing HEPA filters of 89%. These imply that our intervention has sufficient power to identify the impact of purifiers on indoor air pollution.

For absences, we use pilot data for the school year 2022-23 provided by one of the schools and estimate the MDE for Eq.7 using Monte Carlo simulations (see Appendix A.2 for details). Our design allows us to identify a decrease of 4.5% in absences with 80% power and 5% significance. Although we are unaware of any studies that directly test the relationship between air purifiers and absenteeism, previous research has found a clear association between air pollution and school absences. For example, in a quasi-experimental design, Hales et al. (2016) finds that an increase in $PM_{2.5}$ of 10 *micrograms per cubic meter* (μg/m^3) increases absences by 1.7%.[9] In our context, given the average level of $PM_{2.5}$ in the city throughout the school year (26.98 μg/m^3) and the expected effect of purifiers on $PM_{2.5}$ (between 90 and 99%), we expect a reduction between 20 and 25 μg/m^3. Assuming a linear relationship between improved air quality and absences, this would imply a reduction in absences between 3.4% and 4.25%. Although these effects are slightly lower than our MDE, we expect that several aspects will give us larger estimates. First, higher-level filtering provides better air quality not only by curbing $PM_{2.5}$ but also allergens, pathogens, and other contaminants. Second, air pollution levels throughout the year reach extremely high peaks in winter. For example, the average $PM_{2.5}$ value in November 2022 was 46.2 μg/m^3. Third, while most studies on the effect of air pollution on absences focus on the short-term relationship between both variables, our treatment effect would contain the short and cumulative benefits of better air quality.

We estimate the MDE of the Raven Colored Progressive Matrices Tests also with Monte Carlo simulations (see Section A.3). We obtain a MDE of 0.13 standard deviations (σ), with 80% power and 5% confidence level. For the INVALSI test scores, our design grants us an MDE of 0.26σ.[10] To put things in perspective, previous research found that a 1 μg/m^3 reduction in $PM_{2.5}$ increases test scores by approximately 0.02σ (Gilraine and Zheng, 2022) and that the installation of air purifiers in schools in a neighborhood of Los Angeles improved test scores by

[9] Ransom and Pope (1992) find that an increase in monthly PM10 of 100μg/m^3 was associated with an increase in absenteeism 40%.

[10] For the Raven's Colored Progressive Matrices test, we draw information from Facon et al. (2011). For INVALSI, we use the average, standard deviation, and intra-cluster correlation for Northern Italy in 2022, from the Italian Ministry of Education.

0.03σ per μg/m^3 of outdoor $PM_{2.5}$ (Gilraine, 2023). For our expected reduction in $PM_{2.5}$ (20 to 25 μg/m^3), we would expect an effect of 0.4σ to 0.45σ. This suggests that the statistical power to estimate the effects on standardized tests is sufficient.

To the best of our knowledge, this is the first study to experimentally test the effect of air purifiers on children's moods and aggressive behavior. We arranged not to run a power test because we did not have enough information to elicit the prior distribution of children's mood or aggressive behavior. Although some studies link air pollution with mood and aggression, they focus on adults, use different causal variables, and measure mood and aggression in heterogeneous ways. Consequently, linking our study with previous research to estimate its MDE required more assumptions than we were comfortable making.

3.3. Robustness checks

We perform a battery of robustness checks. First, we control for time-varying local conditions, namely temperature and humidity, to reduce concerns about unobservables driving our coefficients when looking at the effects of the purifiers on absences and air pollution. Temperature and humidity can influence the quality of pollution measurement as consumer-grade $PM_{2.5}$ monitors tend to overestimate concentrations with high humidity. Specifically, for the sub-sample of classes left with monitors, we replace day fixed effects with the following controls: temperature, humidity, and class characteristics such as floor and orientation. Similarly, we exploit the variation between classes on the day when the survey is administered and add the same set of controls to the models using survey-based outcomes.

Second, to deal with missing values in the survey questions, we repeat the models using survey-based outcomes by imputing the average class-level mean value and including dummies for missing observations.

Third, we control for multiple hypothesis testing for all main outcomes and hypotheses presented (List et al., 2019). In particular, we calculate and report the FDR adjusted q values (Benjamini et al., 2006).

3.4. Heterogenous effects

We examine the heterogeneity of our results for different sub-samples to test for the extent to which the intervention contributes to reducing inequalities and varies over time and seasonality. We do this by focusing on the different effects for students with higher levels of absences, lower baseline cognitive skills, and socioeconomic backgrounds. For example, previous research has found that the effects of air pollution on absenteeism are particularly stronger for more disadvantaged students (Liu and Salvo, 2018).

Our first hypothesis is that improved air quality due to purifiers reduces absenteeism to a greater extent for students who were absent more in the previous academic year. We test for this by adding the interaction effects of $Post_t$ and $AirPurifier_c * Post_t$ in Eq. 3 with an indicator variable equal to one if the student was absent more than the median number of days in the pretreatment school year and zero otherwise.

Next, given that air pollution is significantly higher during winter in northern Italy, we hypothesize that the effects of purifiers on air pollution and absenteeism are greater compared to Spring and Fall. Similarly, we also check whether the effects are greater in weeks of poor air quality. We test for these heterogeneous effects by adding two interaction variables, one taking the value of one for dates between December 21 and March 20 and zero otherwise, and the other taking the value of one when the average level of outdoor $PM_{2.5}$ in the previous seven days is above the median.

We also expect to see heterogeneous effects on the impact of purifiers on cognitive skills. For this, we hypothesize that the impact of air purifiers in improving cognitive tests is greater for students with lower baseline levels of cognitive skills and socioeconomic status. We test these hypotheses by adding the interaction effects of $Post_t$ and $AirPurifier_c * Post_t$ in Eq. 3 with a dummy variable equal to one if the student scored below the median of the Raven test at baseline and zero otherwise, and a dummy variable equal to one if the socioeconomic index of the student's family is low and zero otherwise.[11]

As with the heterogeneous effects of the Raven test, we also examine whether the impact of air purifiers on INVALSI test scores is greater for students with lower baseline levels of the Raven test scores and socioeconomic status. We apply multiple hypothesis corrections to the test of heterogeneous treatment effects by calculating FDR-adjusted q-values for the interaction coefficients (Benjamini et al., 2006).

[11]INVALSI provides this socio-economic index, which aggregates information concerning parents' occupational status, education, and ownership of goods favoring learning.

3.5. Challanges and limitations

In addition to the mechanical effects of air purifiers acting through improved air quality, there is also the possibility that their presence would affect the behavior of teachers and students, which can eventually affect the outcomes of interest. For example, teachers (or students) in treated or control classes can change the frequency of opening windows or make different decisions about staying or not in class in response to the presence of purifiers.[12] In that scenario, instead of identifying the effects of purifiers through improved air quality, point estimates would include the effects of behavioral adaptations.

There are several strategies to track the degree of behavioral adaptation due to purifiers. First, while purifiers remove $PM_{2.5}$ and other contaminants, they do not remove CO_2. CO_2 is typically found in higher concentrations in indoor environments due to expiration. The opening of windows can reduce the levels of CO_2 while the effect on $PM_{2.5}$ can go in both directions, depending on the outdoor levels. Therefore, we can use the high-frequency measurement of CO_2 and attribute sudden drops in its levels to the opening of windows.We construct a proxy for the daily number of times windows are opened and use a model similar to 2 to estimate the adaptive response to purifiers. Second, we use survey questions on environmental awareness. In particular, we look at the differences in responses to the two questions dedicated to students' assessment of the importance of air pollution as a general issue and the subjective quality of air in a class by treatment. We do this with a model similar to 3 and 5. Significant differences in students' perception and awareness between treatment and control classes could be a signal that teachers reacted differently to the presence or absence of purifiers.[13] In the presence of significant differences in adaptive behavior, the results of the main hypothesis should be considered as a lower bound of the impact of purifiers.

References

Abadie A., Athey S., Imbens G. W., and Wooldridge J. M. (2023) When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1): 1–35. 11

Anderson M. L. (2008) Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American statistical Association*, 103(484): 1481–1495. 4

[12]For instance, teachers can decide to spend the long break after lunch in class or in the garden/courtyard.

[13]We re-iterate that the intervention does not involve any type of information or awareness campaign for students or teachers.

Benjamini Y., Krieger A. M., and Yekutieli D. (2006) Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3): 491–507. 14, 15

Bruhn M. and McKenzie D. (2009) In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, 1(4): 200–232, Publisher: American Economic Association. 11

Cori L., Donzelli G., Gorini F., Bianchi F., and Curzio O. (2020) Risk Perception of Air Pollution: A Systematic Review Focused on Particulate Matter Exposure. *International Journal of Environmental Research and Public Health*, 17(17): 6424. 9

Dean E. B., Schilbach F., and Schofield H. (2017) Poverty and Cognitive Function. in *The Economics of Asset Accumulation and Poverty Traps*: University of Chicago Press. 8

Facon B., Magis D., Nuchadee M.-L., and De Boeck P. (2011) Do Raven's Colored Progressive Matrices function in the same way in typical and clinical populations? Insights from the intellectual disability field. *Intelligence*, 39(5): 281–291. 13, 21

Gignac F., Barrera-Gómez J., Persavento C., Solé C., Tena È., López-Vicente M., Foraster M., Amato F., Alastuey A., Querol X. et al. (2021) Short-term effect of air pollution on attention function in adolescents (ATENC! Ó): A randomized controlled trial in high schools in Barcelona, Spain. *Environment international*, 156: 106614. 13

Gilraine M. (2023) Air Filters, Pollution, and Student Achievement. *Journal of Human Resources*. 14

Gilraine M. and Zheng A. (2022) Air Pollution and Student Performance in the U.S.. Working Paper 30061. 13

Hales N. M., Barton C. C., Ransom M. R., Allen R. T., and Pope III C. A. (2016) A quasi-experimental analysis of elementary school absences and fine particulate air pollution. *Medicine*, 95(9). 13

List J. A., Shaikh A. M., and Xu Y. (2019) Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22(4): 773–793. 14

Liu H. and Salvo A. (2018) Severe air pollution and child absences when schools and parents respond. *Journal of Environmental Economics and Management*, 92: 300–330. 15

Mani A., Mullainathan S., Shafir E., and Zhao J. (2013) Poverty impedes cognitive function. *science*, 341(6149): 976–980. 8

Ransom M. R. and Pope C. A. (1992) Elementary school absences and PM10 pollution in Utah Valley. *Environmental Research*, 58(1): 204–219. 13

Raven J. C. and Court J. H. (1998) *Raven's progressive matrices and vocabulary scales*: Oxford pyschologists Press. 8

Rubalcava L. and Teruel G. (2006) Mexican Family Life Survey, First Wave. 8

Sacchi D., Pasini A., Muti M., Roverselli C., and Spiridigliozzi S. (2023) The Effects of Lockdown 2020 on the Behaviour and Mood of Children Aged Three to Six in Italy. *Journal of Educational, Cultural and Psychological Studies (ECPS Journal)*(27). 9

Schmit K. M., Brown R., Hayer S., Checovich M. M., Gern J. E., Wald E. R., and Barrett

B. (2021) Wisconsin Upper Respiratory Symptom Survey for Kids: Validation of an Illness-specific Quality of Life Instrument. *Pediatric Research*, 90(6): 1207–1214. 8

Appendix

A. Power Calculations

A.1. Fine Particulate Matter

We use a simulation procedure to estimate the MDE of purifiers on air pollution. First, we use the network of monitoring stations in the city of Milan to assign pollution values to schools with inverse distance weighting.[14] After assigning pollution to each school, we select between six and eight classes per school to install air quality monitors and randomly assign 50% to treatment (air purifier) and 50% to control (no air purifier) according to $AirPurifier_c \sim Bernoulli(P = 0.5)$. If a classroom is treated, we multiply the treatment indicator ($AirPurifier_c$) by a parameter $(1 - \beta)$ where $\beta \in (0, 1)$. The size of β represents the simulated treatment effect of the purifiers. For example, $\beta = 0.1$ implies a decrease of $0.1 \times P\hat{M}_{2.5}$. Once we assign treatment effects, we estimate the MDE (80% power and 5% significance.) of purifiers with 500 different simulations and 20 different values of β.

In each simulation, we randomly draw, for each classroom, 171 air pollution observations from the pool of school days (with replacement). After randomly drawing air pollution values, we estimate the treatment effect (for the 20 values of β) with Equation 6. In it, $PM2.5_{ct}$ is the value of $PM_{2.5}$ for classroom c at time t. $AirPurifier_c$ equals one if the classroom is treated and zero if control. λ_t are fixed effects for the observation date that accounts for common shocks across classrooms. ϵ_{ct} is the idiosyncratic error term. We cluster standard errors at the classroom level.

$$PM_{ct} = \alpha + (1 - \beta)AirPurifier_c + \lambda_t + \epsilon_{ct}. \quad (6)$$

Figure A.1 presents the results of the simulation. The Y-axis shows the share of the 500 iterations in which we identify a significant effect ($\alpha = 5\%$), i.e., the level of statistical power. The x-axis presents the effect size (β). Our simulation exercise suggests that our experimental design allows us to identify effects as small as 6.5% of the mean pretreatment values with 80% power and 5% significance.

[14] Here, we assume that indoor pollution levels are the same as outdoor pollution levels.

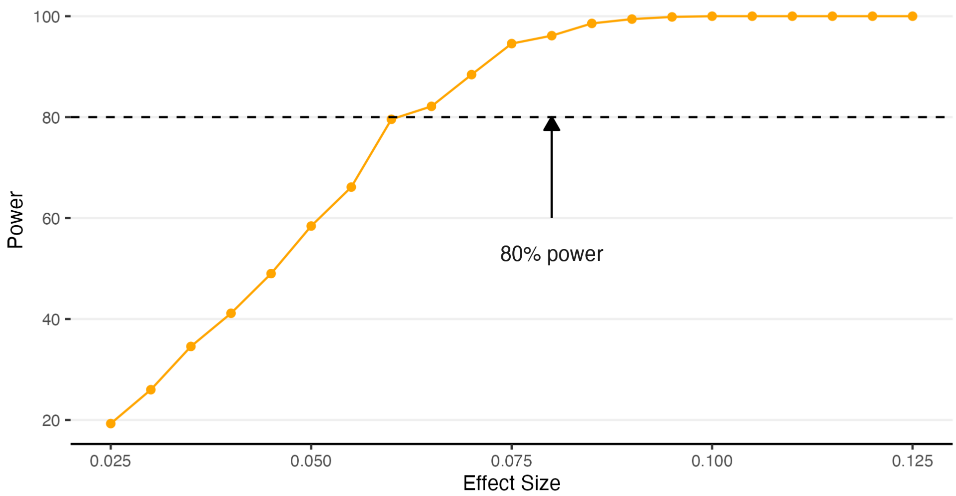


Figure A.1: Power calculations for $PM_{2.5}$

A.2. School Absences

Using pilot data on students' absences provided by one school, we create synthetic data sets for the other schools with a classification and regression tree of absences as a function of the classroom, day of the week, month, and year of observation.[15] This gives us a final data set of 356,022 observations between 09/13/2021 and 06/08/2023. Each observation contains the date, school, class, student id, and an indicator variable if student i missed school on date t.

We assume that the absences of each student follow a binomial distribution of the form $Absence_i \sim Binomial(n, p)$ for all $= p = \frac{\sum_t^T Abs=1}{\sum_t^T Abs=1|0}$. After estimating the binomial distribution of each student, we randomly assign 50% of the classes to the treated group and 50% to the control group with a Bernoulli process of the form: $Prob(Student_{ic} = Treated) = Bernoulli(P_c = 0.5)$. If the student's classroom is assigned to the treated group, we multiply the p value of its binomial distribution by a parameter $\beta \in (0, 1)$ that effectively simulates the effect of purifiers on absences. For example, to simulate a reduction of 5% in absences, we multiply the absenteeism rate (e.g., 10%) by $\beta = 0.95$. After this, we estimate our MDE by running 500 simulations for 20 different values of β and estimating the share of significant point estimates for each β across the 500 estimations.

To estimate β, we use a simple linear probability model following Equation 7. In it, $Absent_{ict}$ is an indicator variable equal to 1 if student i in class c missed school at time t. $AirPurifier_c$

[15] A Classification and Regression Tree (CART) is a versatile decision tree learning technique that can be employed for both classification and regression predictive modeling problems. CART builds a binary tree where each node represents a decision based on the values of the input variables. In the case of classification, the outcome is categorical (e.g., absent or not absent), leading to a discrete label for each terminal node or "leaf" of the tree. The tree can then be used to predict new data by traversing the tree based on the input variable values and arriving at a leaf that provides the prediction.

is a dummy equal to one if the classroom c was assigned to treatment and 0 otherwise. $Post_t$ takes the value of one in the post-treatment period, and zero otherwise; λ_t are date-fixed effects that account for shocks common to all classrooms in t. λ_i are student-fixed effects, considering all student-specific attributes that lead to heterogeneous absence profiles. We cluster standard errors at the classroom level (treatment assignment).

$$Absent_{ict} = \beta_1 Post_t + \beta_2 AirPurifier_c * Post_t + \lambda_t + \lambda_i + \epsilon_{it}. \tag{7}$$

Figure A.2 presents the power calculations for school absences. the Y-axis shows the share of the 500 iterations where we identify a 5% significant effect. The x-axis presents the size of the effect as a percentage decrease in the probability of absences. Each panel presents the results of a different estimator, Logit and OLS. Both estimators include fixed effects for students and dates. Our simulation exercise suggests that our experimental design allows us to identify effects as small as a 4.5% decrease in the probability of absences with 80% power and 5% significance.

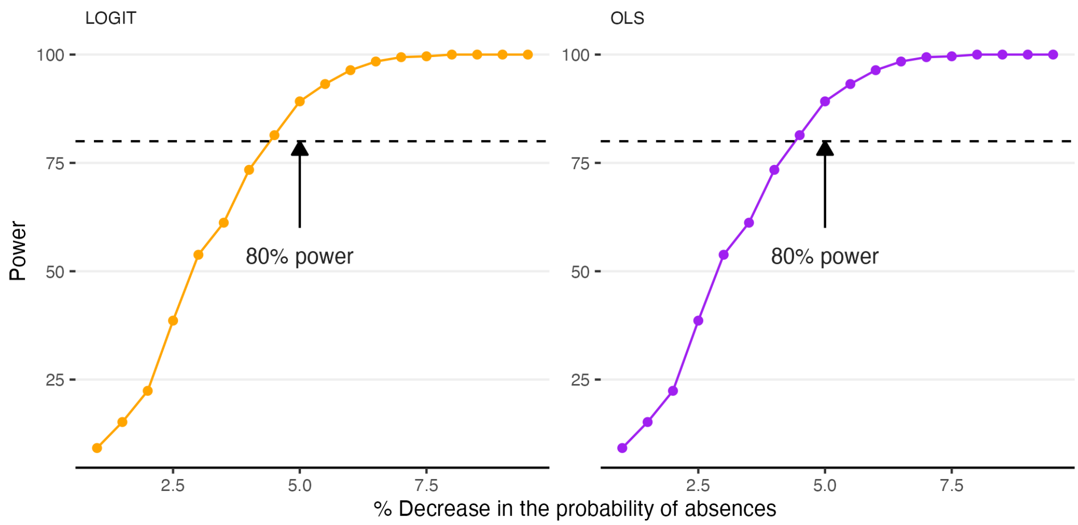


Figure A.2: Power calculations for school absences. Left panel: Logit. Right panel: OLS.

A.3. Raven test

We estimate the MDE of the Raven's Colored Progressive Matrices test through Monte Carlo simulation. In it, we simulate the impact of the purifiers on the test with the following procedure. First, we randomly assign students to the control or treatment group according to $Prob(Student_{ic} = Treated) = Bernoulli(P_c = 0.5)$. Next, for each student, we randomly assign pretreatment test scores based on the mean and standard deviation reported by Facon et al. (2011) assuming a normal distribution ($Raven_i^{pre} \sim \mathcal{N}(\mu, \sigma)$), that is, a mean of 22.4

and a standard deviation of 5.3. Next, we assign the post-treatment test values according to $Raven_i^{post} \sim \mathcal{N}(\mu \times (1+\beta), \sigma)$, in which β is always equal to zero for the control group and positive for the treatment group. After simulating all the data, we estimate the ATT with a simple OLS regression of the form:

$$Raven_{ict} = \beta_1 Post_t + \beta_2 AirPurifier_c * Post_t + \lambda_t + \lambda_s + \epsilon_{ict} \tag{8}$$

We run this simulation 500 times and calculate the share of type 2 errors with a significance threshold of 5%. Figure A.3 presents the relationship between power and different effect sizes in terms of standard deviations across 500 iterations. Our estimated MDE are 0.13 σ, with 80% power and 5% confidence level.

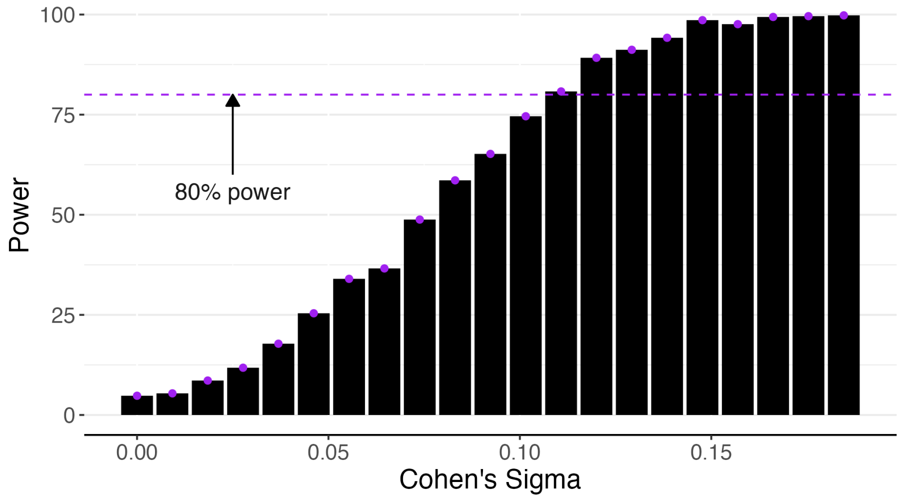


Figure A.3: Power calculations for the Raven's Colored Progressive Matrices test

A.4. INVALSI test scores

To estimate the MDE of the INVALSI test scores, we use data from the Italian Education Ministry on the mean and standard deviation of INVALSI scores. As we estimate the effect of the intervention on INVALSI with a simple difference in means between treated and control students, we estimate the MDE according to:

$$MDE = (\tau_{1-\chi} + \tau_{\alpha/2})\sqrt{\frac{1}{p(1-p)} \times \frac{\sigma^2}{N} \times (1 + (m-1) \times ICC} \tag{9}$$

In it, $\tau_{1-\chi}$ and $\tau_{\alpha/2}$ are the τ-score for the desired level of significance (α) and power (χ). p is the share of units treated, which, in our case, is 50%. σ is the expected standard deviation of the tests. N is the number of treated individuals, m is the number of clusters (classrooms), and ICC is the intracluster correlation. We extract the average, standard deviation, and intracluster correlation of the INVALSI tests in northern Italy from the Italian Ministry of Education, which are equal to 60.4, 23, and 0.08. However, since we are uncertain about the ICC in the sample, we estimate the MDE for different ICC values. A.4 presents the power calculations from ICC between 0.01 and 0.1.

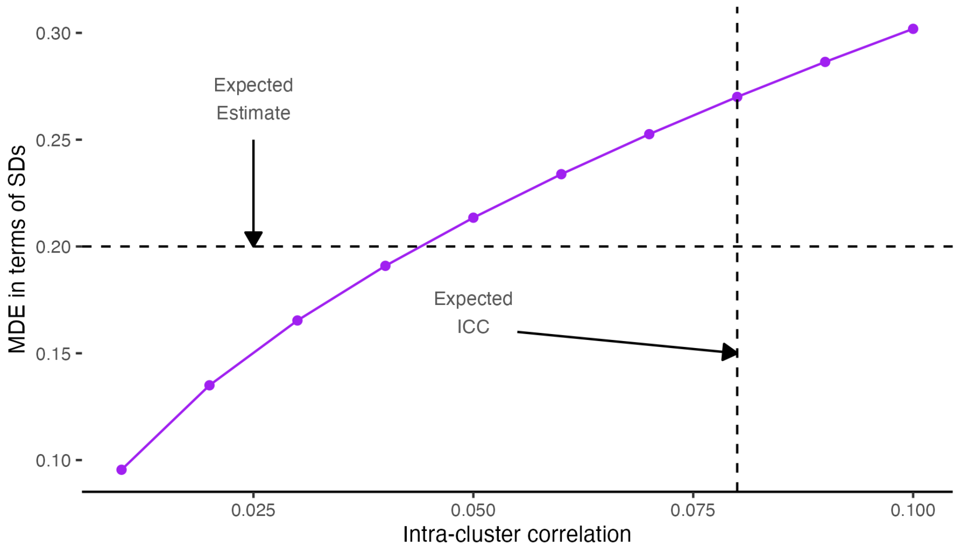


Figure A.4: Power calculations for INVALSI test

Our MDEs range from 0.1 to 0.3 standard deviations, with 80% power and 5% significance level. Given the ICC reported by schools in northern Italy, we expect an MDE of 0.26 σ.

B. Survey

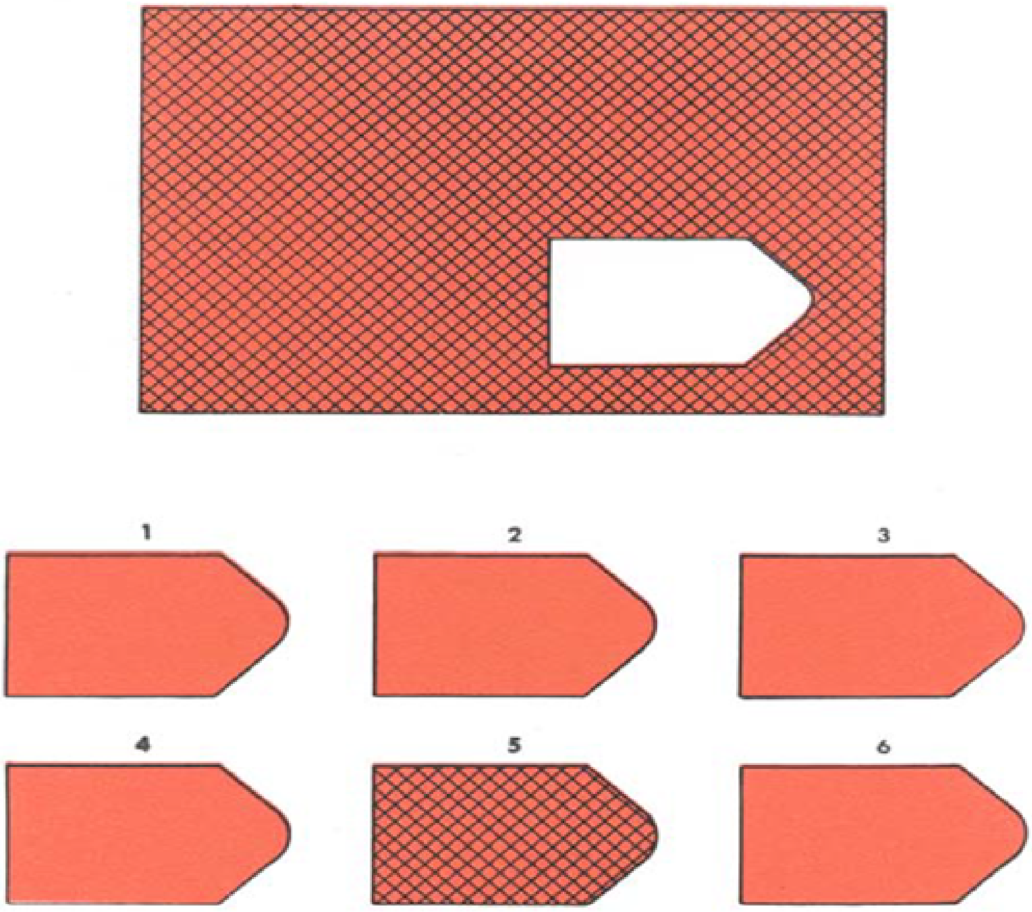
QUESTIONNAIRE

DATE:

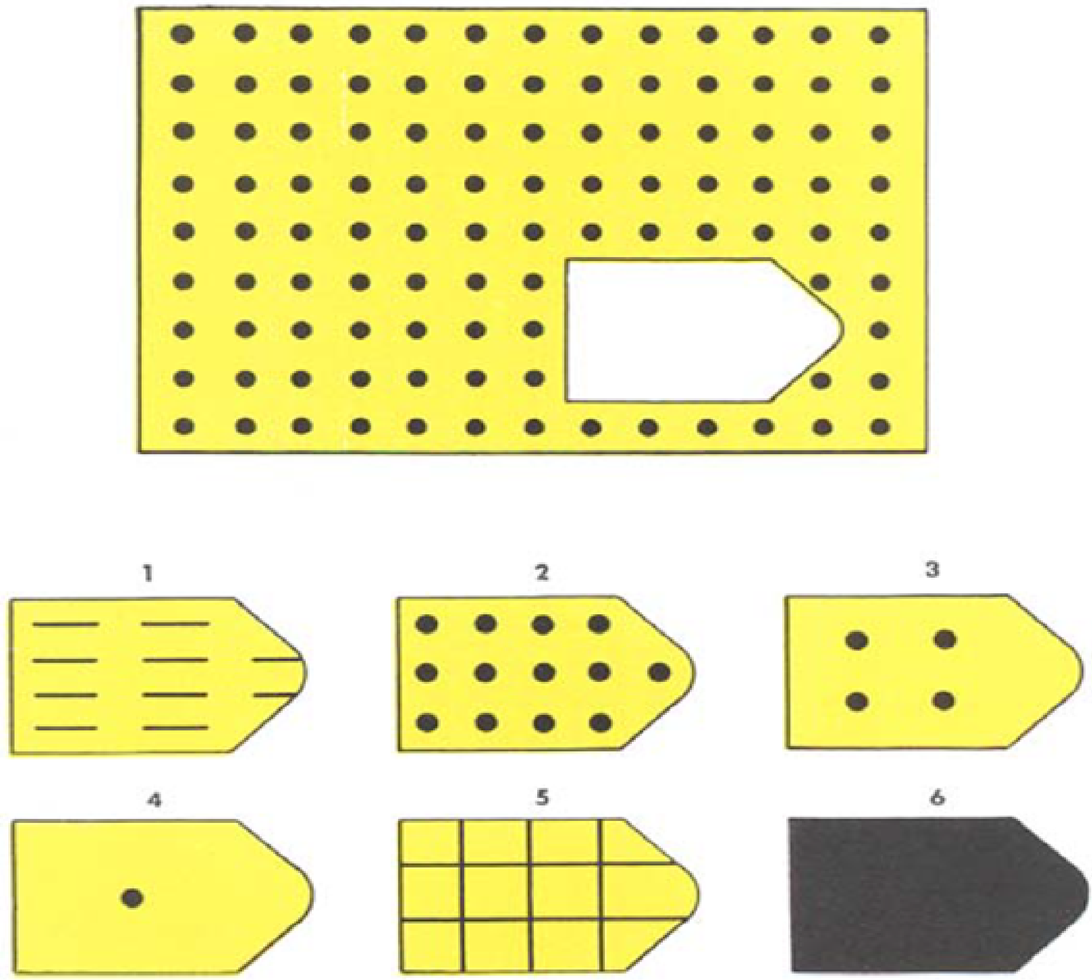
ID:

INSTRUCTIONS: FOR EACH COLORED CARD, SELECT THE MISSING PART AMONG THE GIVEN OPTIONS.

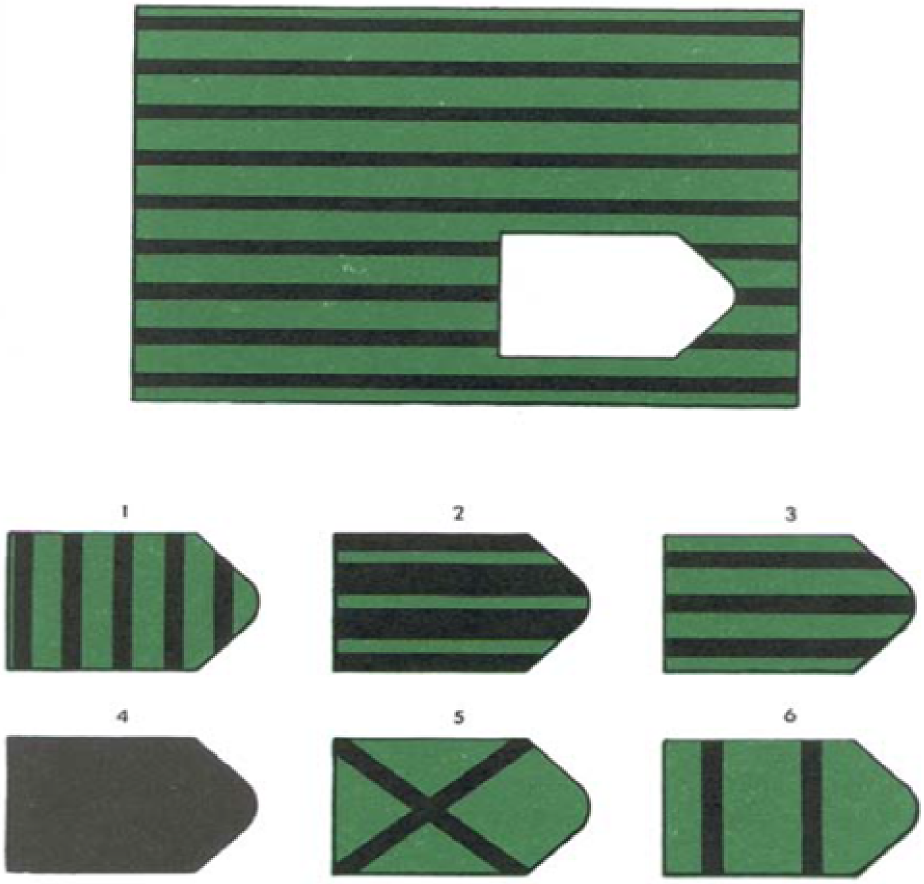
1



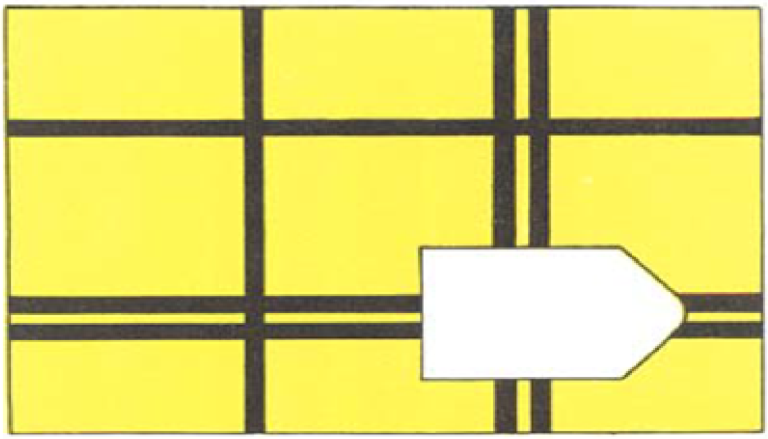
2

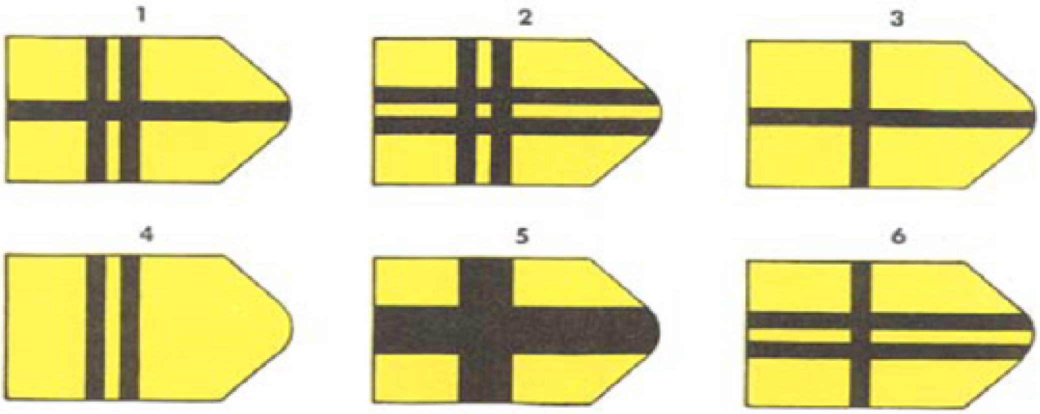


3

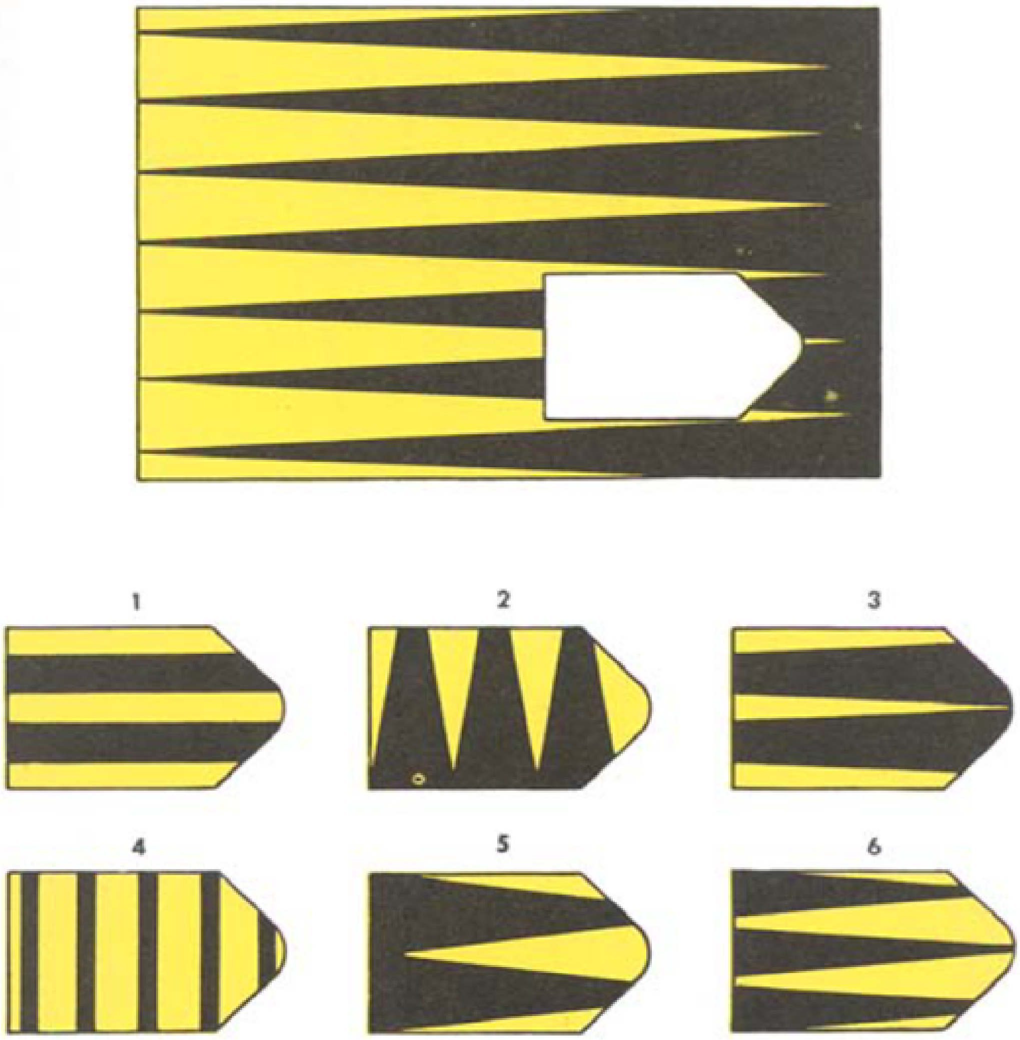


4

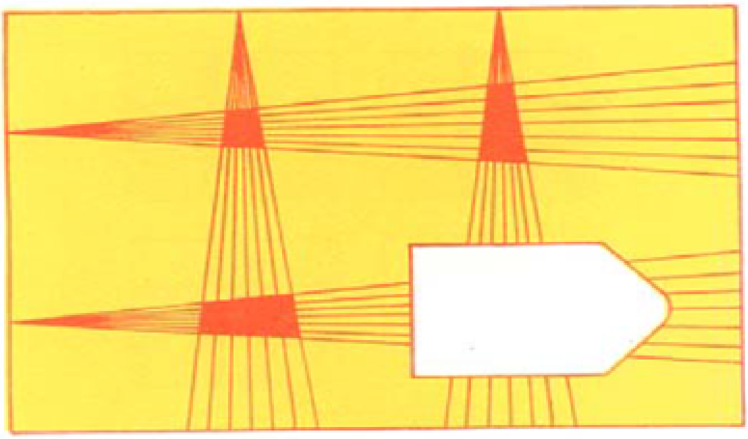


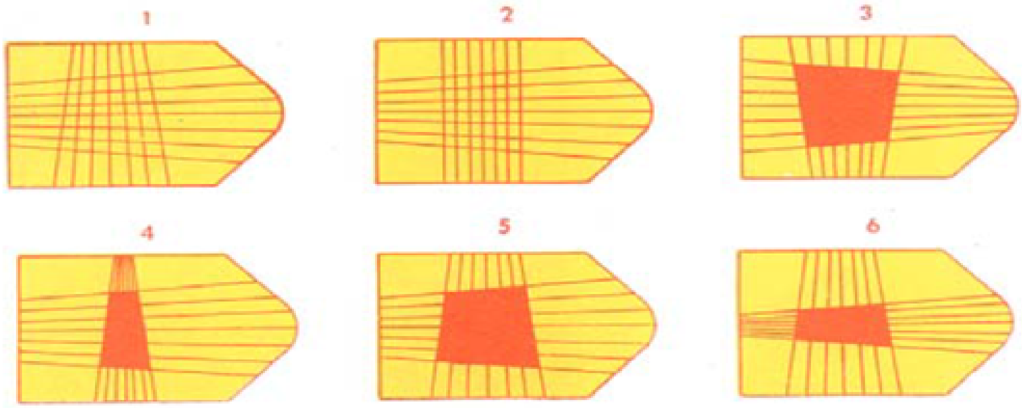


5

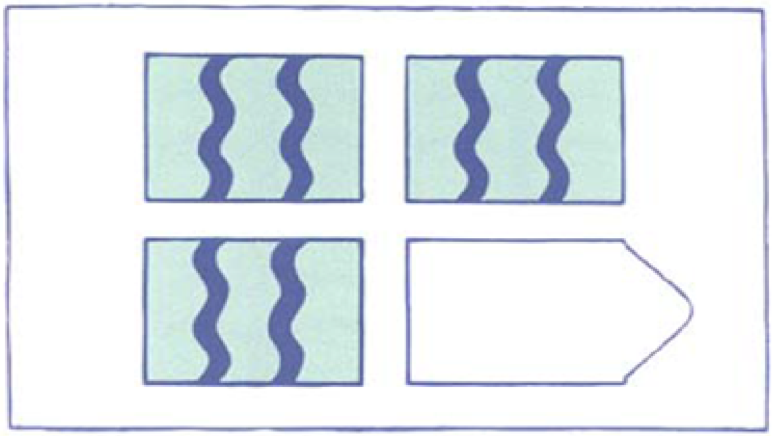


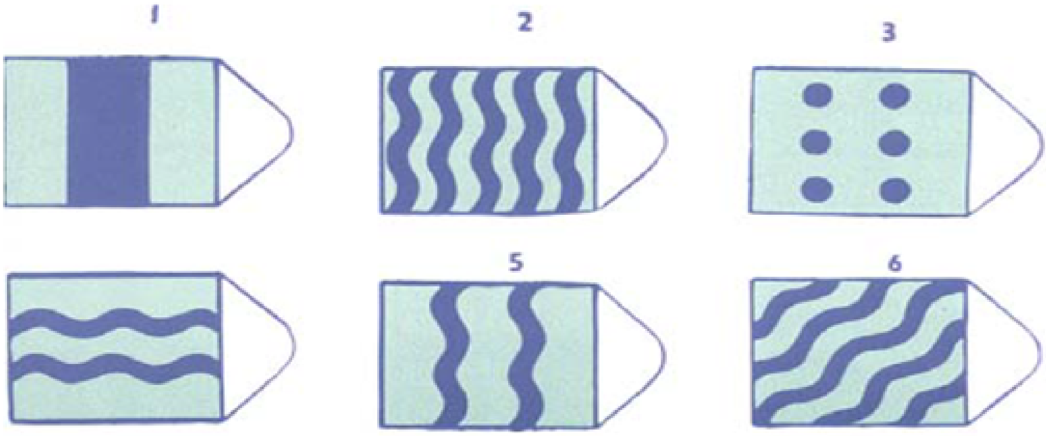
6



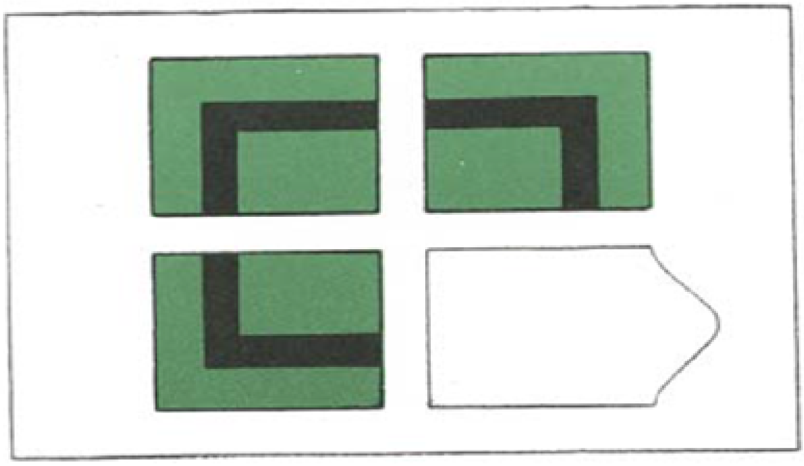


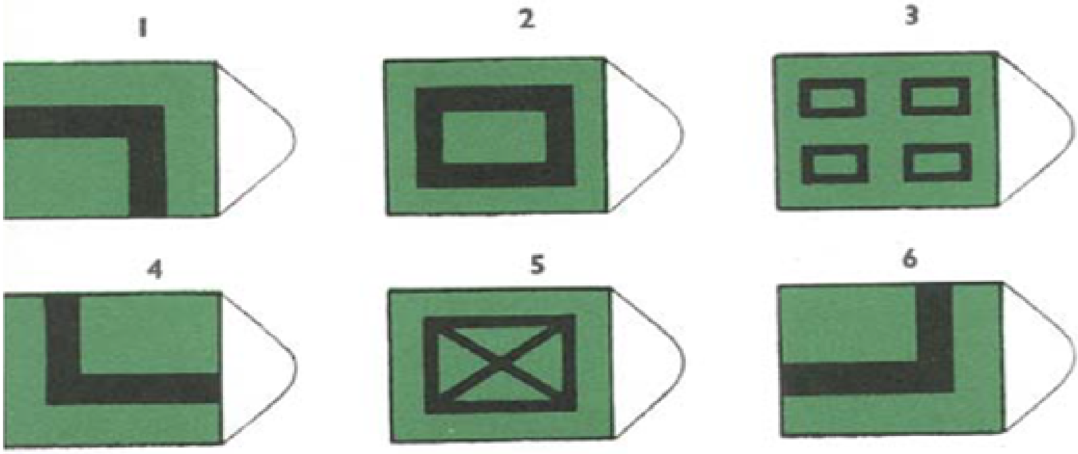
7



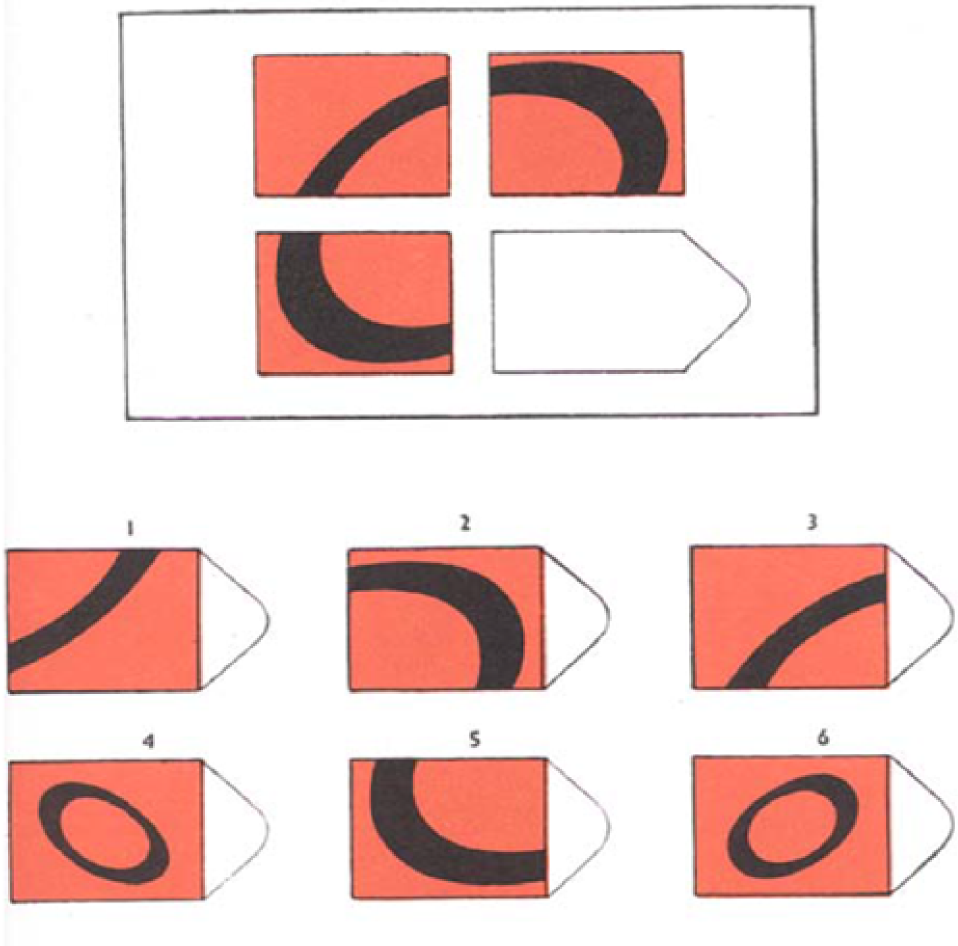


8

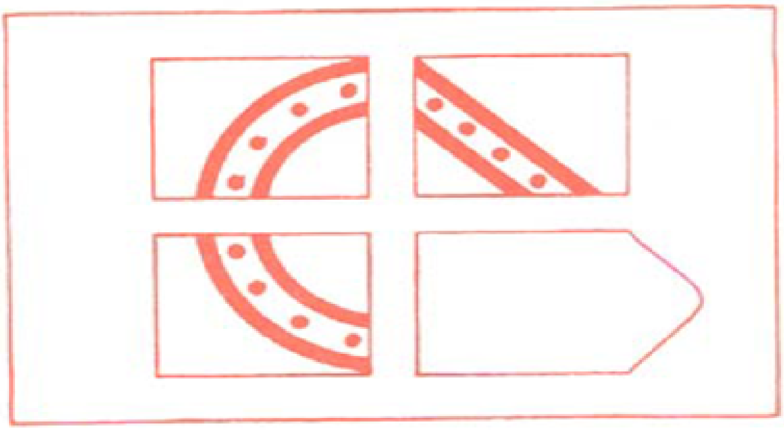


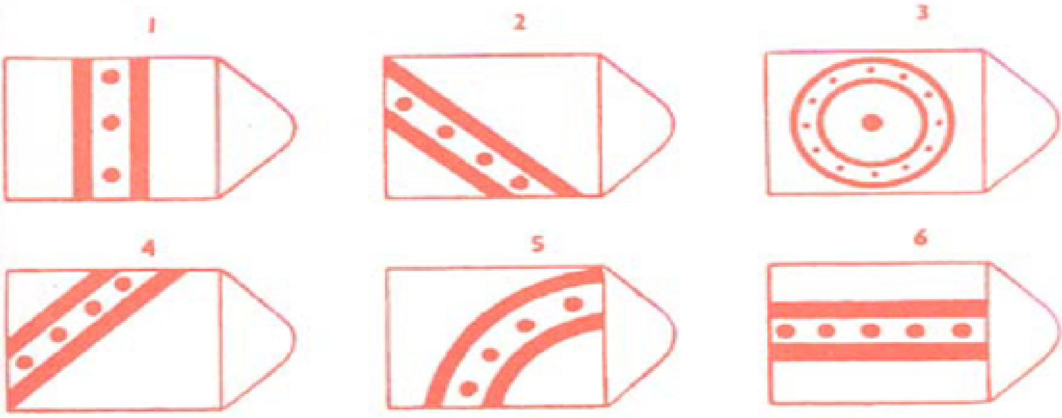


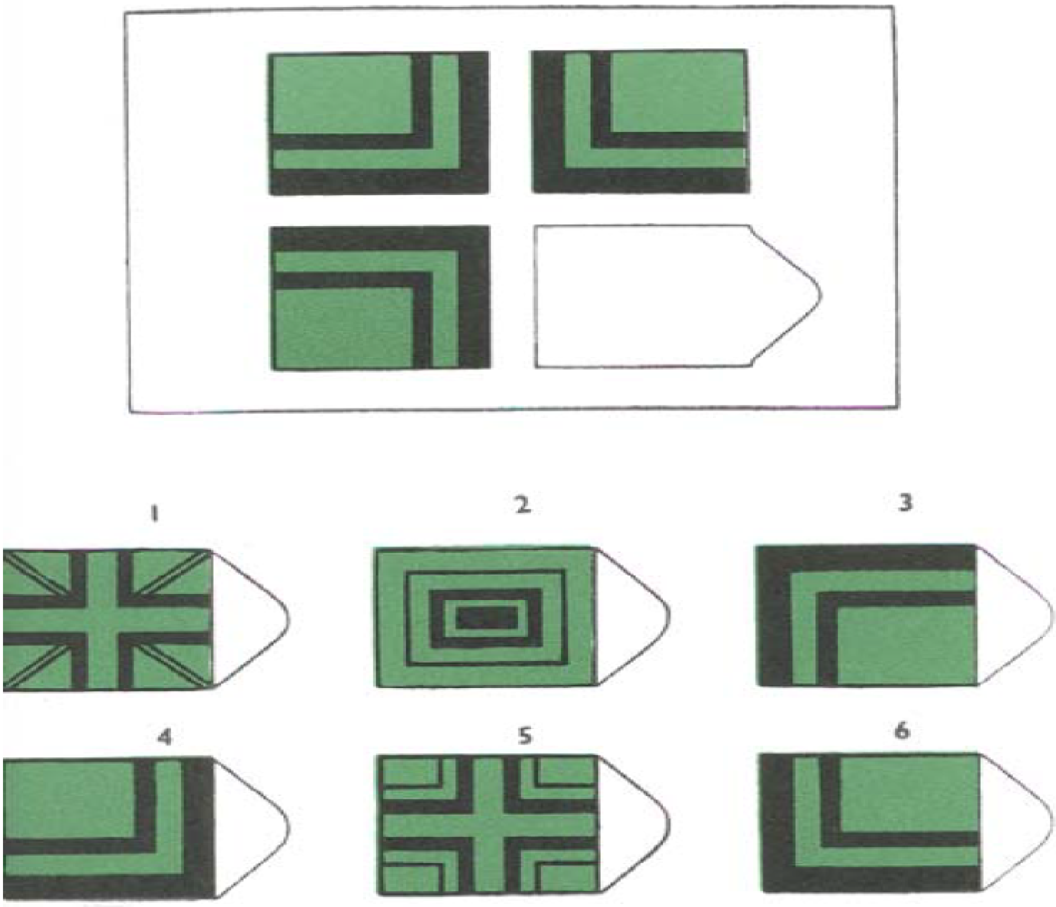
9



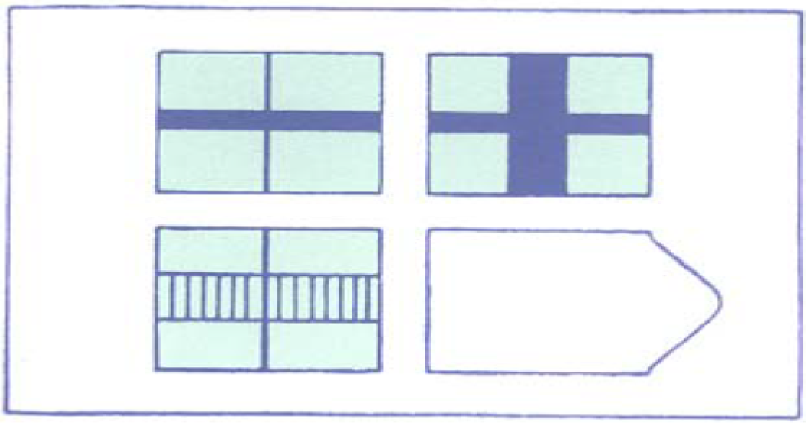
10



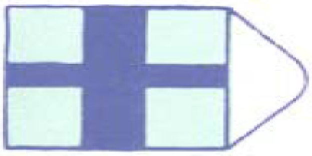



1
2
3
4
5
6

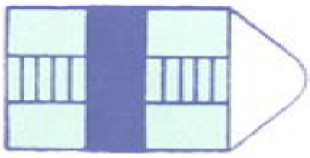
12



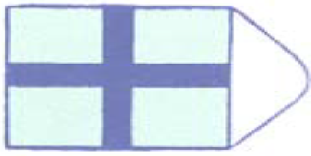
1



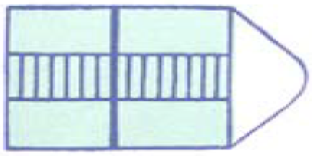
2



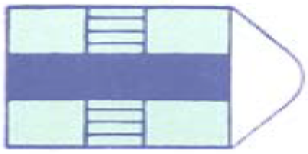
3



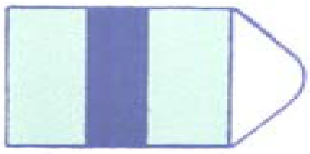
4



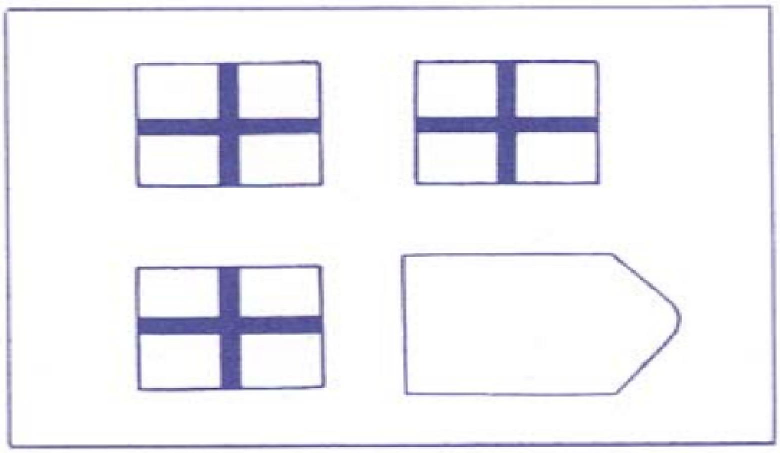
5

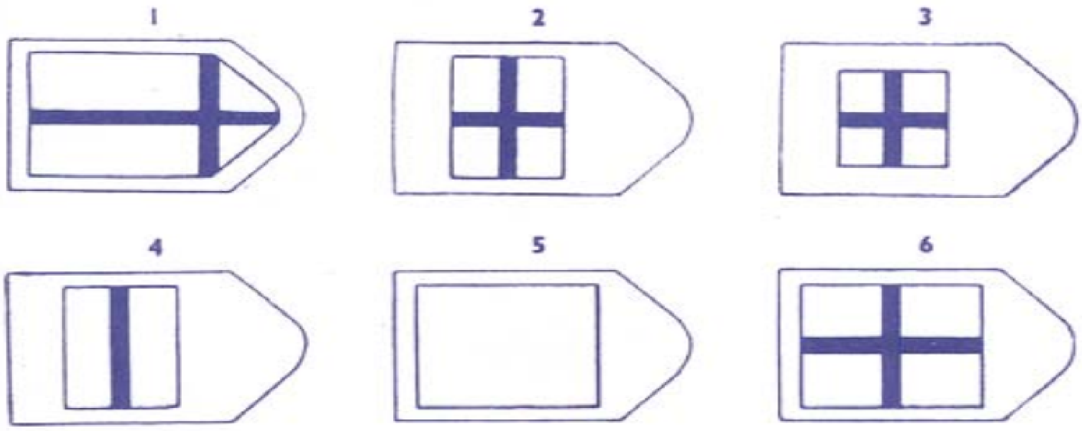


6

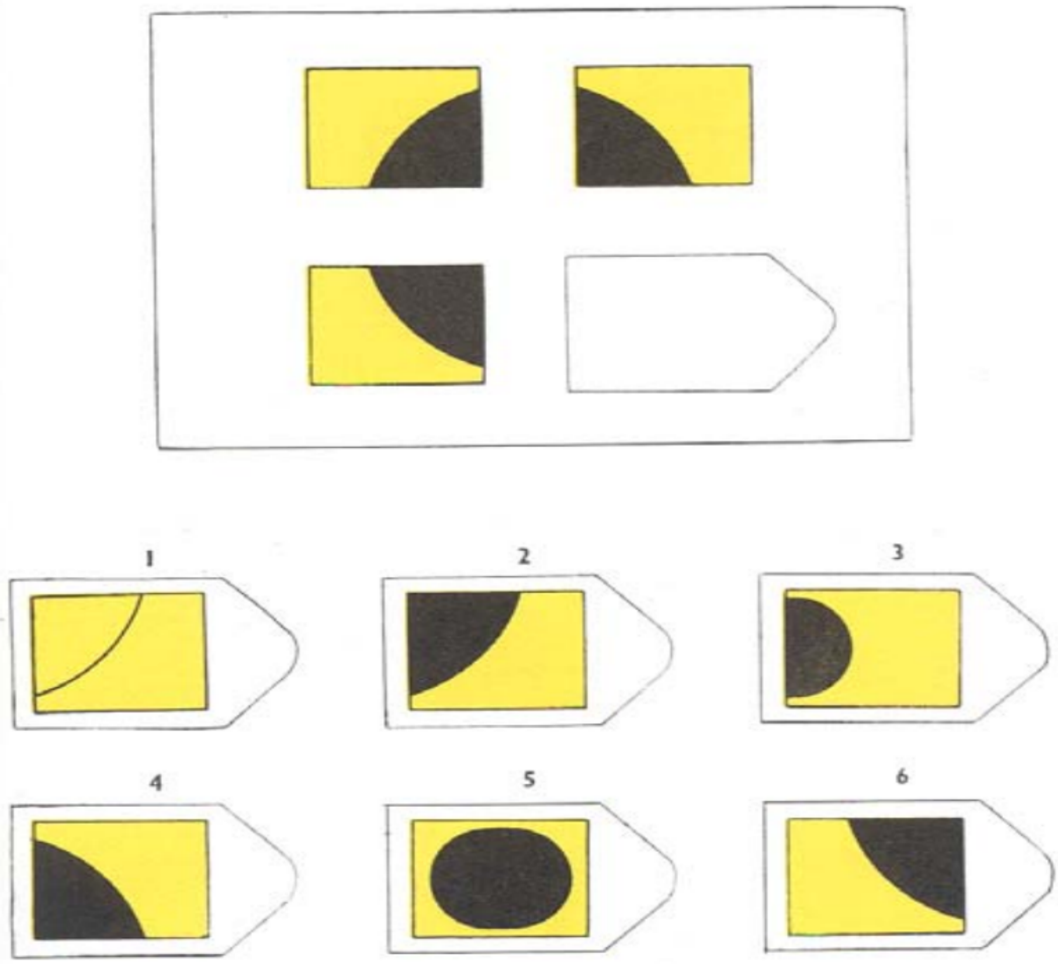


13

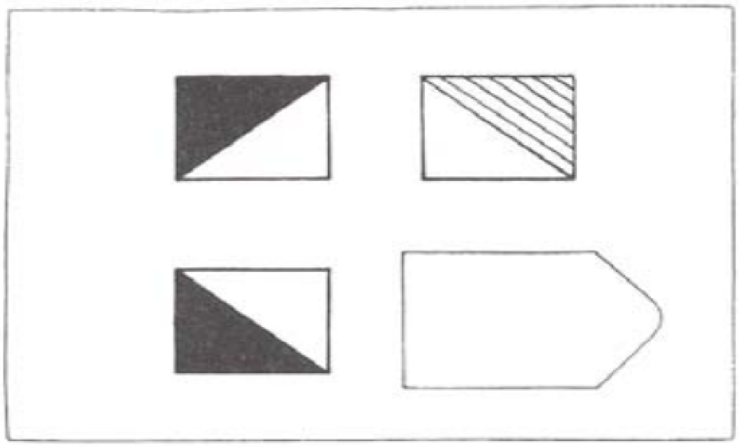


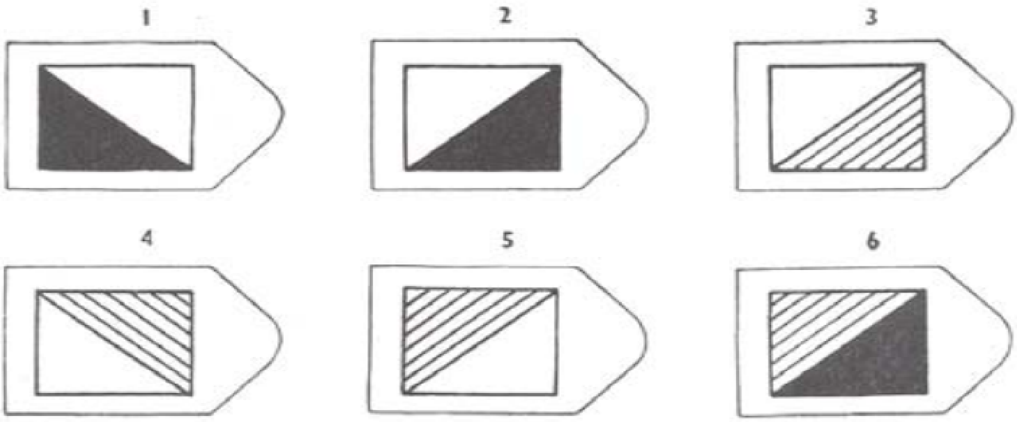


14

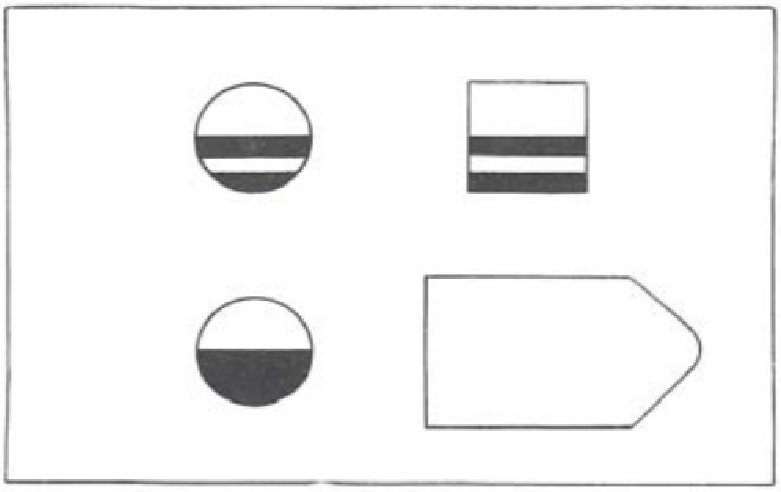


15

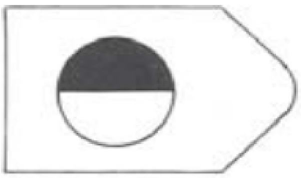




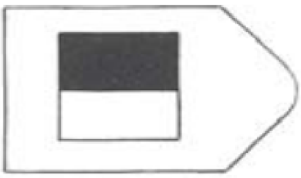
16



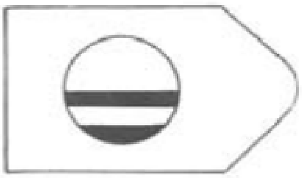
1



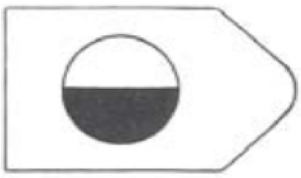
2



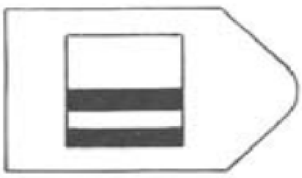
3



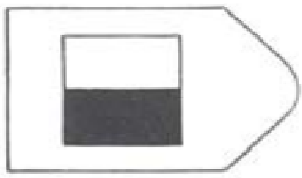
4



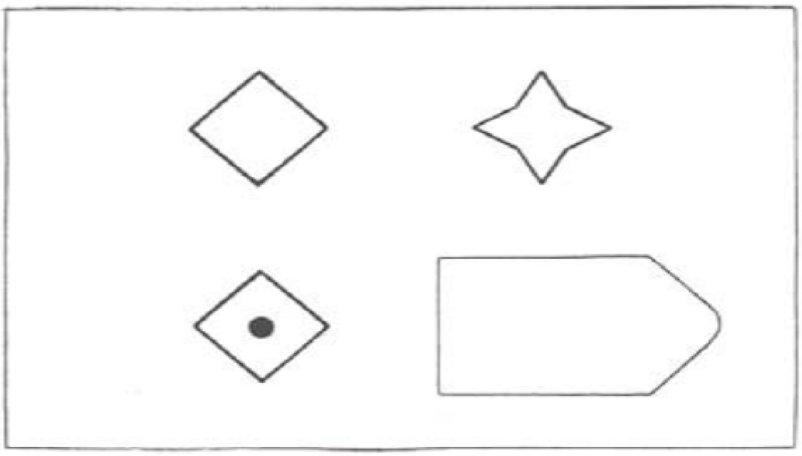
5

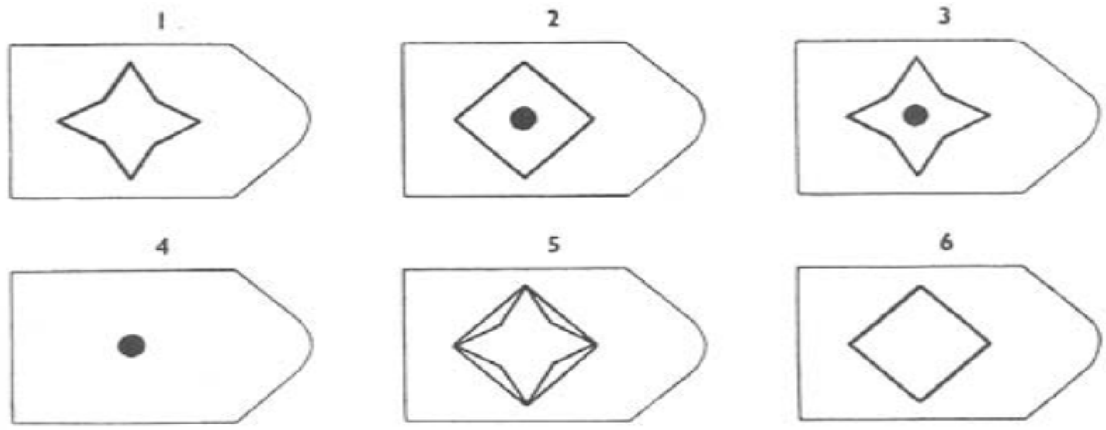


6

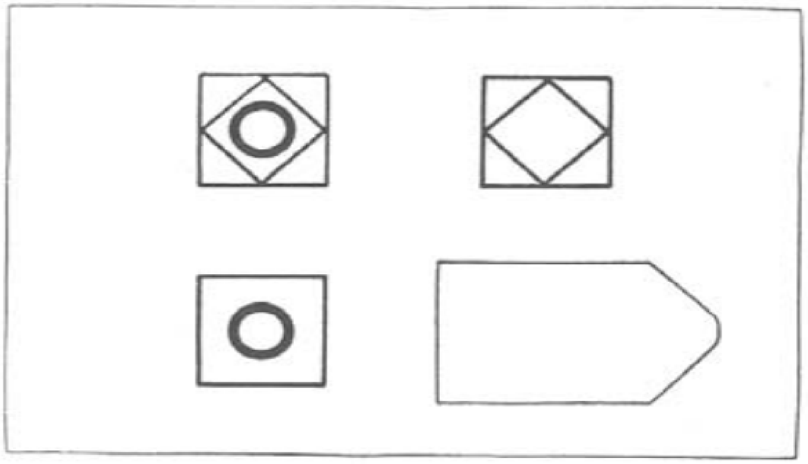


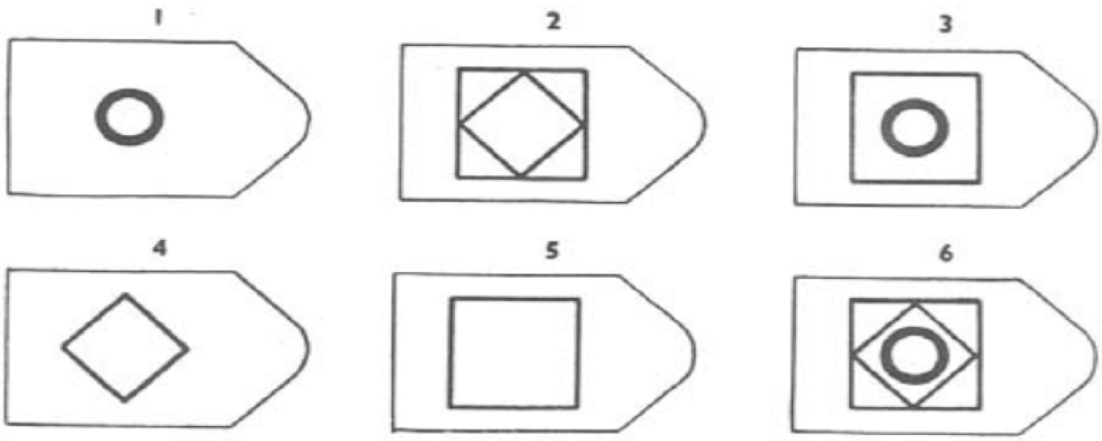
17





18





INSTRUCTIONS: FOR EACH QUESTION, PUT A CROSS IN THE BOX CORRESPONDING TO THE ANSWER YOU WANT TO GIVE.

1.	**NO**	**A BIT SICK**	**SICK**	**VERY SICK**
				
HAVE YOU BEEN SICK IN THE LAST WEEK?	▢	▢	▢	▢

2. **IF YOU HAVE BEEN SICK IN THE LAST WEEK, HOW OFTEN DID YOU HAVE THESE SYMPTOMS?**	**NO SYMPTOMS**	**SOMETIMES**	**OFTEN**	**EVERY DAY**
				
2.1 RUNNY NOSE	▢	▢	▢	▢
2.2 STUFFY NOSE	▢	▢	▢	▢
2.3 SNEEZING	▢	▢	▢	▢
2.4 COUGH	▢	▢	▢	▢
2.5 SHORTNESS OF BREATH	▢	▢	▢	▢
2.6 FEELING TIRED	▢	▢	▢	▢
2.7 HEADACHES	▢	▢	▢	▢
2.8 STOMACH ACHE	▢	▢	▢	▢

3.	NO	NOTHING SERIOUS	YES	YES, SERIOUS
				
DID YOU HAVE ANY ACCIDENTS IN THE LAST WEEK (FOR EXAMPLE, YOU HURT YOURSELF OR FALL)?	▢	▢	▢	▢

4. **IN THE LAST WEEK, WHAT WAS THE MOOD OF:**	**VERY POSITIVE**	**POSITIVE**	**NEGATIVE**	**VERY NEGATIVE**
	☺	😐	☹	😠
4.1 YOUR CLASSMATES	▢	▢	▢	▢
4.2 YOURSELF	▢	▢	▢	▢

5.	NO	SOMETIMES	OFTEN	VERY OFTEN
				
IN THE LAST WEEK, DID YOU EXPERIENCE QUARRELS OR DISAGREEMENTS WITH ANY CLASSMATE?	▢	▢	▢	▢

6. USUALLY, HOW DO YOU COMMUTE TO SCHOOL?
▢ WALKING (INCLUDING SCOOTER, SKATE, ETC.)
▢ BY BIKE
▢ WITH PUBLIC TRANSPORT (TRAM, BUS, SUBWAY, ETC.)
▢ BY CAR

7. HOW LONG DOES IT USUALLY TAKE YOU TO COMMUTE FROM HOME TO SCHOOL?
▢ 0-5 MINUTES
▢ 5-15 MINUTES
▢ 15-30 MINUTES
▢ MORE THAN 30 MINUTES

8. **HOW IMPORTANT IS IT FOR YOU TO FIND SOLUTIONS TO THESE PROBLEMS IN THE CITY?**	**VERY MUCH**	**A BIT**	**NOT MUCH**	**NOT AT ALL**
8.1 GARBAGE IN THE STREET	▢	▢	▢	▢
8.2 LACK OF GREEN AREAS AND PLAYGROUNDS	▢	▢	▢	▢
8.3 LACK OF SPORTS FACILITIES (SWIMMING POOLS, SOCCER FIELDS, VOLLEYBALL FIELDS, BASKETBALL FIELDS)	▢	▢	▢	▢
8.4 AIR POLLUTION	▢	▢	▢	▢
8.5 ROAD TRAFFIC	▢	▢	▢	▢

9.	VERY GOOD	GOOD	BAD	VERY BAD
				
9.1 USUALLY, HOW DOES THE AIR YOU BREATHE FEEL LIKE?	▢	▢	▢	▢
9.2 USUALLY, HOW DOES THE AIR YOU BREATHE IN THE CITY FEEL LIKE?	▢	▢	▢	▢
9.3 USUALLY, HOW DOES THE AIR YOU BREATHE IN THE CLASSROOM FEEL LIKE?	▢	▢	▢	▢
9.4 USUALLY, HOW DOES THE AIR YOU BREATHE IN THE SCHOOL'S COURTYARD FEEL LIKE?	▢	▢	▢	▢

I FILLED IN THIS QUESTIONNAIRE:

▢ BY MYSELF ▢ WITH A BIT OF HELP ▢ WITH A LOT OF HELP

WHO HELPED YOU?

▢ TEACHER ▢ CLASSMATES ▢ SOMEBODY ELSE