

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

---

## Policy problem

The Adult Migrant English Program (AMEP) has been running since 1948. It is a government-funded program for all permanent visa holders and some temporary visa holders who have less than a vocational level of English. Around 50% of AMEP participants leave the program within a year and only 1 in 5 leave with functional levels of English. The Department of Home Affairs commissioned BETA to increase participation in the program.

## Trial aim

This trial aims to test the effectiveness of text messages in English and home languages coming from the Department of Home Affairs or service providers for increasing participation among students who were previously engaged in the AMEP but left the program prior to reaching vocational levels of English. Some of these participants would have left of their own accord and others would have had to leave the program due to reaching the legislative end point, either having studied in the AMEP for 510 hours or for five years. New legislation came into effect in 2021 removing these rules making it possible for people in Australia on or before 1 October 2020 to re-engage with the AMEP.

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>.

Some service providers currently send messages in English to disengaged students, encouraging them to return to the AMEP. In this trial, messages in English and the participants' home languages will be trialled. We will also trial two different messengers, the Department of Home Affairs and service providers.

---

<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.

## Research questions

- 1) Does a text message in participants' home language increase re-engagement in the AMEP, compared with a message in English?
- 2) Does the messenger (the Department of Home Affairs or service providers) sending the communication make a difference on the number of people who reengage in the AMEP?

## Outcome measures

The primary outcome measure is re-engagement in the AMEP, indicated by a scheduled activity date (when the participant contacts the service provider to schedule returning to class). We will be measuring whether people who received a text message re-engage before the end of March 2022. The outcome measure is binary. Participants either did (1) or did not (0) reengage by this date. 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 text message. We will not have individual-level data for this outcome, but we will compare proportions for each main effects group.

## Interventions

We will send text messages to speakers of Mandarin, Cantonese, Vietnamese and Arabic who have previously enrolled in the AMEP and have disengaged.

The trial will be a 2x2 factorial design, thus there will be two independent variables: (A) Language - participants will be randomly assigned to receive text messages in either their home language or English (B) Messenger - participants will be randomly assigned to receive text messages from either the Department of Home Affairs or their previous service provider. The table below shows the notation used to refer to the main effects and individual groups formed from our two independent variables.

Table 1 – Main effects, interventions and treatment group nomenclature

		(A) Language	
		English	Home
(B) Messenger	Department	A0B0	A1B0
	Service provider	A0B1	A1B1

## Hypotheses

H1: There will be a higher proportion of re-engagement among people who received a message in their home language as compared with those who received a message in English ( $A1 > A0$ ).

H2: There will be differences in re-engagement between the Department of Home Affairs messenger group and the service provider messenger group ( $B1 \neq B0$ )

H1 will be assessed using a one-sided test and H2 with a two-sided test.

## Sample size and power calculations

With an available sample size of approximately 26,000 participants, a mean of 1.5 people per cluster, the trial will have 17,054 clusters with an estimated ICC of 0.5. This will give us 80% power to detect a standardised effect size of  $h = 0.036$  with  $\alpha = 0.05$  using a single-sided test. As an example, this is the equivalent of a difference of around 0.4 of a percentage point off a base of one per cent.

## Sample selection

Participants will be drawn from the AMEP client database owned by Home Affairs. Four service providers are involved in the trial, they cover approximately 85 per cent of AMEP students. Participants will be selected based on their home language and service provider region. Participants need to have been previously enrolled in the AMEP but left prior to reaching vocational English. Attempts will be made to contact all disengaged students from the four target language groups in the four service provider regions, based on their last known phone number. There is an estimated sample of approximately 26,000 disengaged students covered by the four service provider regions.

Table 2. Sample sizes by home language and provider contract regions

	Service Provider 1	Service Provider 2	Service Provider 3	Service Provider 4	TOTAL
Arabic	4077	2509	1175	3101	10862
Cantonese	88	656	181	225	1150
Mandarin	634	4598	1856	3468	10556
Vietnamese	1752	500	617	689	3558
Total	6551	8263	3829	7483	26126

## Randomisation

The design is 2 (language) x 2 (messenger) as displayed in Table 1. Disengaged students will be randomised at the address level 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. We will stratify by service provider region to ensure balance in treatment group assignment across the four regions.

### **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 different treatment group. However, other friends or family could be in another treatment group and share the messages 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:** If we receive confirmation that text messages were not delivered the record will be removed from the dataset prior to analysis. All other people will be analysed as randomised.

### **Method of analysis**

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

$$Y = \alpha + \beta_1 A + \beta_2 B + \gamma_1 X + \gamma_2 X * A + \gamma_3 X * B + v + w$$

Where  $Y$  is either a binary variable indicating whether the student re-engaged.  $A$  indicates whether the cluster is allocated to the home language or English language group,  $B$  indicates whether the cluster was allocated to the Home Affairs or provider messenger group,  $X$  indicates the set of de-meaned baseline covariates (years of education) and vector of block indicators,  $X * A$ ,  $X * B$  are the interaction of these covariates with the treatment indicators,  $v$  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 analysis and subgroups**

We may conduct exploratory analyses of outcomes at the cluster level to determine whether the behaviour of one person influences the behaviour of other cluster members.

There are a number of subgroups that we may investigate: language group, service provider/location, gender, age, visa category and status, years of schooling, country of birth, people who had to leave as they were no longer eligible vs. people who disengaged of their own volition

### **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.