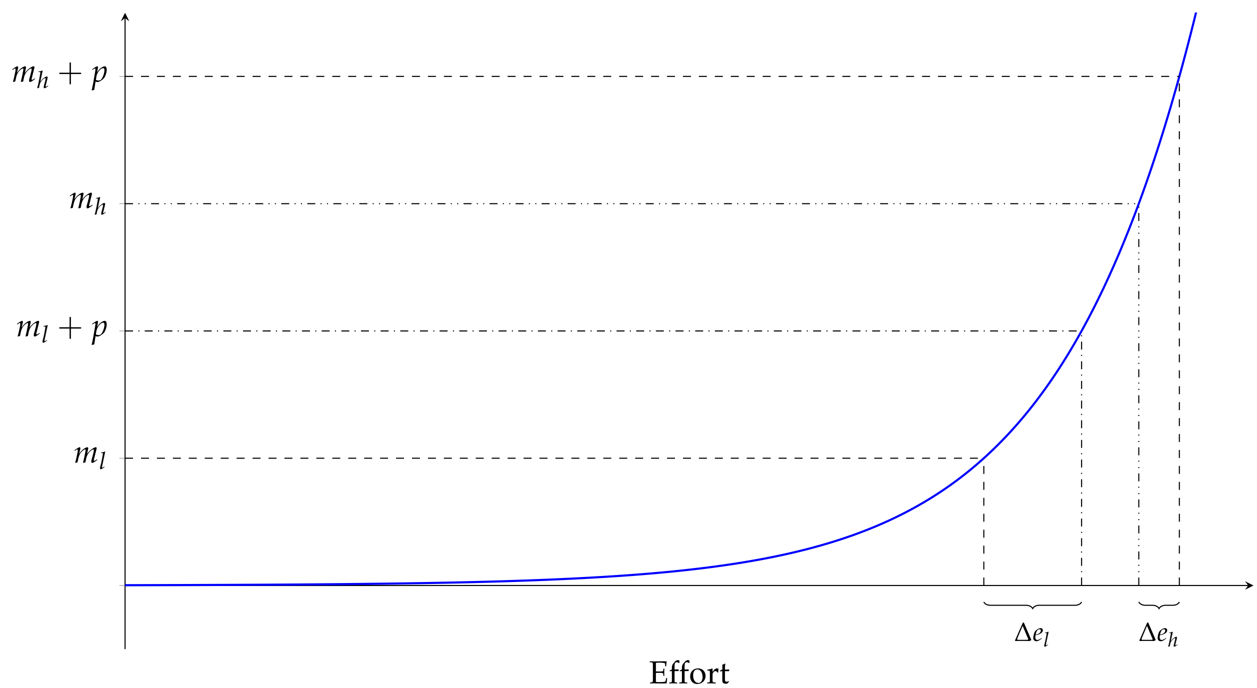
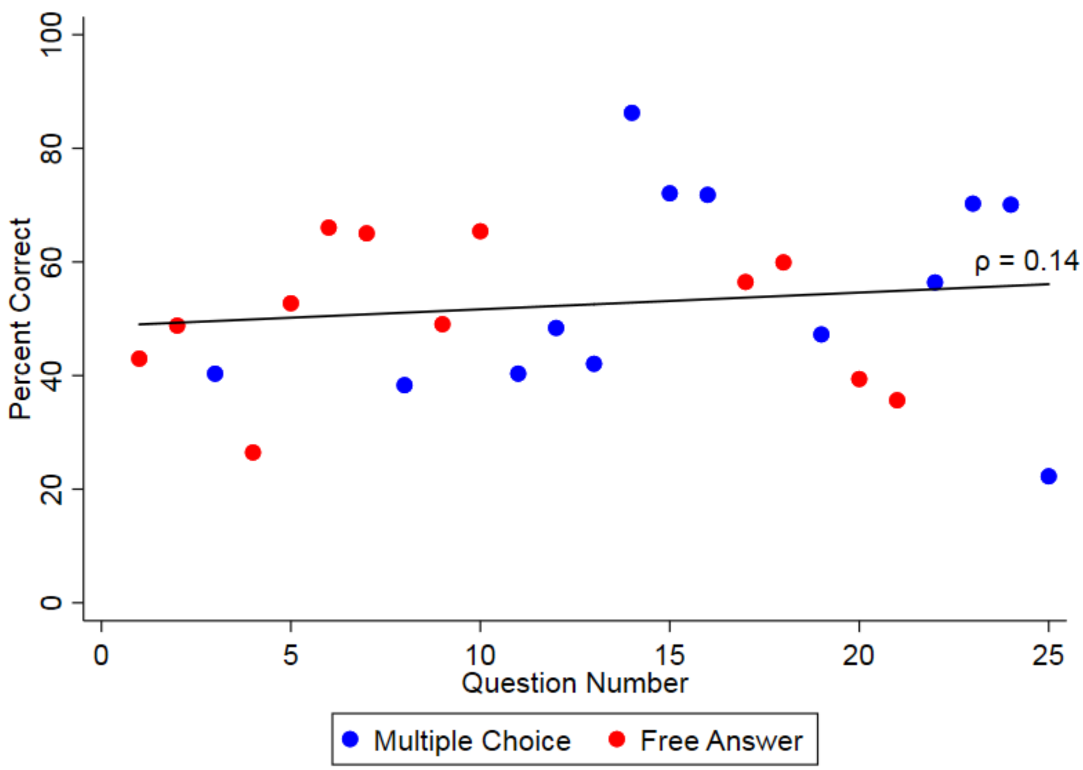
A Appendix Figures and Tables

Figure A.1: Marginal benefits and cost curves



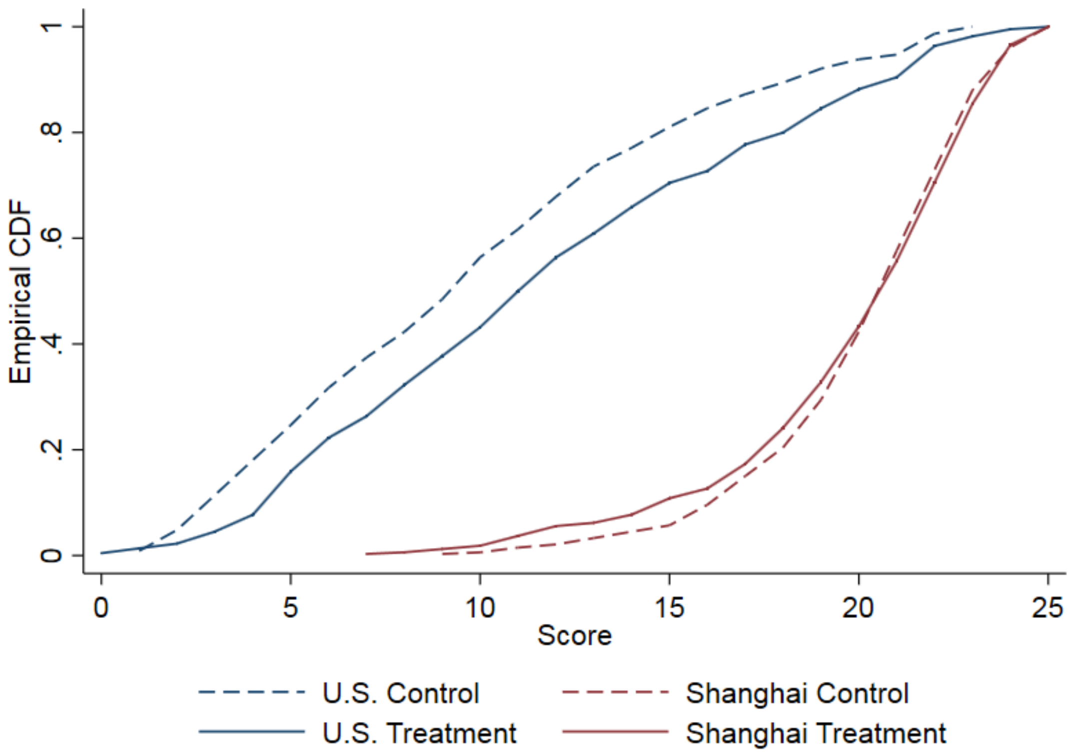
The figure plots the determination of effort at the intersection of marginal benefits (dashed lines) and marginal cost (solid line). Following DellaVigna and Pope (2018), a student chooses effort, e, to maximize: $\max_{e \geq 0} (m + p)e - c(e)$ where m is intrinsic motivation (i.e., motivation absent extrinsic incentives), p is the extrinsic incentive and $c(e)$ is the cost of effort with $c'(e) > 0$ and $c''(e) > 0$. We compare the changes in effort Δe in response to the same extrinsic incentive p for a student with high intrinsic motivation, m_h, compared to a student student with low intrinsic motivation, m_l, where $m_h > m_l$. Due to the convexity of the cost function, the student with low intrinsic motivation will experience larger effort responses than the student with high intrinsic motivation, $\Delta e_l > \Delta e_h$

Figure A.2: PISA worldwide percentage correct



The figure plots the worldwide percentage of students who answered each question correctly when the questions were administered as part of official PISA exams. We calculate the percentage correct using individual-level data available from the OECD.Individual-level responses to every question given on each iteration of the PISA by every participant are available at `http://www.oecd.org/pisa/data`. The percentage correct ranges from 25.7 to 87.3. The correlation between question difficulty and question position on the test is $\rho = 0.14$.

Figure A.3: Distribution of test scores, by treatment group



Tests of equality of Treatment and Control distributions within country: U.S. $p < 0.01$, Shanghai $p = 0.6869$. We estimate p-values using the following non-parametric permutation tests of differences between the test score distributions in each country. We construct test statistics using permutation methods based on Schmid and Trede (1996) and run one-sided tests for stochastic dominance and separatedness of the distributions (see also Imas, 2014). The test statistic identifies the degree to which one distribution lies to the right of the other, and takes into account both the consistency of the differences between the distributions (i.e. how often they cross) and the size of the differences (i.e., the magnitudes). We compute p-values by Monte-Carlo methods with 100,000 repetitions.

Figure A.4: Proportion of questions answered by question and treatment group

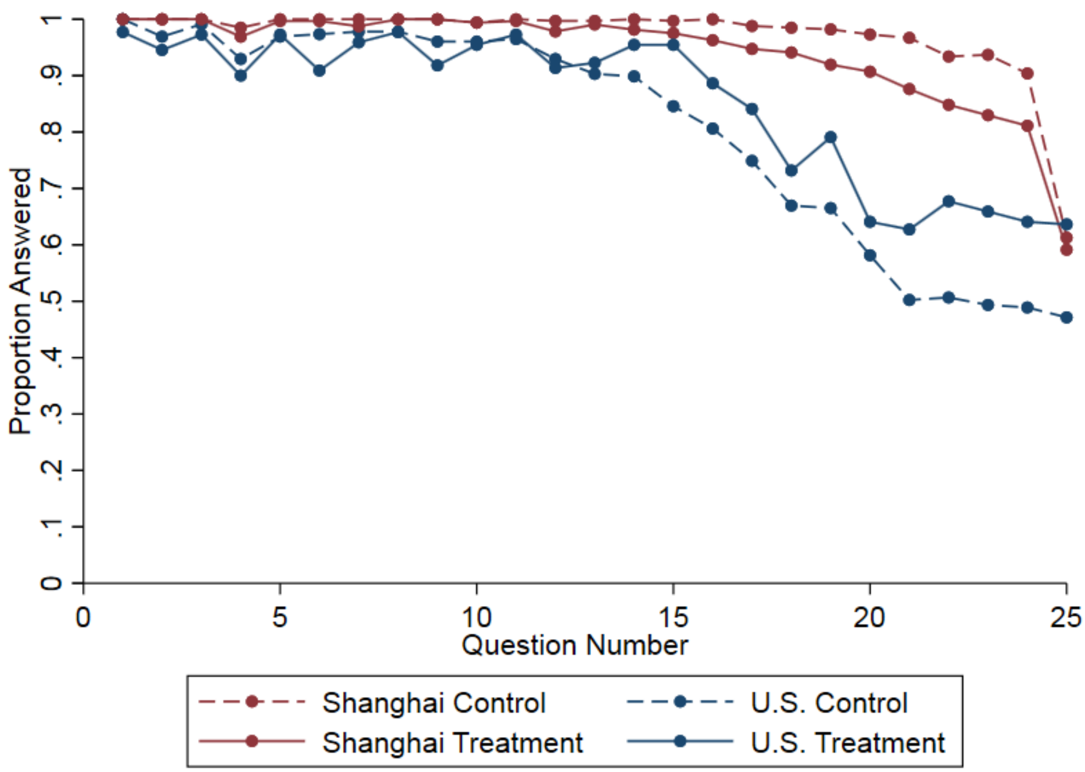


Table A.1: Sensitivity of U.S. treatment effect to sample changes

	Main sample (1)	drop if missing age (2)	keep non-10th (3)	keep ELL (4)	Control for baseline exam score (5)
Treatment	1.34	1.34	1.37	1.37	1.38
(Std. error)	(0.34)	(0.33)	(0.32)	(0.33)	(0.62)
[*p*-value]	[0.001]	[0.001]	[0.001]	[0.001]	[0.007]
Control mean	10.22	10.22	9.59	9.91	11.09
(Standard dev.)	(5.64)	(5.64)	(5.58)	(5.69)	(5.71)
Students	447	446	534	469	348
Clusters	133	132	132	135	123

Notes: OLS estimates of equation (1). Robust standard errors clustered by class (except U.S. school 2, which was randomized at the individual level) in parentheses. *p*-values in brackets. Inference in each column is based on a randomization test using the procedure of Young (forthcoming). All columns include school-track fixed effects and the following covariates: age, gender, and race/ethnicity. Column 3 includes a control for non-10th graders, and Column 4 includes a control for ELL students.

Table A.2: Treatment effects by predicted test score: Threshold regressions, U.S.

	Score		Attempted	Proportion Correct	
Predicted Score Threshold:	< 10.66	≥ 10.66	n/a	< 10.15	≥ 10.15
	(1)	(2)	(3)	(4)	(5)
Treatment	0.76	2.45	1.01	0.026	0.060
(Std. error)	(0.61)	(0.70)	(0.48)	(0.023)	(0.026)
[p-value]	[0.241]	[0.003]	[0.048]	[0.259]	[0.045]
Control mean	7.34	15.04	20.19	0.380	0.688
(Std. deviation)	(3.66)	(5.05)	(5.00)	(0.161)	(0.171)
School-track FE	No	No	No	No	No
Covariates	Yes	Yes	Yes	Yes	Yes
Std. effect size	0.11	0.33	0.23	0.11	0.22
Students	250	197	447	250	197
Clusters	29	123	133	29	123

Notes: The table reports results from threshold regressions where the number of thresholds is estimated by minimizing the Bayesian Information Criterion and the threshold is the value of γ that minimizes the sum of squared residuals $S_{T1}(\gamma) = \sum_{t=1}^{T}\{Y_{ics} - (\alpha^1 + \beta_1^1 Z_c + \beta_2^1 X_i)I((-\infty < w_{ics} \leq \gamma) - (\alpha^2 + \beta_1^2 Z_c + \beta_2^2 X_i)I(\gamma < w_{ics} < \infty)\}^2$. Robust standard errors clustered by class (except U.S. school 2, which was randomized at the individual level) in parentheses. p-values in brackets. Inference is adjusted for multiple hypothesis testing of estimates from two subsamples by controlling the family-wise error rate using the free step-down resampling methodology of Westfall and Young (1993) in columns 1-2 and columns 4-5. Inference in column 3 is based on a randomization test using the procedure of Young (forthcoming). All columns include the following covariates: age, gender, race/ethnicity. School-track fixed effects are not controlled for because they are collinear with predicted score.

Table A.3: Effect of incentives on test scores, by gender

	U.S.		Shanghai	
	Male	Female	Male	Female
	(1)	(2)	(3)	(4)
Treatment	1.67	0.97	0.07	-0.41
(Std. error)	(0.55)	(0.37)	(0.38)	(0.34)
[*p*-value]	[0.019]	[0.019]	[0.870]	[0.477]
Control mean	10.36	10.08	20.51	20.50
(Std. deviation)	(5.97)	(5.31)	(2.93)	(2.96)
Observations	226	221	308	348
Clusters	86	71	179	213

Notes: OLS estimates of equation (1). Robust standard errors clustered by class (except U.S. school 2 and Shanghai schools visited in 2018, which were randomized at the individual level) in parentheses. p-values in brackets. We adjust p-values within each country for multiple hypothesis testing of estimates from two subsamples by controlling the family-wise error rate using the free step-down resampling methodology of Westfall and Young (1993). All columns include school-track fixed effects and the following covariates: age, race/ethnicity (U.S. only), and a wave fixed effect (Shanghai only).

B Test Questions

Question 1

Mark (from Sydney, Australia) and Hans (from Berlin, Germany) often communicate with each other using "chat" on the internet. They have to log on to the internet at the same time to be able to chat.

To find a suitable time to chat, Mark looked up a chart of world times and found the following:

At 7:00 PM in Sydney, what time is it in Berlin?

NOTE: In your answer, please specify the hour, minutes, and whether it is AM or PM. For example, if your answer is 3 PM, write your answer as 3:00 PM.

<< >>

Question 2

To complete one set of bookshelves a carpenter needs the following components:

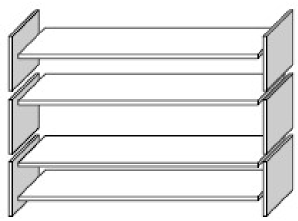
4 long wooden panels,

6 short wooden panels,

12 small clips,

2 large clips and

14 screws.



The carpenter has in stock 26 long wooden panels, 33 short wooden panels, 200 small clips, 20 large clips and 510 screws.

How many sets of bookshelves can the carpenter make? (units not required)

<<

>>

Question 3

A documentary was broadcast about earthquakes and how often earthquakes occur. It included a discussion about the predictability of earthquakes.

A geologist stated: "In the next twenty years, the chance that an earthquake will occur in Zed City is two out of three".

Which of the following best reflects the meaning *of the geologist's statement*?

- ○ 2/3 x 20 = 13.3, so between 13 and 14 years from now there will be an earthquake in Zed City.
- ○ 2/3 is more than 1/2, so you can be sure there will be an earthquake in Zed City at some time during the next 20 years.
- ○ The likelihood that there will be an earthquake in Zed City at some time during the next 20 years is higher than the likelihood of no earthquake.
- ○ You cannot tell what will happen, because nobody can be sure when an earthquake will occur.

<<   >>

Question 4

Infusions (or intravenous drips) are used to deliver fluids and drugs to patients.

Nurses need to calculate the drip rate, D, in drops per minute for infusions.

They use the formula:

$$D = \frac{dv}{60n}$$

, where

d is the drop factor measured in drops per milliltre (mL)

v is the volume in mL of the infusion

n is the number of hours the infusion is required to run

Nurses need to calculate the volume of the infusion, v, from the drip rate, D.

An infusion with a drip rate of 50 drops per minute has to be given to a patient for 3 hours. For this infusion, the drop factor is 25 drops per milliliter.

What is the volume in mL of the infusion? (units not required)

<<

>>

Question 5

You are making your own dressing for a salad.

Here is a recipe for 100 milliliters (mL) of dressing.

Salad Oil:	60 mL
Vinegar:	30 mL
Soy sauce:	10 mL

How many milliliters (mL) of salad oil do you need to make 150 mL of this dressing? (units not required)

<<

>>

Question 6

A car magazine uses a rating system to evaluate new cars, and gives the award of "The Car of the Year" to the car with the highest total score. Five new cars are being evaluated, and their ratings are shown in the table.

Car	Safety Features (S)	Fuel Efficiency (F)	External Appearance (E)	Internal Fittings (T)
Ca	3	1	2	3
M2	2	2	2	2
Sp	3	1	3	2
N1	1	3	3	3
KK	3	2	3	2

The ratings are interpreted as follows:

3 points = Excellent
2 points = Good
1 point = Fair

To calculate the total score for a car, the car magazine uses the following rule, which is a weighted sum of the individual score points:

$$\text{Total Score} = (3 \times S) + F + E + T$$

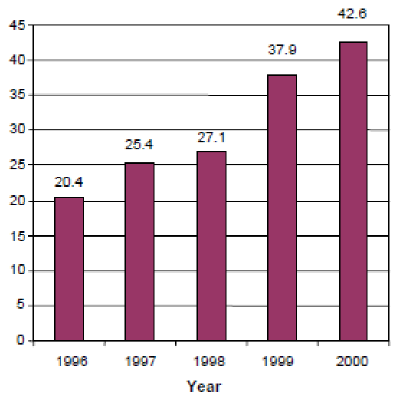
Calculate the total score for Car "Ca". Write your answer in the space below. (units not required)

<< >>

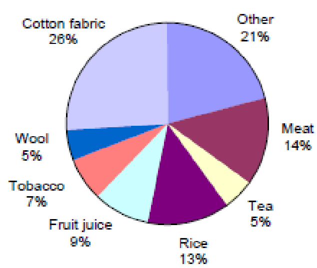
Question 7

The graphics below show information about exports from Zedland, a country that uses zeds as its currency.

Total annual exports from Zedland in millions of zeds, 1996-2000



Distribution of exports from Zedland in 2000



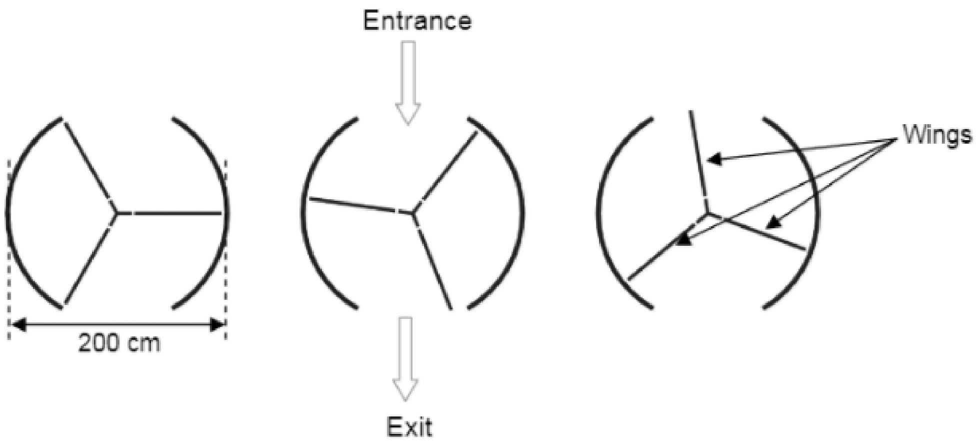
What was the total value (in millions of zeds) of exports from Zedland in 1998? (units not required)

<<

>>

Question 8

A revolving door includes three wings which rotate within a circular-shaped space. The inside diameter of this space is 2 meters (200 centimeters). The three door wings divide the space into three equal sectors. The plan below shows the door wings in three different positions viewed from the top.



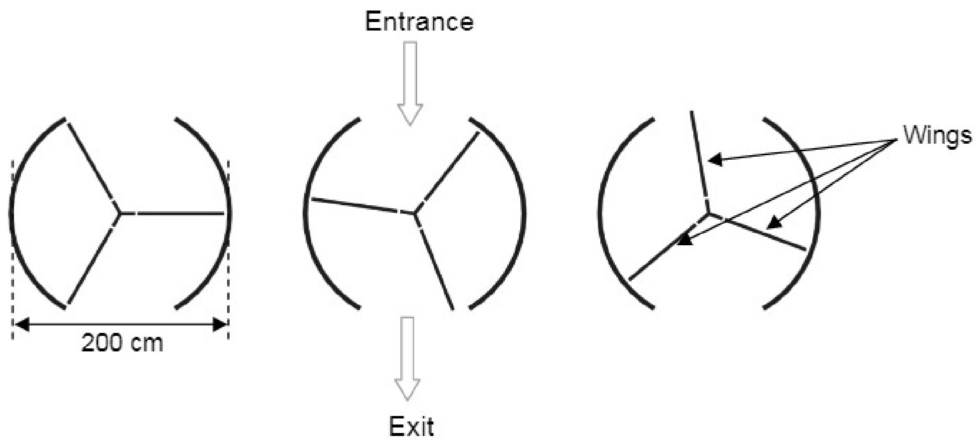
The door makes 4 complete rotations in a minute. There is room for a maximum of two people in each of the three door sectors.

What is the maximum number of people that can enter the building through the door in 30 minutes?

- ○ 60
- ○ 180
- ○ 240
- ○ 720

<< >>

Question 9

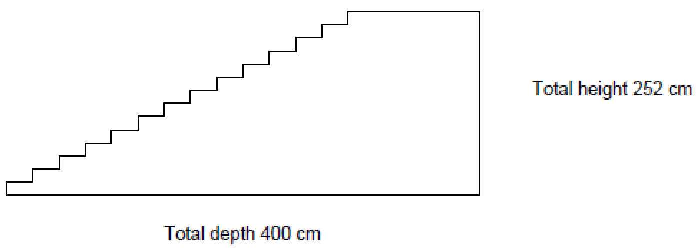


What is the size in degrees of the angle formed by two door wings? (units not required)

<< >>

Question 10

The diagram below illustrates a staircase with 14 steps and a total height of 252 cm:

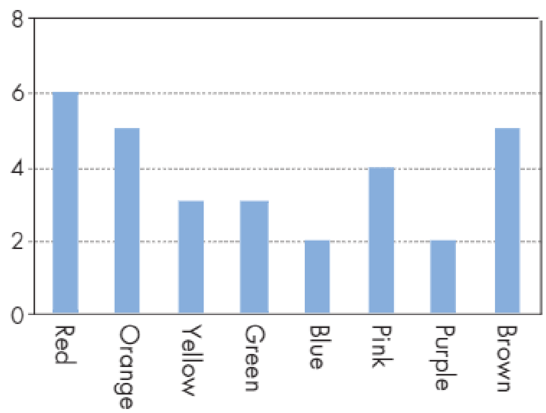


What is the height of each of the 14 steps (in cm)? (units not required)

<< >>

Question 11

Robert's mother lets him pick one candy from a bag. He can't see the candies. The number of candies of each color in the bag is shown in the following graph.



What is the probability that Robert will pick a red candy?

- 10%
- 20%
- 25%
- 50%

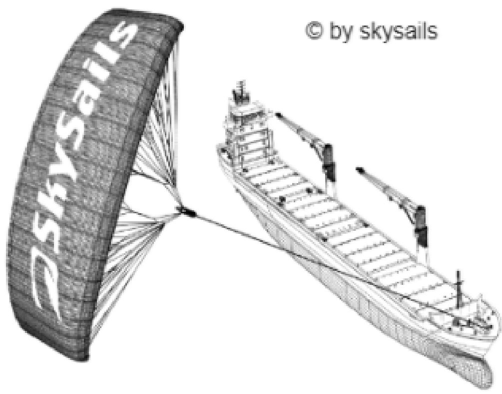
<<

>>

Question 12

Ninety-five percent of world trade is moved by sea, by roughly 50,000 tankers, bulk carriers and container ships. Most of these ships use diesel fuel.

Engineers are planning to develop wind power support for ships. Their proposal is to attach kite sails to ships and use the wind's power to help reduce diesel consumption and the fuel's impact on the environment.



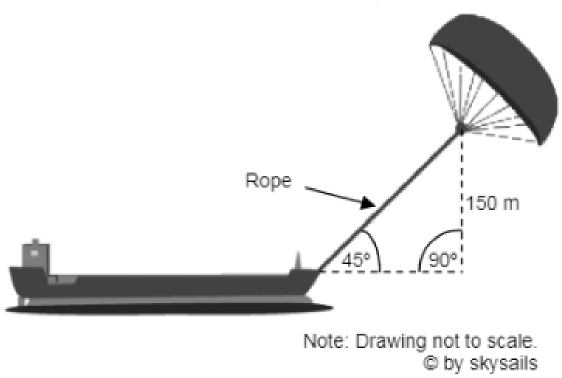
One advantage of using a kite sail is that it flies at a height of 150 m. There, the wind speed is approximately 25% higher than down on the deck of the ship.

At what approximate speed does the wind blow into a kite sail when a wind speed of 24 km/h is measured on the deck of the ship?

- ○ 6 km/h
- ○ 18 km/h
- ○ 25 km/h
- ○ 30 km/h
- ○ 49 km/h

<< >>

Question 13



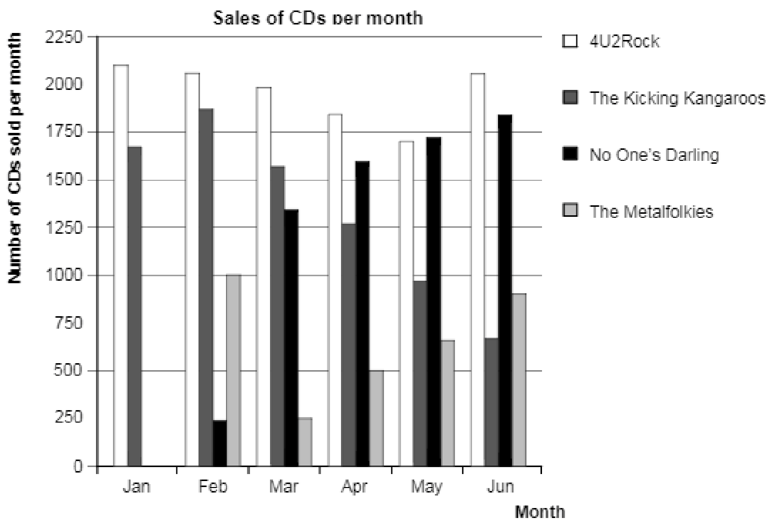
Approximately what is the length of the rope for the kite sail, in order to pull the ship at an angle of 45 degrees and be at a vertical height of 150 m, as shown in the diagram above?

- 173 m
- 212 m
- 285 m
- 300 m

<< >>

Question 14

In January, the new CDs of the bands 4U2Rock and The Kicking Kangaroos were released. In February, the CDs of the bands No One's Darling and The Metalfolkies followed. The following graph shows the sales of the bands' CDs from January to June.

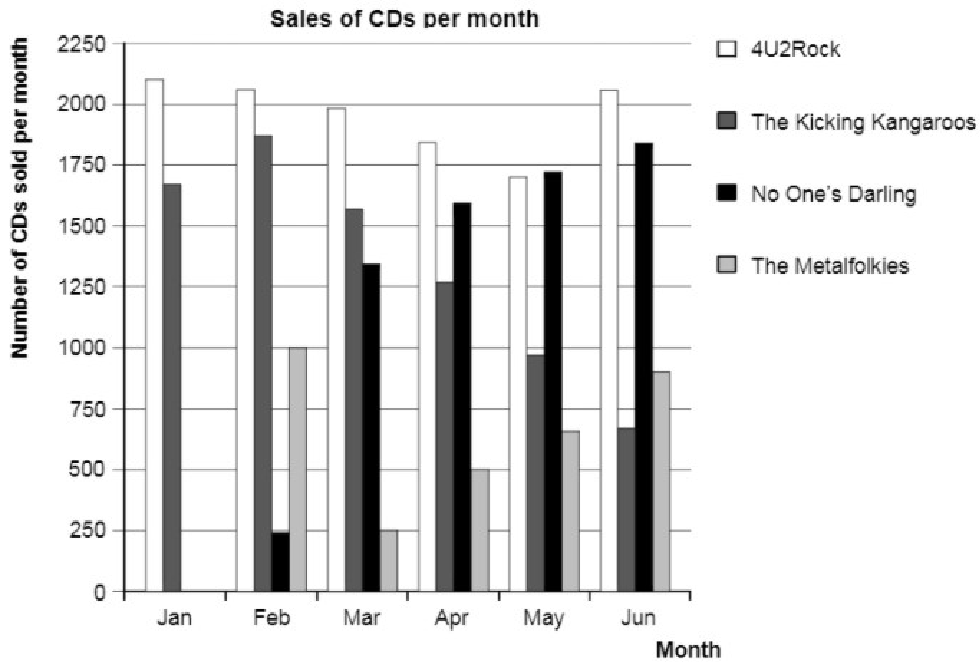


How many CDs did the band The Metalfolkies sell in April?

- ○ 250
- ○ 500
- ○ 1000
- ○ 1270

<< >>

Question 15



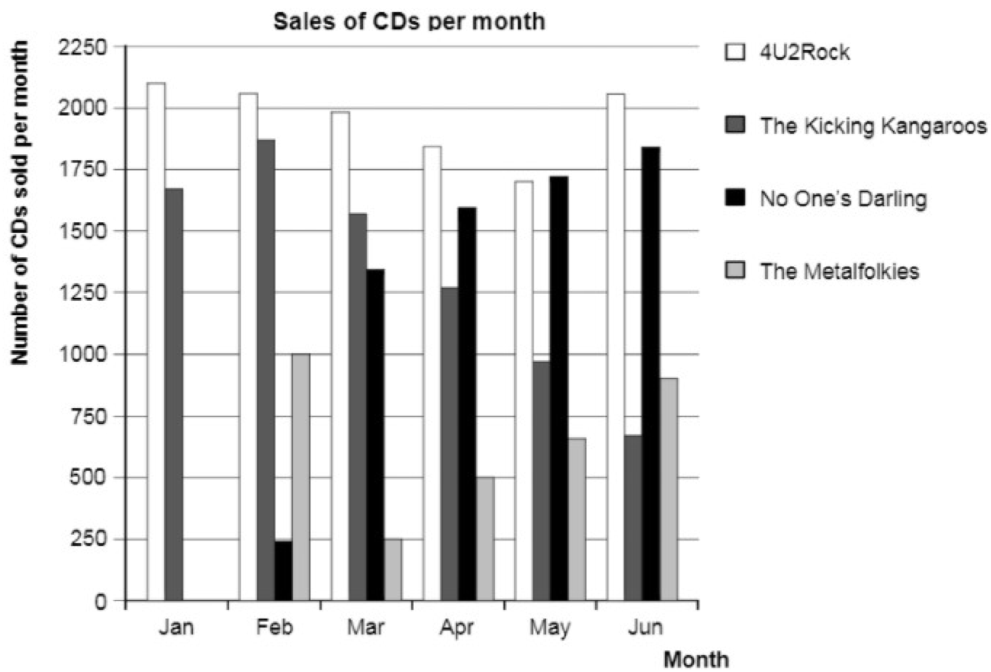
In which month did the band No One's Darling sell more CDs than the band The Kicking Kangaroos for the first time?

- ◯ No Month
- ◯ March
- ◯ April
- ◯ May

<<

>>

Question 16



The manager of *The Kicking Kangaroos* is worried because the number of their CDs that sold decreased from February to June.

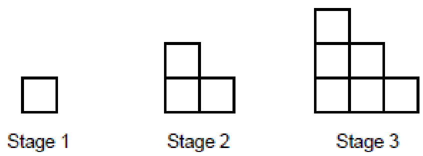
What is the estimate of their sales volume for July if the same negative trend continues?

- ○ 70 CDs
- ○ 370 CDs
- ○ 670 CDs
- ○ 1340 CDs

<<   >>

Question 17

Robert builds a step pattern using squares. Here are the stages he follows.



As you can see, he uses one square for Stage 1, three squares for Stage 2 and six for Stage 3.

How many squares should he use for the fourth stage? (units not required)

<< >>

Question 18

On returning to Singapore after 3 months, Mei-Ling had 3,900 ZAR left. She changed this back to Singapore dollars, noting that the exchange rate had changed to:

1 SGD = 4.0 ZAR

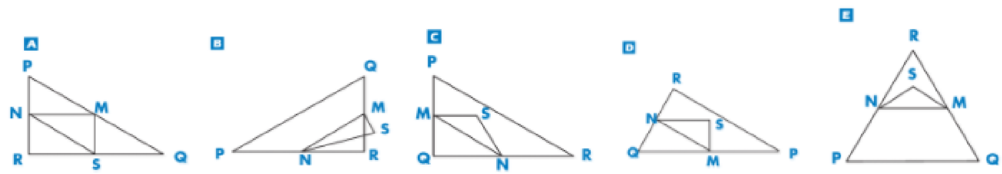
How much money in Singapore dollars did Mei-Ling get? (units not required)

<< >>

Question 19

Choose the one figure below that fits the following description.

Triangle PQR is a right triangle with right angle at R. The line RQ is less than the line PR. M is the midpoint of the line PQ and N is the midpoint of the line QR. S is a point inside the triangle. The line MN is greater than the line MS.



○ ○ ○ ○ ○

<<   >>

Question 20

In a pizza restaurant, you can get a basic pizza with two toppings: cheese and tomato. You can also make up your own pizza with **extra** toppings. You can choose from four different extra toppings: olives, ham, mushrooms and salami.

Ross wants to order a pizza with two different **extra** toppings.

How many different combinations can Ross choose from? (units not required)

<<   >>

Question 21

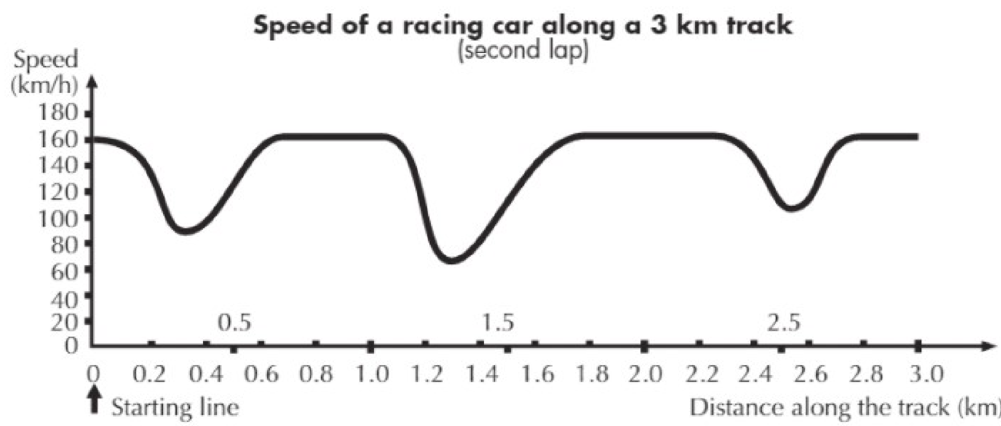
In Mei Lin's school, her science teacher gives tests that are marked out of 100. Mei Lin has an average of 60 marks on her first four Science tests. On the fifth test she got 80 marks.

What is the average of Mei Lin's marks in Science after all five tests? (units not required)

<< >>

Question 22

This graph shows how the speed of a racing car varies along a flat 3 kilometre track during its second lap.

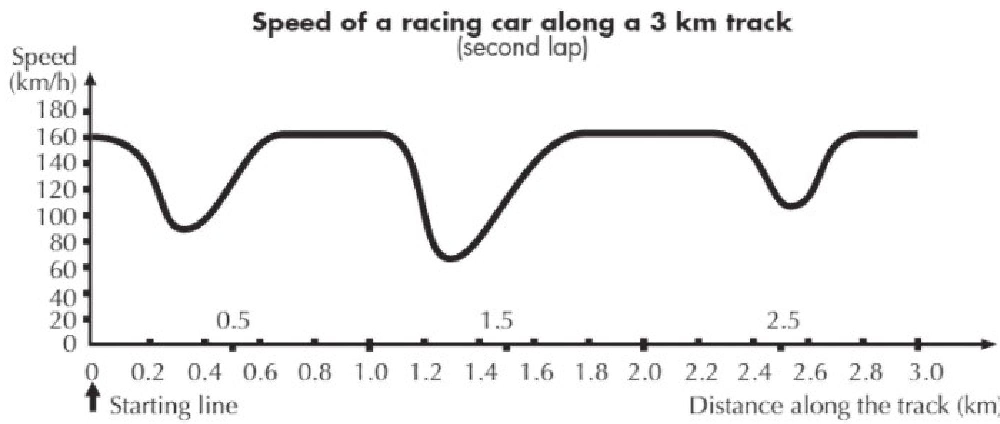


What is the approximate distance from the starting line to the beginning of the longest straight section of the track?

- ○ 0.5 km
- ○ 1.5 km
- ○ 2.3 km
- ○ 2.6 km

<< >>

Question 23



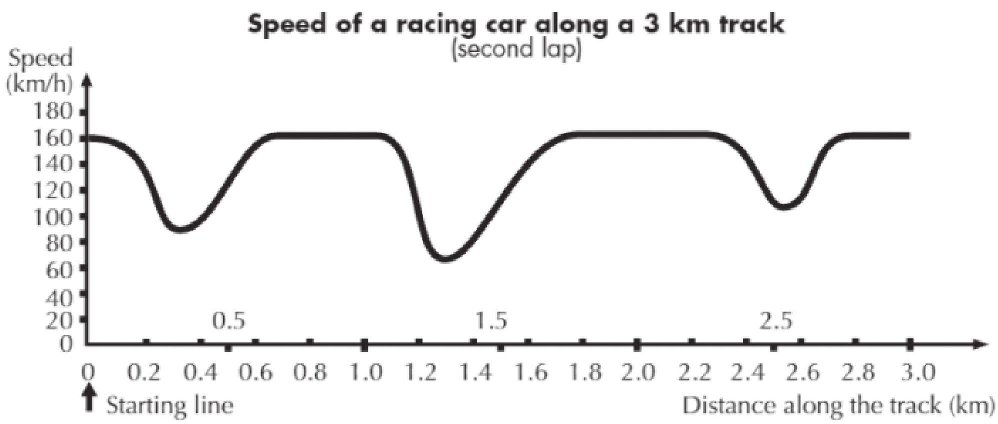
Where was the lowest speed recorded during the second lap?

- ○ at the starting line.
- ○ at about 0.8 km.
- ○ at about 1.3 km.
- ○ halfway around the track.

<<

>>

Question 24

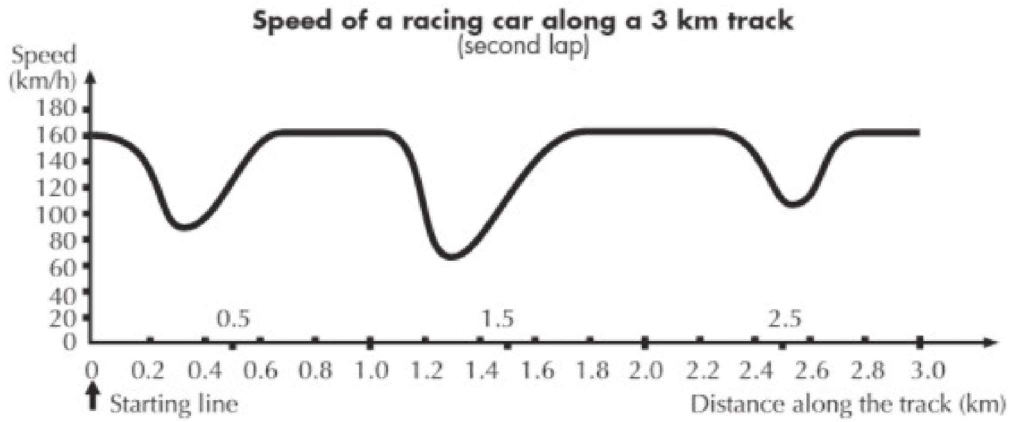


What can you say about the speed of the car between the 2.6 km and 2.8 km marks?

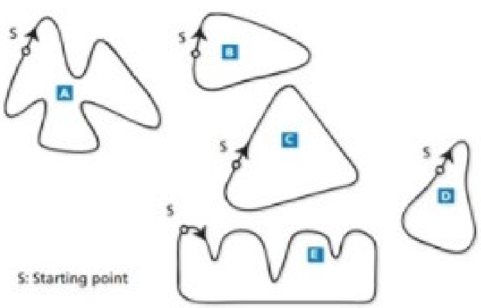
- ◯ The speed of the car remains constant.
- ◯ The speed of the car is increasing.
- ◯ The speed of the car is decreasing.
- ◯ The speed of the car cannot be determined from the graph.

<< >>

Question 25



Here are pictures of five tracks:



Along which one of these tracks was the car driven to produce the speed graph shown earlier?

- A
- B
- C
- D
- E

<< >>

C Instructions

C.1 Control Instructions

Hello and thank you for participating in our study. Today you will be asked to complete a 25-question math quiz. Your payment today will not depend on your performance on the quiz.

You may use the pen, paper and calculator provided, but all answers must be entered on the computer. You will have 25 minutes to complete the math quiz. Do not start the quiz until you are told to do so, and stop when you are asked to. During the quiz you can go back and change your answers, but after the last question your answers will be submitted and you will be unable to change them.

If you have questions about these instructions, please raise your hand and a test administrator will come to you. However, once the exam begins, you will be unable to ask questions until it is over.

When you are finished, the final screen will show you your score on the quiz and the ID number that you were assigned. When called upon to do so, please write your score and ID number on a piece of paper and bring this piece of paper to the front of the room so we can make sure your score is recorded correctly.

Please enter your ID number below and press the button at the bottom right of the screen to continue.

C.2 Treatment Instructions

Hello and thank you for participating in our study. Today you will be asked to complete a 25-question math quiz. Your payment today will depend on your performance on the quiz. You are being given an envelope that contains $25. Please open the envelope to make sure there is $25 inside. Then write your ID number that was assigned to you on the outside of the envelope.

While this money is yours, we will take away $1 for each question you answer incorrectly. Unanswered questions count as questions answered incorrectly. At the end of the quiz, we will subtract the number of questions you answered incorrectly from 25 and that will be your final payment.

Please sign the form that says this is your $25, but that you may have to give some back depending on your score on the quiz.

You may use the pen, paper and calculator provided, but all answers must be entered on the computer. You will have 25 minutes to complete the math quiz. Do not start the quiz until you are told to do so, and stop when you are asked to. During the quiz you can go back and change your answers, but after the last question your answers will be submitted and you will be unable to change them.

If you have questions about these instructions, please raise your hand and a test administrator will come to you. However, once the exam begins, you will be unable to ask questions until it is over.

When you are finished, the final screen will show you your score on the quiz along with the ID number that you were assigned. When asked to do so, write your score and ID number on a piece of paper and bring it to the test administrators at the front of the room so we can make sure your score is recorded correctly.

Please enter your ID number below and press the button at the bottom right of the screen to continue.

References

DellaVigna, S. and D. Pope, "What motivates effort? Evidence and expert forecasts," *The Review of Economic Studies*, 2018, *85* (2), 1029–1069.

Imas, Alex, "Working for the "warm glow": On the benefits and limits of prosocial incentives," *Journal of Public Economics*, 2014, *114*, 14–18.

Schmid, F. and M. Trede, "Testing for first-order stochastic dominance: A new distribution-free test," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 1996, *45* (3), 371–380.

Westfall, P.H. and S.S. Young, *Resampling-based Multiple Testing: Examples and Methods for p-value Adjustment*, New York: John Wiley and Sons, 1993.

Young, A., "Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results," *Quarterly Journal of Economics*, forthcoming.