

# Pre-analysis plan: Increasing participation in the Adult Migrant English Program –potential students

---

## Policy problem

The Adult Migrant English Program (AMEP) is a government-funded program for all permanent visa holders and some temporary visa holders who have less than a vocational level of English. Not all people who are eligible for the AMEP participate in it. The Department of Home Affairs commissioned BETA to increase participation in the program.

## Trial aim

This trial aims to test the effectiveness of letters and emails in English and home languages for increasing participation among eligible people who have never engaged with the AMEP. Previously, new migrants had to enrol in the AMEP within six months of arrival and begin classes within 12 months. New legislation, which took effect in April 2021, removed the timeframes for registration for those in Australia on or before 1 October 2020. This provides migrants who didn't enrol when they first arrived in Australia a second chance to join the AMEP and learn English.

The research will focus on speakers of Mandarin, Cantonese, Vietnamese and Arabic in four service provider regions which cover 85 per cent of AMEP students. These four language groups self-identified in the 2016 census as having lower levels of English proficiency. These languages are highly linguistically different to English which is likely to result in it being harder for speakers of these languages to learn English<sup>1</sup>.

The trial will involve sending letters and emails to people from the four language groups identified above, who have eligible visas and have never enrolled in the AMEP. The messages will be sent in English or participants' home languages.

## Research question

Do letters and emails in participants' home language increase participation in the Adult Migrant English Program compared with letters and emails in English?

---

<sup>1</sup> Chiswick, B. R., & Miller, P. W. (2005). Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development*, 26(1), 1-11.

## Interventions

We will send letters and emails to speakers of Mandarin, Cantonese, Vietnamese and Arabic who are eligible for the AMEP and have never enrolled.

There will be one independent variable, Language - participants will be randomly assigned to receive letters and emails in either their home language or English.

## Outcome measures

The primary outcome measure is enrolments in the AMEP, as measured by registration date. We will be measuring whether people who receive the letters and emails register before the end of March 2022. The outcome measure is binary. Participants either did (1) or did not (0) enrol. From this binary measure we will calculate sample proportions.

The secondary outcome will be the proportion of participants who click on the link embedded in the email or type the link in from the letter. We will not have individual-level data for this outcome, but we will compare proportions for each group.

## Hypothesis

H1: There will be a higher proportion of enrolments among people who receive the letter/email in their home language as compared with those who receive the letter/email in English ( $A1 > A0$ ).

H1 will be assessed using a one-sided test.

## Sample size and power calculations

With an available sample size of approximately 17,000 participants, and an estimate of 1.5 people per cluster, the trial will have around 11,300 clusters with an estimated ICC of 0.4. This will give us 90% power to detect a standardised effect size of  $d = 0.04$  with  $\alpha = 0.05$  using a one-sided test. This is equivalent to approximately 1.5 percentage points difference between groups with the assumption that one per cent of control participants will enrol.

## Sample selection and exclusion criteria

Participants will be drawn from databases owned by the Department of Home Affairs. Participants will be selected based on their visa status, indicating their potential eligibility, home country, language and their address falling within the four service provider regions in Qld, NSW, the ACT or Victoria. Attempts will be made to contact all potential students from the four target language groups in the four service provider regions in Qld, NSW, the ACT or Victoria, based on their last known address and email address. There is an estimated sample of approximately 17,000 potential students covered by the four service provider regions.

Table 1. Sample sizes by home language and provider contract regions

	Provider region 1	Provider region 2	Provider region 3	Provider region 2 & 3	Provider region 4	TOTAL
<b>Arabic</b>	263	58	344	357	60	<b>1082</b>
<b>Cantonese</b>	409	71	483	30	223	<b>1216</b>
<b>Mandarin</b>	4034	1036	4745	220	1875	<b>11910</b>
<b>Vietnamese</b>	745	203	661	742	431	<b>2782</b>
<b>Total</b>	<b>5451</b>	<b>1368</b>	<b>6233</b>	<b>1349</b>	<b>2589</b>	<b>16990</b>

### Randomisation

Potential students will be randomised at the cluster level. Clusters will be constructed so people at the same addresses are in the same treatment group (clustered). To match free text address fields, we will use a ‘fuzzy’ matching strategy, matching similar addresses into clusters. This yields clusters largely comprised of individuals at the same individual addresses, but does cluster some similar addresses together. For example, different units in the same apartment block may be clustered together, or houses on the same street. Participants without an address will be clustered on email address. This means that people using the same email address will receive the same intervention. We will stratify by service provider region and language to ensure balance.

### Trial threats

**Blinding:** Individuals enrolled in the trial will be aware of the messages they receive, but unaware they are involved in a trial.

**Spillovers:** Randomisation will occur at a ‘fuzzy’ address level (to account for errors in address data) to avoid spillovers of people in the same household being in a different treatment group. However, other friends or family could be in a different treatment group and share their letters/emails with each other. We expect that this will not be a common occurrence and will have a negligible impact on our treatment effect estimates.

**Attrition/missing data:** Letters and emails will be sent to the addresses and email addresses last given to the department by potentially eligible visa holders. People may have moved since then making some letters unlikely to reach their intended targets. For this reason emails are also being used as people will be less likely to change their email address.

Letters and emails may also be sent to people who already speak vocational English and are not eligible for the AMEP. Any letters sent back as Return to Sender or

emails that bounce back as undelivered will be taken out of the sample and not analysed.

## Method of analysis

For the primary outcome measure, we will estimate treatment effects using Ordinary Least Squares (OLS) regression with standard errors (CR2) clustered using the following specification:

$$Y = \alpha + \beta_1 A + \gamma_1 X + \gamma_2 X * A + v + w$$

Where  $Y$  is a binary variable indicating whether the student enrolled.  $A$  indicates whether the cluster is allocated to the home language or English language group,  $X$  indicates a vector of block indicators,  $X * A$  is the interaction of this vector with the treatment indicator,  $v$  is a cluster-level error term and  $w$  is the individual-level error term.

We will not adjust p-values for multiple comparisons. We will conduct logistic regression as a robustness check for our primary OLS specification and report average marginal effects.

## Exploratory and subgroup analyses

We may conduct exploratory analyses of outcomes at the cluster level to determine whether the behaviour of one person influences the behaviour of other household members. We may also explore differences between groups that opened the email and those that did not.

There are a number of subgroups that we may investigate: language group, location, gender, age, visa category and status, country of birth.

## Pre-analysis plan commitments

No trial data have been collected/no analysis has been undertaken prior to the completion of this pre-analysis plan.

We will be transparent about, and provide justification for, any deviations (additions or omissions) from this plan.