

Academic Reading Paper 16

Canals, Consonants and Pilots

Time allowed: 1 hour

INSTRUCTIONS TO CANDIDATES

- Do not open this booklet until you are told to do so.
- Write your name and the date at the top of the answer sheet.
- Answer **all** the questions.
- Write your answers **on the answer sheet**, not in this booklet.
- You must transfer your answers within the hour. **No extra time is given for transferring answers in the Reading test.**
- Do not remove this booklet from the examination room.

INFORMATION FOR CANDIDATES

- There are 40 questions in this paper.
- Each question carries one mark.
- There are three passages, with 13, 13 and 14 questions.

These passages were written for this library. They are **not** reproduced from any published past paper. The format, timing and task types follow the published specification for the Academic Reading test so that sitting this on paper feels like the real thing.

You should spend about 20 minutes on Questions 1–13, which are based on Reading Passage 1 below.

Two mistakes at Panama

The French attempt to cut a canal through Panama began in 1881 under the man who had built the one at Suez. Ferdinand de Lesseps was seventy-five, he was famous, and he had raised the money from several hundred thousand small investors who trusted the name, many of whom had put in sums they could not afford to lose. The project was abandoned eight years later with the company bankrupt, the excavation less than half done, and something in the region of twenty-two thousand people dead. Nothing about that failure was mysterious afterwards. It rested on two assumptions, one about earth and one about disease, and both were wrong in ways that were knowable at the time.

The first was the design. Suez runs through flat desert at sea level, and de Lesseps insisted that Panama be cut the same way, straight through from ocean to ocean with no locks. Panama is not flat. A ridge of mountains runs down the isthmus, and the Chagres River crosses the line of the canal and rises many metres in a single storm. A sea-level cut through that country meant removing a mountain and controlling a river that could not be controlled, and the deeper the excavation went the more the sides of it slid back in. The rainy season stopped work for months at a time, rusted the machinery where it stood, and turned the spoil heaps into mud that ran back into the trench. Engineers on the ground said so. The design was not changed.

The second assumption was medical and was shared by almost everybody. Yellow fever and malaria were believed to come from bad air rising off wet ground, which is why the hospitals were built in the healthiest-looking places and why the wards were kept scrupulously clean. The legs of the hospital beds stood in shallow pans of water, to stop ants climbing into them. Those pans were mosquito nurseries, in a building full of infected patients, tended by nurses who had no reason to think anything was wrong.

The correction was already available. In 1881, the same year the digging began, a Cuban physician named Carlos Finlay proposed that yellow fever was carried from person to person by a mosquito, and identified the species. The proposal was published, was not obviously absurd, and was ignored for nineteen years. Part of the reason was that Finlay could not demonstrate transmission convincingly; his own attempts to infect volunteers produced results that were too mild and too irregular to persuade anyone, because he did not know that the insect has to feed on a patient early in the illness and then wait almost two weeks before it can pass the disease on. In 1900 an American commission led by Walter Reed tested it in Havana with human volunteers and established that Finlay had been right.

The Americans, who took the canal over in 1904, therefore began with something the French had not had. William Gorgas, who had run the campaign that removed yellow fever from Havana, was put in charge of sanitation and did the same thing at a larger scale: fumigating buildings, screening windows and doors, draining or oiling every pool of standing water within a set distance of habitation, and re-

building the water supply so that households stopped storing water in open jars. Yellow fever disappeared from the isthmus within about two years. Malaria, whose mosquito breeds in a wider range of places, was greatly reduced and never eliminated.

The engineering was rethought at the same time and by an equally direct argument. John Stevens, the chief engineer from 1905, said that a sea-level canal was not buildable in that terrain in any reasonable time, and proposed instead to dam the Chagres, use the resulting lake as most of the crossing, and lift ships up to it and down again with locks. Damming the river turned the principal obstacle into the principal asset, and the lake it made was for some years the largest artificial body of water in the world. The plan was approved in 1906.

Even then the hardest section stayed hard. The cut through the ridge at Culebra was excavated for years against continuous landslides, in which the walls flowed back into the channel faster than they could be cleared, and the total volume removed went on rising long after the engineers had stopped predicting it. Whole hillsides moved as single masses, carrying the railway tracks laid to remove the spoil down into the cut with them, so that the first job after each slide was to dig out the equipment for digging. The canal opened in August 1914. Deaths during the American period were about five thousand six hundred.

The comparison is usually drawn as American competence against French failure, and that is not quite what the record shows. The French had worse luck with timing: they began in the year the mosquito hypothesis was published and gave up before it was proved. What they did not have was any mechanism for letting the men on the ground overrule a decision taken in Paris by somebody whose previous project had made the decision look obvious.

Questions 1–6

Complete the table below.

Choose **NO MORE THAN TWO WORDS AND/OR A NUMBER** from the passage for each answer.

Two attempts at the same canal		
	French, from 1881	American, from 1904
Canal design	a 1 cut, with no locks, as at Suez	a dammed river, a lake and locks
The river	the Chagres was an obstacle that rose many metres in a single 2	the river was dammed and became the crossing
Theory of disease	fever was thought to come from bad 3 rising off wet ground	fever was known to be carried by mosquitoes

Hospital practice	bed legs stood in pans of water to keep off 4	buildings fumigated and windows screened
Result for yellow fever	a major cause of the deaths	gone from the isthmus within about 5 years
Deaths	around twenty-two thousand	about five thousand 6

Questions 7–13

Do the following statements agree with the information given in Reading Passage 1?

Write **TRUE** if the statement agrees with the information, **FALSE** if the statement contradicts the information, or **NOT GIVEN** if there is no information on this.

- 7** Engineers working in Panama warned that the French design would not work.
- 8** Finlay identified the mosquito as the carrier of yellow fever nineteen years after the French began digging.
- 9** Gorgas eliminated malaria from the isthmus as well as yellow fever.
- 10** Gorgas had worked in Panama before he was put in charge of sanitation there.
- 11** Stevens turned the canal's main problem into part of its solution.
- 12** The amount of material removed at Culebra was accurately predicted in advance.
- 13** Culebra was the most expensive section of the canal to build.

You should spend about 20 minutes on Questions 14–26, which are based on Reading Passage 2 below.

How an infant narrows down a language

- A** A newborn can hear the difference between speech sounds that its parents cannot. The distinctions that separate one language's consonants from another's — the two kinds of t in Hindi, the several k sounds of some languages of the Pacific Northwest, the tones that change a word's meaning in Mandarin — are all audible to an infant a few weeks old, regardless of what is being spoken around it. Adults hear only the distinctions their own language uses. What an infant cannot do is report any of this, so the ability is measured indirectly: the baby is trained to turn its head when a repeating sound changes, and the turn is counted. Somewhere between those two states, something is lost, and the loss is the beginning of competence rather than a decline from it.
- B** The timing was established by Janet Werker and Richard Tees. Testing infants from English-speaking homes, they found that at six to eight months the babies could reliably distinguish contrasts from Hindi and from an indigenous language of British Columbia, neither of which they had ever heard. Tested again between ten and twelve months, the same infants could not. Infants of the same age growing up with those languages could, and did so throughout. The window closes on a schedule and closes on whatever the child has not been given a reason to keep.
- C** Patricia Kuhl described what replaces the lost sensitivity. Exposure to a language builds categories, and a category behaves like a magnet: sounds near the centre of it become harder to tell apart, because the perceptual system has begun treating them as the same thing. This is exactly what a listener needs. A useful hearer of English must treat a hundred slightly different productions of the same vowel as one vowel — spoken by a child, by a stranger, through a bad telephone line, at the end of a tired sentence — and the price of that ability is being unable to hear the differences between them.
- D** The experiment that changed how the process is understood was about who is doing the talking. Kuhl's group exposed nine-month-old American infants to Mandarin for a small number of short sessions, in three conditions. One group heard a live Mandarin speaker who played with them, read to them and responded to what they did. A second group saw the same speaker, saying the same words, on a television screen. A third heard the audio alone. The infants with the live tutor kept their sensitivity to Mandarin contrasts at a level indistinguishable from infants raised in Taiwan. The video group and the audio group showed nothing at all — the same result as infants who had heard no Mandarin.
- E** That finding is awkward for any account in which the infant is simply a statistical device processing whatever arrives at the ear, because the acoustic input in the two conditions was the same. Something about a person being present, responding, and looking where the infant looks appears to be a condition on the learning happening at all, rather than an accompaniment to it.

- F** It is awkward and it is not the whole story, because the statistical account has evidence of its own that is very hard to argue with. Jenny Saffran, Richard Aslin and Elissa Newport played eight-month-olds two minutes of a continuous stream of invented syllables, spoken by a synthesiser in a monotone with no pauses, no stress and no intonation. Every cue that ordinarily marks the edge of a word in speech had been deliberately removed, which was the point of the design. Hidden in the stream were three-syllable sequences that recurred, so that within a sequence one syllable predicted the next reliably and across a boundary it did not. After two minutes the infants distinguished the recurring sequences from equally familiar syllables that spanned a boundary. They had extracted word-like units from nothing but the probabilities.
- G** Put beside each other, the two results describe a learner rather than a recorder. The infant is running a statistical procedure on the stream of sound, and that procedure is powerful enough to find structure in two minutes of nonsense. But in the ordinary case it appears to run only when a person is in front of the child, which is why the second result does not simply cancel the first.
- H** The practical version of this gets stated badly in both directions. It is not that recordings are harmful, and it is not that screen time in itself explains anything. It is that the specific job of sorting a language's sounds into categories, in the first year, seems to require the thing that a recording by definition cannot supply, which is somebody who reacts.

Questions 14–18

Reading Passage 2 has eight sections, A–H.

Which section contains the following information?

Write the correct letter, A–H.

NB You may use any letter more than once.

- 14** a warning against two opposite misreadings of the evidence
- 15** an explanation of why the loss is necessary for ordinary listening
- 16** a description of material that contained no pauses or stress
- 17** a group whose performance matched that of children raised where the language is spoken
- 18** a statement that one result does not cancel another

Questions 19–22

Look at the following findings and the list of researchers below.

Match each finding with the correct researcher, A–C.

NB You may use any letter more than once.

List of Researchers

- A Werker and Tees
- B Patricia Kuhl
- C Saffran, Aslin and Newport

- 19 They established when the ability to hear foreign contrasts disappears.
- 20 She showed that a learned category makes nearby sounds harder to separate.
- 21 They showed that infants can find word boundaries using probability alone.
- 22 She found that identical sound delivered without a person present produced no learning.

Questions 23–26

Complete the summary using the list of words, A–H, below.

List of Words

- A categories
- B vocabulary
- C probabilities
- D repetition
- E recorder
- F learner
- G tutor
- H schedule

What the first year does

During the first year an infant stops hearing distinctions its own language does not use, and builds **23** in their place. The two-minute syllable experiment shows that infants can extract structure from **24** alone. Taken together with the tutoring result, the evidence describes a **25** rather than a passive receiver. The tutoring study showed further that identical material delivered by a screen produced nothing at all, so the procedure appears to run only in the presence of a live **26**

You should spend about 20 minutes on Questions 27–40, which are based on Reading Passage 3 below.

Why pilots succeed and programmes fail

A familiar sequence runs through education, health, policing and development work. Somebody designs an intervention, runs it in a few places, and measures a clear improvement. The result is published, funded and expanded. Some years later a larger evaluation reports that the effect is small, or absent, and the programme is quietly reduced. The usual explanations are that the original result was a fluke or that the implementation was poor. Both happen. Neither is the main thing.

The main thing is that a pilot and a programme are different objects, and the differences all point the same way. A pilot runs in sites that agreed to take part, which are not a random sample of sites; the ones that volunteer tend to be better organised and more interested. It is staffed by people who chose to be involved, often including the person who designed it, working with an attention that cannot be paid to two thousand sites at once. It is small enough to be supervised, so problems are noticed and fixed. At two thousand sites nobody notices, and the version of the programme that is actually being delivered drifts away from the version that was tested without anybody deciding that it should. Every one of those conditions inflates the measured effect, and every one of them disappears on expansion.

There is a statistical contribution as well, and it is not a matter of anyone cheating. A small study needs a large apparent effect to reach significance at all. So among small studies, the ones that get published are disproportionately those in which the effect looked large, whether or not it is. Repeat the same study at scale and the estimate falls towards the truth, which will be read as the programme failing when it is the first estimate correcting itself. The effect is well known and has a name — the winner's curse — and it applies to whichever studies are selected for attention by the size of their result.

Some interventions also stop working for reasons that have nothing to do with measurement. A programme that helps a few hundred people find work may be moving them up a queue rather than creating jobs, in which case running it for two million people cannot produce two million times the result. A tutoring scheme that recruits the best available teachers can do that in one district and not in forty, because at forty districts it is hiring from a different part of the distribution. These are not implementation failures. They are consequences of the size, and they are invisible at pilot scale by construction.

It is worth saying plainly that none of this is an argument against piloting. The alternative to a pilot is a national rollout based on a hunch, and the record of that is far worse. What the argument establishes is narrower: that a pilot answers the question of whether an intervention can work under favourable conditions, and that this is a different question from whether it will work under ordinary ones. Treating the first answer as the second is the error, and it is made by well-informed people constantly.

The remedies follow from the diagnosis and are mostly about design rather than statistics. Pilot with the staff who will actually deliver the programme rather than with the people who invented it. Include some sites that did not volunteer. Run it at a scale where the designer cannot visit, because the thing being

tested is not the idea but the idea as operated by somebody who has read a manual. State in advance what a scaled version would look like, and test that version, rather than testing a version that will never exist again. Where the intervention depends on a resource that is scarce, say what happens when it runs out, and test the programme under that condition rather than the one in which it is abundant.

One further habit is worth breaking. A large evaluation that finds nothing is currently treated as a failure, both by the people who ran the programme and by those who funded it, which creates a strong incentive not to run one. But a null result at scale is a genuine finding about the world: it says that this thing, delivered this way, at this size, does not do what was hoped, which is information that the alternative — continuing on the strength of the pilot — never produces.

The underlying point is not about pilots at all. It is that scale is a property of an intervention rather than a quantity of it, and that a result obtained at one scale carries no automatic licence at another. That is an unwelcome thing to be told by anybody trying to decide what to fund, and it is true in every field where it has been checked.

Questions 27–31

Do the following statements agree with the claims of the writer in Reading Passage 3?

Write **YES** if the statement agrees with the claims of the writer, **NO** if the statement contradicts the claims of the writer, or **NOT GIVEN** if it is impossible to say what the writer thinks about this.

- 27** The writer thinks poor implementation is the main reason programmes disappoint at scale.
- 28** The writer argues that pilot studies should not be carried out.
- 29** Most large evaluations are funded by the organisations that ran the original pilot.
- 30** The writer accepts that a null result from a large study tells us something real.
- 31** The writer believes tutoring schemes are the clearest example of the problem.

Questions 32–36

Complete each sentence with the correct ending, A–H, below.

Endings

- A are better organised than sites in general.
- B needs a large apparent effect in order to be published at all.
- C may be moving people up a queue rather than creating opportunities.
- D tests the idea as operated by somebody who has only read a manual.
- E is information that continuing on the strength of a pilot never produces.
- F shows that the original researchers were dishonest.
- G proves that small studies should be abandoned.
- H is caused entirely by poor supervision.

- 32 Sites that volunteer for a pilot
- 33 A small study
- 34 A job programme that succeeds at small scale
- 35 Running a trial at a size the designer cannot visit
- 36 A large evaluation that finds no effect

Questions 37–40

Choose the correct letter, A, B, C or D.

- 37 According to the passage, when a scaled-up estimate falls, this usually means
 - A the first estimate is correcting itself.
 - B the programme was implemented badly.
 - C the original researchers were dishonest.
 - D the intervention has stopped being useful to anyone.
- 38 Why does the writer mention hiring teachers in forty districts?
 - A To show that teaching quality varies by region.
 - B To give an example of an effect that disappears with size.
 - C To argue that tutoring does not work.
 - D To explain why volunteers make poor staff.

- 39** The writer says a pilot answers the question of whether an intervention
- A** will work under ordinary conditions.
 - B** can work under favourable conditions.
 - C** is cheaper than the alternative.
 - D** is popular with the people who run it.
- 40** Which of the following best states the writer's conclusion?
- A** Scale is a property of an intervention, not simply an amount of it.
 - B** Only large studies should be believed.
 - C** Public money should not be spent on pilots.
 - D** Statistical significance is a meaningless standard.