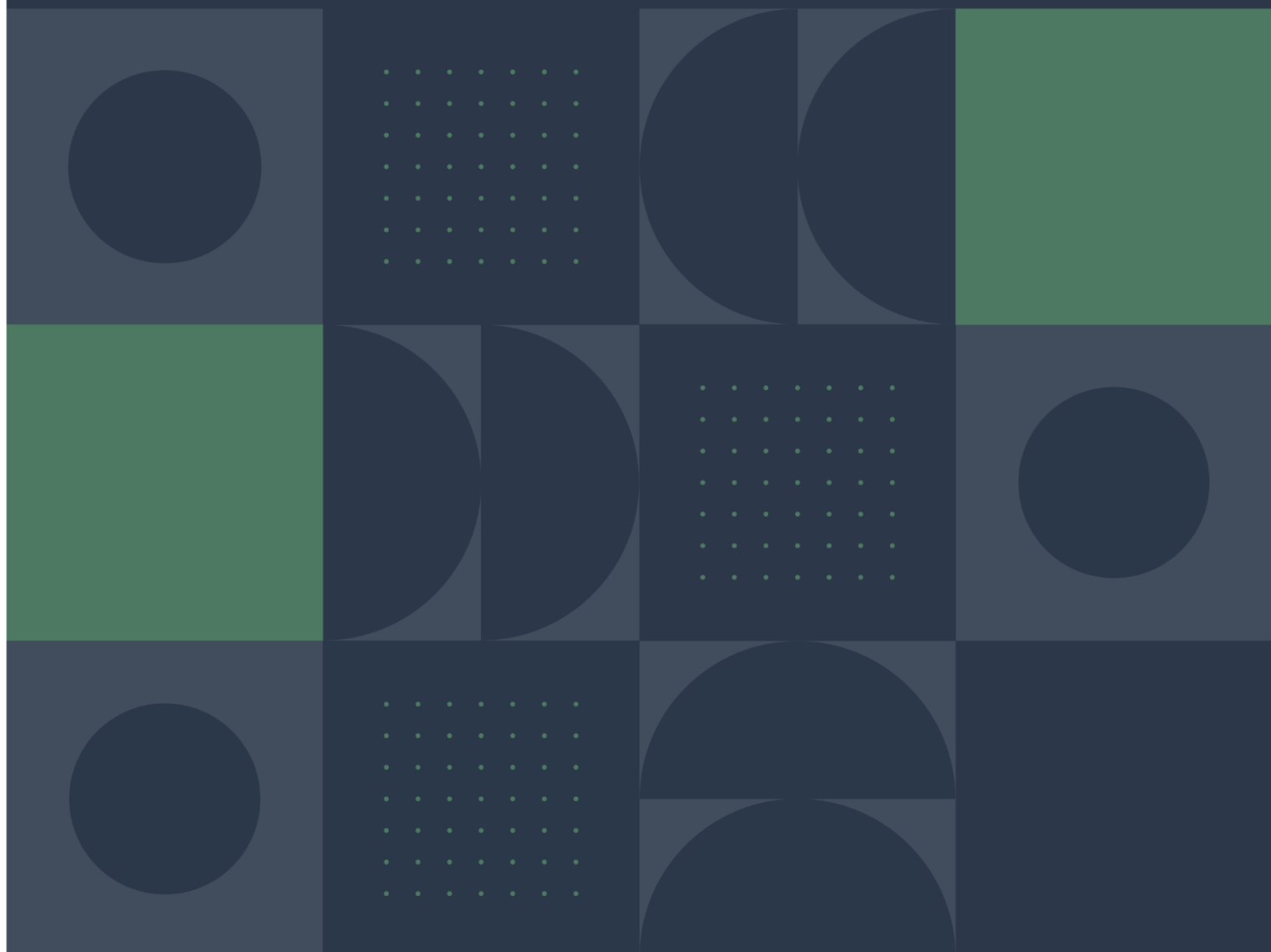# The effect of online training formats on engagement: a randomised trial

December 2025

Ethan Slaven, Peter Bowers, Shantanu Sheshgir, Georgia Kennedy, Scott Copley and Harry Greenwell

Contents

Executive summary

Background: Online training courses are an important and widely-used tool for professional development in the Australian Public Service (APS). These courses include short, introductory modules, which are designed to give participants a brief overview of key concepts as a foundation for further learning, or to help engage with experts in that topic.

This evaluation tested the effectiveness of different *delivery formats* for short, introductory online training developed by the Australian Centre for Evaluation (ACE) and targeted at APS staff. The 3 different formats were: a traditional click-through e-learning module; a micro-learning video; and a podcast. While the courses were all on the same topic (the role of randomisation in evaluation), the formats varied in several ways – the medium (module, video or podcast), the duration, and whether an initial sign-in was required.

Methods: To test the effectiveness of the different training course formats, the ACE conducted a rapid randomised trial over 3 months. Over 750 public servants were recruited to the trial in January and February 2025 through existing email distribution lists. They were then randomised to complete training on the same topic but in one of the 3 different training formats: a 3-minute video, a 15-minute podcast or a 30-minute e-learning module.

This trial measured the impact of different delivery formats on a 'training effectiveness index' constructed from measures of engagement, completion, learning comprehension, and whether the participant would recommend the training to others. Since click-through modules are common for APS online training, they were treated as the benchmark against which the video or podcast training were compared. The study also involved 2 focus groups with 13 participants.

Randomised trial results: The 3-minute video and 30-minute module training formats were similarly effective and both out-performed the podcast format. The training effectiveness score for the podcast group was 7.9% lower than for the click-through module. By contrast, for people in the video group, the effectiveness index was slightly higher than for the module, but this difference was not statistically significant.[1]

Survey results: A follow-up survey question showed strong support for the video format: 52% of respondents preferred a 'short explainer video', compared to 39% for an 'online click-through module' and 9% for a 'podcast format'.

Focus group results: The evaluation included 2 focus groups with 13 participants. Most focus group participants said they generally preferred a video to click-through modules or a podcast, consistent with the survey results. Additional insights from focus group participants included:

- Video participants reported that they found it fast-moving and engaging, and the video's visuals effectively reinforced the points being made by the voice over. They also appreciated that it was short. Another observation was that videos may also help neurodiverse people.

- Participants found the requirement to log in to the click-through module was a barrier to commencing (and hence completing) it.

- When participants listen to a podcast in a work environment it is much easier to get distracted by other work tasks than when completing training in a video or click-through module format.

---

1   This result was not the primary analysis specified in the pre-analysis plan. Instead, it was pre-specified as exploratory analysis. The primary analysis compared the video and podcast groups, combined, against the click-through module group, and found no apparent difference between these two groups. Nonetheless, the authors believe it is reasonable to treat the video versus module result as a 'confirmatory' result: see Appendix A for details.

Training Effectiveness Index Score by training format
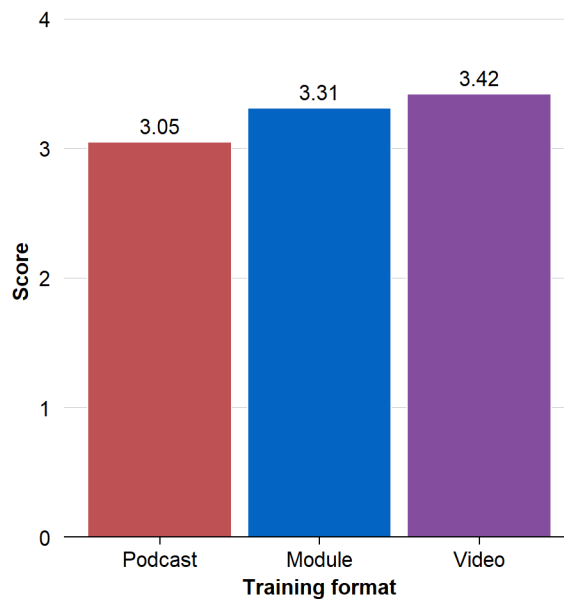
Note: The training effectiveness scores ranged from 0 to 5. This analysis was exploratory: see Results and Appendix A for more details. Sample size: click-through module=216, podcast=250 and video=286.

Limitations: The evaluation evidence had 4 main limitations. First, the trial focused on short, voluntary introductory training and so the findings may not generalise to other forms of training. Second, the randomised trial and the supplementary survey question both suffered from missing outcome data, which may have introduced non-response bias into the results. Third, the trial analysis deviated from the pre-analysis plan by treating a pre-specified exploratory analysis as 'confirmatory', however the authors believe this deviation was justified (see footnote 1 and Appendix A). Finally, the randomisation procedure did not produce an even balance between the 3 treatment arms as intended. This was most likely due to a bug in the survey that changed the assignment ratio approximately halfway through the trial. The authors accounted for the change in assignment ratio in their analysis and proceeded on the basis that participants were randomly allocated.

Conclusion: Notwithstanding these limitations, the randomised trial results suggested the video format is unlikely to be worse than the module format in terms of training effectiveness. Given the video format was preferred in the supplementary survey question and the focus groups, and was cheaper to produce, the ACE judged that the evidence was sufficient to guide an immediate decision about which training format to use for the online training courses it was developing. However, the study's limitations, and the context in which it was conducted, should be borne in mind before applying the results elsewhere.

Background and policy context

The Australian Public Service (APS) currently offers numerous online training courses covering a wide range of content, and delivered in a variety of formats. These courses are an important and widely-used tool for APS officers' professional development and learning.

In 2025, the Australian Centre for Evaluation (ACE) developed a series of short, introductory training activities on evaluation topics, and targeted at APS staff. The aim of the training was to give participants a brief overview of key evaluation concepts as a foundation for further learning, or to help engage with experts in that topic.

As part of the training development, the ACE wanted to determine whether a different *delivery format* could improve engagement with the training. Consequently, the ACE conducted a rapid evaluation, over January to March 2025, to test the effectiveness of different delivery formats. The 3 different formats tested were: a traditional click-through e-learning module, a micro-learning video and a podcast.

Despite the investment in these online training courses across the APS and their importance to professional development, the authors were not aware of any counterfactual-based evaluations of the effectiveness of different delivery formats for similar courses delivered to similar professional audiences.

All 3 courses that were tested as part of the evaluation covered the same topic – the role of randomisation in evaluation. However, as discussed in more detail in the next section, the courses varied in several ways:

- Medium – module, video or podcast

- Duration – about 3 minutes for the video, 15 minutes for the podcast, and 30 minutes for the module, and

- Sign-in requirements – necessary for the module only.

These differences reflected the specific formats relevant to the training developed by the ACE, but other organisations may have different needs or choices available (for example, in relation to sign-in requirements or course duration). Furthermore, this evaluation studied delivery formats for a short, introductory and voluntary course. These considerations may affect the generalisability of these results to other forms of online training.

Evaluation design and implementation

This evaluation used a randomised controlled trial (RCT) and focus groups to test the effectiveness of different online learning formats for APS staff.

Key evaluation question: Can micro learning videos or podcasts improve training effectiveness compared to traditional online click through training modules?

Trial design and analysis

The evaluation included a 3-arm randomised controlled trial with the following groups:

- **Control Group:** Click-through training module (approximately 30 minutes)

- **Treatment Group 1:** Podcast (approximately 15 minutes)

- **Treatment Group 2:** Micro-learning video (approximately 3 minutes)

While the format of each training was different for each group, all trainings taught participants about the same topic: How randomised controlled trials (RCTs) work, and what they are useful for. The click-through module was a course titled 'Randomisation' (APS Academy 2024). The video was 'Why Randomize?', developed by the Poverty Action Lab (J-PAL) (J-PAL 2024). The podcast was titled 'Impact Evaluation Podcast', developed by the Australian Centre for Evaluation (ACE 2025a).

Figure 1: Images from click-through module and micro-learning video
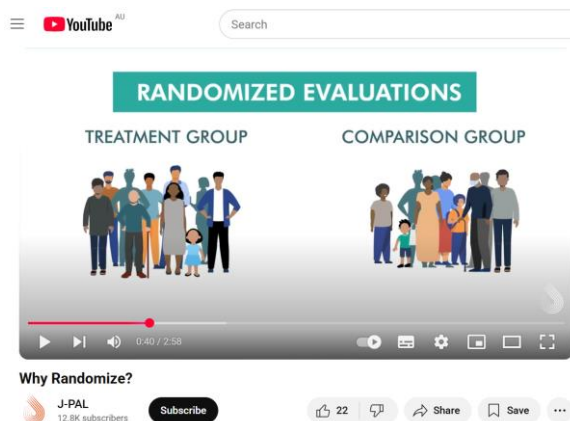
Trial video source: J-PAL (2024)   Trial module source: APS Academy (2024)
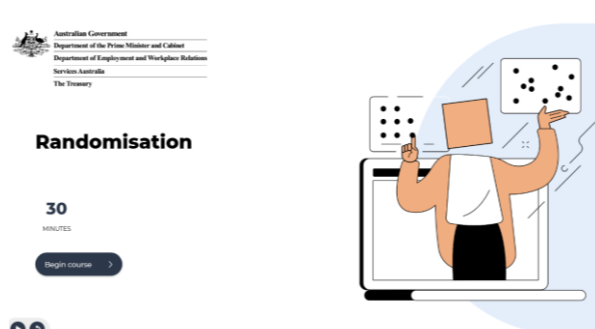
The trial's primary outcome measure was a training 'effectiveness index' which was a score out of 5, constructed from the following questions.

- *Participants' engagement rating*: 'To what extent do you agree with the following statement: The training content was engaging'. This was measured using a Likert scale (strongly disagree, disagree, neither agree nor disagree, agree, strongly agree).

- A *'net promoter score'*: 'How likely would you be to recommend this training to a friend/colleague?'. This was initially measured on a 0 to10 scale but was converted into a binary variable: scores of 7 to 10 ('promoter') were assigned a score of one, others ('detractors') were assigned a score of zero.
- *Self-reported training completion*: 'How much of the training module/video/podcast did you complete?'. Response options: little or none (0%–24%), some (25%–74%), most (75%–99%) and all of it (100%).
- *Learning outcome 1 (control group)*: A multiple-choice question designed to test understanding of the concept of a control group.
- *Learning outcome 2 (randomisation)*: A multiple-choice question designed to test understanding of the role of randomisation in causal inference.

Treatment effects were calculated using ordinary least squares regression model with robust standard errors. Various robustness checks were undertaken. For further detail on the trial design and analysis, including the definition of outcome variables, see Appendix A. For further detail on the results of the robustness checks, see Appendix C.

Ethics approval for this study was granted by Macquarie University's Human Research Ethics Committee (Project ID: 18683). The trial was pre-registered on the American Economic Association's Social Science Registry (AEARCTR-0015387) on 14 February 2025 (ACE 2025b). This registration was completed after the trial was launched but prior to data collection and analysis.

Trial sample and randomisation

Public servants from a range of different federal government departments were invited, in January and February 2025, to take part in the trial by email using the APSC's existing email distribution lists. The email encouraged public servants to take part in the training to help provide evidence about the effectiveness of different online training formats. The email also mentioned that participants would, in appreciation for their participation, be entered into a prize draw to win one of 5 gift cards valued at $50 each.

The email included a link to a Qualtrics survey. After starting the survey, participants were randomly assigned to one of the 3 training formats using the 'Randomizer' functionality of the Qualtrics survey software. They were asked to complete the training and then return to the survey to answer the 5 questions listed in the previous section.

While 805 participants were randomised into one of the 3 treatment arms, there were 53 duplicate responses, where a participant commenced the training survey more than once. These duplicates were removed, leaving a total of 752 unique participants who received the assigned training. The assignment ratio was meant to be evenly balanced (that is, 1:1:1) however the final allocation was unbalanced: 216 (29%) received the click-through module, 250 (33%) received the podcast and 286 (38%) received the video. This was most likely due to a bug in the survey software that changed the assignment ratio approximately halfway through the trial. The authors accounted for the suspected change in assignment ratio using a covariate and proceeded on the basis that participants were randomly allocated (albeit not in an even 1:1:1 ratio). For further discussion, see Appendix A.

Outcomes were collected through a follow-up survey, meaning there was missing outcome data for those who did not respond. Overall, 173 participants (or 23% of the total sample of 752) did not answer any of the survey questions. This was distributed unevenly between the 3 groups, with the following non-response rates for each group: 30% for the click-through module, 27% for the podcast, and 15% for the video. This differential attrition mirrored differences in completion rates: 56% for the

click-through module, 52% for the podcast, and 74% for the video. (Some participants answered the survey even though they did not complete the training itself.) For a full CONSORT diagram and discussion of differential attrition, see Appendix A.

Missing values were imputed through the following 2 steps, as set out in the pre-analysis plan. First, the 'completion' value was assumed to be zero (that is, if the person hadn't completed the survey, it was assumed that they had not completed the training either). Second, the remaining components were imputed using the 'grand mean' for that component (that is, the mean across all 3 groups). A robustness check tested how this imputation affected the results: see Appendix C for details.

Participants from some government departments could not access the video or podcast due to IT security restrictions that limited their access to external links. This only affected a small number of participants, but they were still included in the analysis using the conventional 'intent-to-treat' approach, and missing values were imputed following the steps described above.

Focus groups

The evaluation included 2 focus groups with 13 trial participants. The focus groups aimed to generate further insight into participants' experience and format preferences. The discussions followed a discussion guide centred around the following 5 themes:

1. Experience of accessing e-Learning

2. Engagement with, and understanding of learning content

3. Challenges with the learning format

4. Comparison of video learning format to click-through module format

5. Suggestions for improvements

For the complete discussion guide, see Appendix E.

At the start of the trial, survey participants were asked to opt-in to a respondent pool for further research. Of these, 18 were randomly selected to be invited to attend a focus group discussion. Of these, 13 participants accepted the invitation and attended one of the 2 focus groups. Text transcripts of the focus group discussions were generated using Microsoft Teams. The research team had clearly defined research questions and specific gaps arising from the quantitative analysis, so the transcripts were analysed with an 'a priori thematic analysis' approach. Specifically, the 'Miles-Huberman matrix' was used to organise and analyse the focus group data (Miles et al. 2014).

Randomised trial results

This section describes the overall trial results from the 'training effectiveness index' and then unpacks these results for each of the index's individual components. Next, it presents the results of a supplementary survey question that asked all trial participants about which delivery format they preferred. The section concludes with a discussion of missing data and robustness checks. (For more details on these results and robustness checks, see Appendices B and C.)

The video format had the highest 'effectiveness score'

This trial measured the impact of different delivery formats on a training effectiveness index constructed from measures of engagement, completion, learning comprehension, and whether the participant would recommend the training to others. (See previous section for details.)

The pre-registered primary analysis for this trial compared the effectiveness score of the click-through module with the score for the *combined* sample of podcast and video users. There was no statistically significant difference in the scores for these 2 groups.

To better understand this result, the authors conducted 2 further comparisons: click-through module versus video and click-through module versus podcast. When the video and podcast effects were estimated individually, it turned out they had roughly opposing effects (Figure 2). The index score for the podcast format was 7.9% lower than for the click-through module. The score for the video format was 3.3% higher than for the module however this difference was small and not statistically significant.[2]

Figure 2: Training Effectiveness Index Score by training format
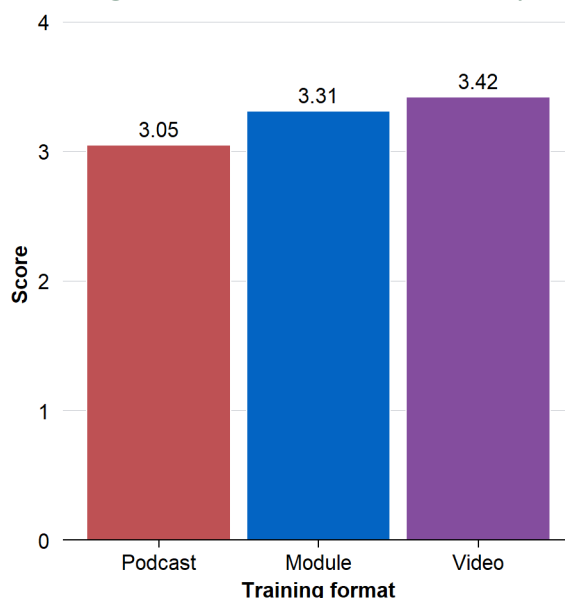
Note: This analysis was exploratory, and treatment effects were covariate adjusted. N: click-through module=215, podcast=251, video=286. The difference between the module and podcast was statistically significant, while the difference between the module and video was not. Survey completion rates varied: of the module recipients, 30% did not answer any of the survey questions, compared to 27% and 15% for the podcast and video formats, respectively. The training effectiveness scores ranged from 0 to 5. See Appendix A for technical details, and Appendix B for more detailed results.

---

2   Because the primary analysis did not produce a statistically significant result, treating this subsequent analysis as 'confirmatory' is a deviation from the pre-analysis plan. However, the authors' assessment is that this deviation is justified. See Appendix A for further discussion.

Strong video performance driven by high completions

Exploratory analysis suggested the video training performed better than the click-through module on the self-reported completion rate (Figure 3). The completion rate for those assigned to the video training was 74%, as compared to 56% for the click-through module. Although this was based on exploratory analysis, the difference is so large it is unlikely to be due to chance alone. No doubt this was partly due to their differing lengths: the video was 3 minutes whereas the module was closer to 30 minutes. Focus group participants also shared that they found it easy to stay attentive to the video due to the visual cues it provided.

There was no clear evidence of a meaningful difference in the other measures of effectiveness (even though it appeared that the module had a somewhat higher 'net promoter score' than the video). See Appendix B for further discussion on these results.

Figure 3: Effectiveness index components by training format
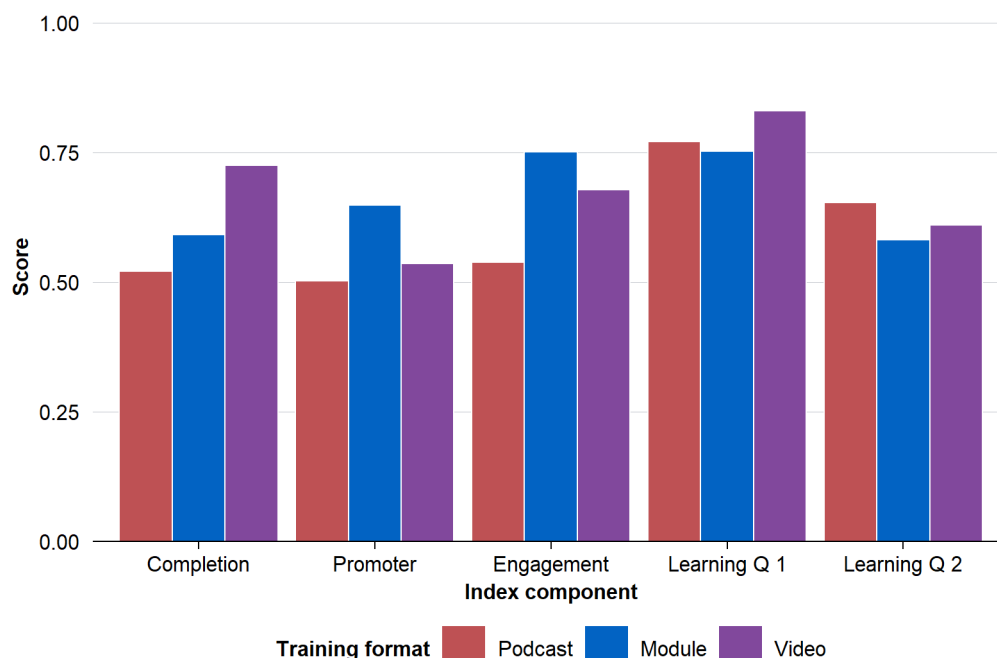
Note: The results are based on exploratory analysis and all treatment effects are covariate adjusted. Nonetheless, they provide further insights about the different index component outcomes for each of the training formats. See Appendix A for technical details.

Poor podcast performance driven by 'promoter score' and engagement

The podcast training performed worse than the click-through module because it had a lower engagement and 'net promoter score'. There was no clear evidence of a meaningful difference between the podcast and the module in the other measures of effectiveness (although, when compared to the *video*, the podcast had a much lower completion rate).

The podcast format was inferior to both other formats on engagement: only 54% of podcast participants agreed the training was engaging, compared to 71% for the video and 75% for click-through module. Focus group participants shared that the podcast was particularly unsuitable in a work environment because they listened to the podcast while at their desk and so were easily distracted by other work. However, they did acknowledge that the language used in the podcast was easy to understand.

The podcast also performed worse on the 'net promoter score', which asked: 'How likely would you be to recommend this training to a friend/colleague?' (0 to 10 score). People gave a score of 7 to 10 were classified as a 'promoter'. While 65% of participants assigned to the click-through module were 'promoters', this fell to 51% for podcast recipients. For further detail, see Appendix B.

Supplementary question confirmed video preference

After participants had completed the training and answered the survey questions related to the training, a final, optional question asked what format they thought was best to deliver the online training. Responses aligned with the trial and focus group results with 52% preferring a 'short explainer video', 39% an 'online click-through module', and just 9% preferring a 'podcast format'.

Figure 4: What format do you think is best to deliver this training?
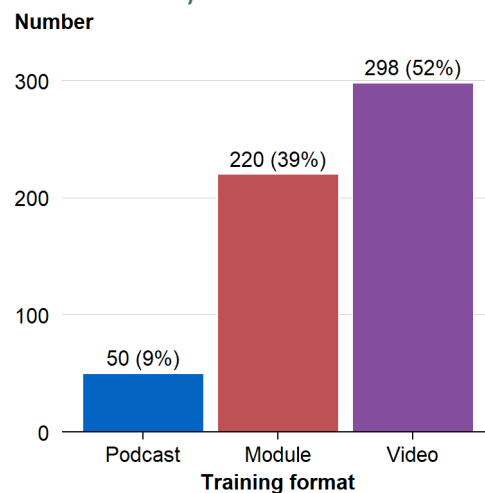
Note: Within the sample of 752 participations, 568 responded (76% response rate).

Video and podcast participants both preferred the video over the other 2 options however people who were assigned the click-through module preferred the module. More generally, participants were more likely to say they preferred the training format that they had just completed (Figure 5).

Because this supplementary question was optional, there were 184 missing responses. These missing responses were unevenly spread across the 3 groups, with fewer missing responses from the video group: 71 podcast, 66 module and 47 video. While the stronger response rates from video participants may have partly skewed the results in favour of video, this is unlikely to fully explain the substantial differences shown in Figure 4.

Figure 5: Preferred format, by treatment arm
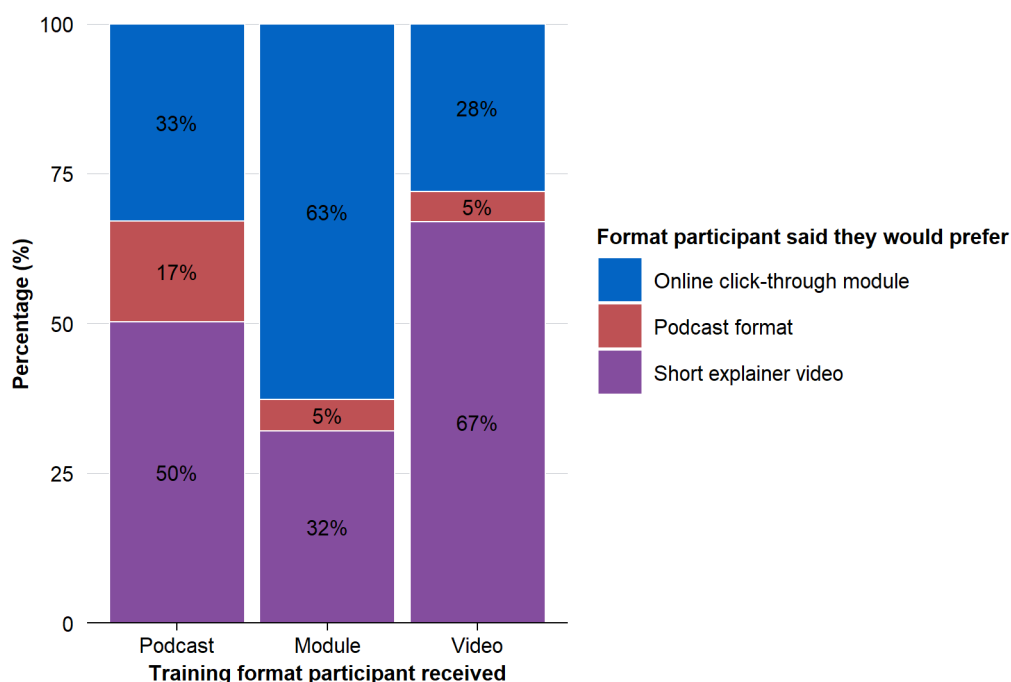
Note: Within the sample of 752 participations, 568 responded (76% response rate).

Missing data and robustness checks

Some participants (173 or 23%) did not answer any of the survey questions that were used to collect outcome data. As discussed in the Methods section, there was differential attrition in missing outcomes between groups: more were missing from the module and podcast groups than the video group. For the main analysis, the missing values were imputed, consistent with the pre-analysis plan.

An initial robustness check tested whether the results changed if the participants with missing data were removed from the analysis. The finding that the podcast was inferior remained, with the treatment effect becoming slightly more negative. Similarly, the video versus module result remained small and not statistically significant.

Missing survey data was closely correlated with training completion (that is, the video had the highest training completion rate and the highest survey completion rate). Consequently, a second robustness check tested whether the training completion component of the effectiveness index was the main driver of the overall result. Again, both results were unaffected.

See Appendix A for further discussion of missing data, and Appendix C for the results of the robustness checks.

Focus group results

The research team held 2 focus groups with a mix of 13 participants from each of the 3 trial arms. The full discussion guide can be found in Appendix E.

The role of video

Most focus group participants said they generally preferred videos to click-through modules due to their short duration, and the succinct and focused message. Participants were also more inclined to say that the visuals effectively illustrated the point being made. Another observation was that videos may also help neurodiverse people because they are visually more engaging, and often have captions or transcripts.

Focus group participants outlined that, in their view, an effective and engaging micro-learning video should include:

- A short length and focused topic

- Engaging graphics/animation that illustrate the point being made in the audio

- Speech with human voices rather than voices that sound computer generated

- Real world examples

- Transcripts and video captions (also good for neurodivergent viewers)

- A one-pager that summarises the video's content for future reference

- A content summary so viewers know why to watch it at the start

The role of click-through modules

There was mixed feedback in the focus groups on whether the click-through module was engaging. Some people who received the module found it engaging because it used both visual and audio features, while others found it less engaging because the visuals were not well integrated into the learning. Some people shared that they got in the habit of just clicking to get through the module, which decreased their concentration on the actual content.

People also shared that a contributing factor to poor engagement with the click-through module was that it required either setting up or remembering their login details. Further, it required additional clicks to navigate. For example, once the participant opened the Randomisation module, they had to click 5 more times – through introduction pages and 'how to navigate' pages – before viewing the actual content. In some people's view, this decreased the user experience.

The role of podcasts

Focus groups largely agreed with the result from the randomised trial that podcasts were less engaging than the module. Some focus group participants explained that although the language in the podcast was easy to understand, it was hard to stay engaged because there was nothing to look at. Participants received the link to the podcast via their work email and hence listened to it on their work computer. This meant some participants ended up getting distracted by emails or other work tasks. Few participants went to the trouble of sending the link to their personal device so they could listen at other times.

Some participants in the focus group suggested podcasts work better as a format for topics that the listener already has an interest in and some background knowledge as this can make them easier to understand during a commute. People also shared that a podcast can be better for a topic one listens to for leisure as there is no requirement to absorb every detail.

Limitations

The ACE conducted a rapid evaluation to inform an immediate decision about which training format to use for future online training courses it was developing. Although this evaluation had several limitations, the ACE judged the evidence was sufficient to guide that decision. This section outlines the 4 main limitations of the evaluation. It also discusses limits on the generalisability of the results.

First, the randomised trial and the supplementary survey question both suffered from missing outcome data for 23% of the study population, which may have introduced non-response bias into the trial estimates and the response to the survey question (see Appendix A for details).

Second, the trial analysis deviated from the pre analysis plan by treating exploratory (but pre-specified) analysis as confirmatory. However, as discussed in Appendix A, the authors believe this deviation was justified.

Third, the randomisation procedure did not produce an even allocation across the 3 treatment arms. This was most likely due to a bug in the survey software that changed the assignment ratio approximately halfway through the trial. The authors accounted for the change in assignment ratio in our analysis using the usual approach for analysing data with a change in assignment ratio. (That is, the authors added a 'dummy variable' that identified survey responses submitted after the change in assignment ratio from those submitted before it). The analysis then proceeded on the basis that participants were randomly allocated. This was a deviation from the pre analysis plan and is detailed in Appendix A.

Finally, the findings might not generalise to other online trainings. The study compared 3 different forms of short, introductory, voluntary training intended for Australian public servants on a technical topic: how to conduct randomised trials. Furthermore, the study varied several features of the training at once: training format (module, video or podcast), length of training (ranging from 3 to 30 minutes) and presence of sign-in requirements (only the module has sign-on requirements). For these reasons, care should be taken when generalising these findings beyond this setting.

Conclusion

Despite the limitations outlined above, the combined evidence from the trial, the supplementary survey question and the focus groups was sufficient to support the ACE's immediate decision to use a video format for future short, introductory online training courses. An additional consideration was that the videos were likely to be cheaper to produce.

Given the video format was preferred in the supplementary survey question and the focus groups, and was cheaper to produce, the ACE judged that the evidence was sufficient to guide an immediate decision about which training format to use for the online training courses it was developing.

Further, the focus groups suggested that the video was preferred due to a combination of the medium (video), duration (about 3 minutes) and the lack of a sign-in requirement. These different features reflected the specific formats that were relevant to the ACE, but other organisations may have different needs or choices available (for example, in relation to sign-in requirements or course duration).

The evaluation results were also consistent in finding that the podcast was an inferior format. Focus groups suggested podcasts may be better suited to settings where the learner can listen while multi-tasking (for example, driving or exercising) rather than when they are at work. Podcasts may also be better for communicating information that the listener is already very interested in, or for learning about concepts where they do not need to absorb every detail.

A final consideration is the ease of access to voluntary training. Users had to login to access the click-through module, meaning they had to remember their login details, or create an account. Once they had logged in and started the module, users had to click through several pages of introductory text and explainers on how to use the module before they came to the actual training content. This likely contributed to module participants being less likely to complete the training (although the duration of the module, about 30 minutes, was undoubtedly relevant too). Requiring participants to login can be valuable for various reasons, such as tracking course progress, providing completion certificates, or sending follow-up surveys. Nonetheless, there may be a case for making some non-compulsory courses available without the need to login, if further evidence suggests that this will improve take-up or completion rates.

Appendix A: Technical details

This appendix provides further detail on how the randomised trial was conducted and analysed, in accordance with the pre-analysis plan. It details:

- pre-registration and ethics
- deviations from the pre-analysis plan
- a CONSORT diagram, showing the flow of participants through each stage of the trial
- the study population
- randomisation
- definition of outcome variables
- the method of analysis
- the risk of bias

Pre-registration and ethics

This evaluation was reviewed and approved by one of Macquarie University's Human Research Ethics Committees (Project ID: 18683) on 13 January 2025.

The trial was publicly pre-registered with the American Economic Association's Social Science Registry (AEARCTR-0015387) on 14 February 2025 (ACE 2025b). The pre-analysis plan (PAP) and registration were completed after the trial was launched but prior to data collection and analysis.

Deviations from pre-analysis plan

There were 2 deviations from the pre-analysis plan.

The first was to treat exploratory (but pre-specified) analysis as confirmatory. The primary analysis specified a pooled comparison of video and podcast formats against the click-through module, on the expectation that both would prove superior (or at least not inferior), and the pooled analysis would have greater statistical power. However, there was no statistically significant difference between the combined video/podcast group and the module.

Subsequent exploratory analysis of the effects of the video or podcast, separately, versus the click-through module, suggested that the authors' initial expectation was mistaken: only the video had a positive effect on the effectiveness index (although not statistically significant), while the podcast was inferior. Since the initial choice to base the primary analysis on the pooled comparison appears to have been based on an error of judgement, and the exploratory analysis was pre-specified (and a natural corollary of the primary analysis), the authors felt it was reasonable to treat this analysis as confirmatory.

The second deviation was to add a 'post assignment ratio change' mean-centred dummy variable that was interacted with treatment. This variable was included in the analysis to account for a suspected change in the assignment ratio where far fewer participants were allocated to the click-through module in the second half of the survey than would likely be possible due to randomisation. Full details on the suspected change in assignment ratio are detailed in the 'Randomisation and balance checks' section below.

CONSORT diagram

Participants were recruited into the randomised trial through emails that were sent to APS employees on distribution lists. From these emails:

- **858 potential participants** clicked on the training link included in the email

- **53 potential participants** did not consent to participate or provide an email address

- This left **805 participants** who were randomly allocated to one of the training formats.

Table 1 shows the allocation of the 805 participants to the 3 treatment arms (module, podcast or video).

Table 1: CONSORT diagram: random allocation to treatment arms (n=805)

	Module	Podcast	Video
Allocation (A)	234	267	304
Duplicate responses (B)	18	17	18
Unique participants (C = A − B)	216	250	286
Training: Completed 100%	121	131	213
Training: Completed some or none	95	119	73
Survey: Answered all 5 questions	151	178	241
Survey: Did not answer survey	64	67	42
Survey: Answered first 3 questions but not the 2 learning comprehension questions	1	5	3

Study population

The study population was recruited using emails sent to several of the APSC's APS-wide distribution lists. This meant that the email was received by APS employees in over 60 different agencies. Some of these agencies included Services Australia, Prime Minister and Cabinet, the National Disability Insurance Agency, Department of Industry Science and Resources, Department of Health, Disability and Aged Care, Department of Social Services, Treasury, Defence, the Australian Taxation Office, APSC and the Australian Bureau of Statistics.

Randomisation and balance checks

Participants were randomly assigned to one of the 3 training formats using the 'Randomizer' functionality of the Qualtrics survey software. While 805 participants were randomised into one of the 3 treatment arms, there were 53 duplicate responses, where a participant commenced the training survey more than once. These duplicate responses were dealt with as follows:

1.   If both responses were allocated to the same treatment arm *and* the participant had not completed the survey in the first response *and* had done so in their duplicate response, then the first response was deleted.

2.   Otherwise, the duplicate response was deleted.

This left a total of 752 unique participants who received the assigned training.

The assignment ratio was meant to be evenly balanced (that is, 1:1:1) however the final allocation was unbalanced: 216 (29%) received the click-through module, 250 (33%) received the podcast and 286 (38%) received the video. Such an imbalance is very unlikely to have occurred by chance.

This imbalance was not due to the removal of duplicates, which were fairly evenly allocated between the 3 treatment arms (see CONSORT diagram). By analysing when training commencements occurred, the research team discovered that for the first half of the study period, the assignment ratio was approximately evenly balanced. However, around the 430th survey response (out of a total of 805 responses), the likelihood of being randomised to the 'module' arm decreased noticeably.

The research team investigated the cause of the change with the survey provider's (Qualtrics) technical support team. The authors had modified the survey slightly around the 430th response in ways that were unrelated to the random assignment ratio (for example, an additional question was added to the survey although this was not subsequently used in this report). The survey provider's technical support team could not provide a definitive answer on how or why the allocation ratio changed. However, they noted that it is not recommended to change the survey after it has gone live and that it was possible this introduced a bug into the software that affected the assignment ratio.

One way to diagnose the risk of a randomisation failure is to conduct balance checks. The survey did not ask for information on demographic characteristics or agency, so it was not possible to conduct a balance check on these sorts of variables. However, the survey did ask whether participants would be willing to be contacted for further research opportunities. This variable seemed to be broadly balanced across treatment assignment (Table 2).

Table 2: Balance check – characteristic composition by treatment group

Characteristic	Module	Podcast	Video
Willing to be contacted for opportunities to participate in further research	66.7%	62.5%	64.1%

Note: Balance check conducted on the randomised data before duplicates were removed, sample = 805.

Notwithstanding this balance check, overall, the authors were concerned the unusual random assignment ratio could have introduced bias into the study. As a result, they included a mean-centred dummy variable for post-assignment ratio change that was interacted with treatment. They then proceeded on the basis that participants were randomly allocated (albeit possibly not in an even 1:1:1 ratio). This is the usual procedure for dealing with a change in assignment ratio in a randomised trial if a change in assignment ratio was intentionally implemented by the researchers. Full details on the model specification can be found under the 'Method of analysis' section.

Outcome variables

The table below outlines how each of the survey response options were coded for analysis.

Dependent Variable	Survey question and response options	Variable coding
Training effectiveness index	Combination of dependent variables to form an index (engagement rating, net promoter score, training completion, and learning comprehension tests)	Continuous variable from 0 to 5. Each of the secondary outcome variables described below were converted to a binary measure [0,1] and summed together to create a score out of 5.
Self-reported training completion*	How much of the training module/video/podcast did you complete? a. Little or none (0%–24%) b. Some (25%–74%) c. Most (75%–99%) d. All of it (100%)	Binary variable, where {a, b, c} = 0 and d = 1
Net Promotor Score	How likely would you be to recommend this training to a friend/colleague? (1–10 scale)	Binary variable, where {1–6} = 0 and {7–10} = 1
Engagement rating	To what extent do you agree with the following statement: "The training content was engaging." (5-point scale from Strongly Disagree to Strongly Agree)	Binary variable, where {a, b, c} = 0 and {d, e} = 1
Learning outcome 1	Consider this scenario: The Australian Government is studying a new approach to language learning through flash cards for primary school children. The government wants evidence to decide whether this is a good approach. Children from the same suburb, with similar NAPLAN scores, and demographics are randomly assigned to 2 groups. Group A receives the flash cards, while group B doesn't. In this scenario, what is the control or comparison group? Options: a. NAPLAN. b. Flash cards. c. Group A. d. Group B. [(d) is correct.]	The 2 learning outcomes were scored as correct (1) or incorrect (0) and were then summed to create a score from 0 to 2.
Learning outcome 2	Which **ONE** of the following correctly completes the sentence below? Randomising between the treatment and control groups helps find the true causal effect of a program because it ensures a. Participants in the trial are representative of the overall population of interest. b. People in the treatment group are not systematically different to people in the control group [(b) is correct.] c. The sample size of the study is large enough.	See above.
Preferred format	What format do you think is best to deliver this training? a. Short explainer video b. Online click-through module c. Podcast format	Not applicable

* Trial participants who were randomised but who left at the first page of the survey, without commencing the training, had their training completion recorded as 'little or none'.

Method of analysis

A conventional Intent-to-Treat approach was used to estimate the effect of being assigned to one of the treatment groups (micro-learning video or podcast) on the primary outcome variable: *Training Effectiveness Index*. The ITT approach was particularly relevant because a small number of participants from some government departments could not access the video or podcast due to IT security restrictions that limited their access to external links. Consistent with ITT, these participants were still included in the analysis, and missing values were imputed.

The pre-specified threshold for statistical significance was $p < 0.1$. This is a less stringent test than is conventionally used but was judged appropriate in this context given the consequence of a false positive is not severe and since the authors wanted to avoid a large 'false-negative' error rate. All tests for statistical significance were 2-sided.

The estimate, confidence intervals and p-values were derived from an ordinary least squares (OLS) regression model using robust (heteroskedasticity consistent type 2, HC2) standard errors with the following specification:

$$Y_i = \beta_0 + \beta_1 Treatment_i + \beta_2 \text{Ratio Change}_i + \beta_3 \text{Treatment} * \text{Ratio Change}_i + \in_i$$

Where:

- i is an index for each individual in the trial

- Y is the Training Effectiveness

- β_0 is the intercept

- $Treatment$ is a binary assignment indicator (where {0} = online training module; {1} = podcast format or micro-learning video)

- β_1 is a coefficient representing the average treatment effect when 'Ratio Change' is at its mean value.

- $Ratio\ Change$ is a binary indicator to indicate whether the observation is before (0) or after (1) the suspected assignment ratio change which was then mean centred

- $Treatment * Ratio\ Change$ is an interaction term that captures whether the effect of treatment differs after the mean centred post assignment ratio change indicator

- $\in$ is the standard error term.

Risk of bias

While the aim of randomisation is to eliminate the risk of bias in estimated outcomes, bias may re-emerge in various ways. This section discusses 3 potential threats to internal validity: randomisation failure, missing outcome data, evaluation-driven effects, or spillovers.

Randomisation failure is a possibility, as discussed in the section above on 'Randomisation and Balance Checks'. However, as described above, the authors accounted for the apparent change in assignment ratio during analysis with a covariate 'Ratio Change', and then proceeded on the basis that participants were randomly allocated (albeit possibly not in an even 1:1:1 ratio).

Missing outcome data introduced a risk of selection bias into this trial, as discussed in the Methods and Results sections. Of the 752 trial participants, 168 participants started the survey but did not complete any of the questions and a further 9 participants completed the first 3 questions only but not the 2 learning comprehension questions (see CONSORT diagram in Figure 6). To deal with the missingness, the research team imputed a 'completion' score of 0 and then imputed the other scores using the mean value from all non-missing respondents. This was consistent with the pre-analysis plan. A robustness check tested how the results changed when missing values were dropped rather than imputed (see Results section and Appendix C for details).

Evaluation-driven effects were unlikely to be an important source of bias. This study implemented a pragmatic trial where participants were fully aware of the trial and not blinded to treatment. Therefore, it is possible that knowledge of the other training format options could have changed their behaviour (that is, how they approached the training) and their subsequent survey responses. However, the research team's judgement is that the risk of such behavioural effects, driven by awareness of being in a trial, is small.

Finally, **spillover effects** were unlikely to cause bias in this trial given the outcomes largely relate to the participants' experience of the training course to which they were assigned. In particular, the Qualtrics survey questionnaire platform ensured that a participant could only get access to a different treatment arm if they re-entered the survey (and subsequent training) from a different device, with a different IP address. In those cases, they still had to enter their email address (for follow-up reminders), which allowed the authors to identify 'duplicate' respondents. See Methods section for further details.

Appendix B: Detailed results

This appendix section provides the detailed results of the primary and exploratory analysis.

Primary analysis

The primary analysis for this trial compared the effectiveness score of the click-through module (n=216) with the score for the podcast and video *combined* (n=536) and accounted for the change in the random assignment ratio, as described in Appendix A. The difference between these 2 groups was small and was not statistically significant (Table 3).

Table 3: Effect of training format on the Training Effectiveness Index Score (0–5)

	Grouped (video and podcast)
Marginal mean for click-through module	3.31
Treatment effect	-0.059
P-value	0.515
Total sample size (n)	752

Note: * = p < 0.1. The treatment effect is covariate adjusted. The p-value is 2-sided and calculated using HC2 robust standard errors. For technical details refer to Appendix A.

Exploratory analysis

A pre-specified, exploratory analysis separately tested the difference between the video and the click-through module, and between the podcast and the click-through module. The video format had a somewhat higher index score by 0.109 points (3.3%) relative to the click-through module but this difference was not statistically significant (p-value > 0.1). By contrast, the podcast format decreased the index score by 0.263 points (7.9%) and was statistically significant (Table 4).

Table 4: Effect of training format on the Training Effectiveness Index Score (0–5)

	Grouped (video and podcast)	Video	Podcast
Marginal mean for click-through module	3.31	3.31	3.31
Treatment effect	-0.059	0.109	-0.263*
P-value	0.515	0.280	0.013
Total sample size (n)	752	502	466

Note: * = p < 0.1. Treatment effects are covariate adjusted. The p-value is 2-sided and calculated using HC2 robust standard errors. For technical details refer to Appendix A.

The click-through module was used as the benchmark for the tests in Tables 1 and 2 however the exploratory analysis also included a direct test of the video against the podcast. The video had a training effectiveness index that was 12% (or 0.372 index points) higher than the podcast (Table 5). This test was not pre-specified.

Table 5: Pairwise comparison of video to podcast format

	Grouped (video and podcast)
Marginal mean for podcast	3.05
Treatment effect for video format	0.372*
P-value	<0.001
n	536

Note: * = p < 0.1. The treatment effect is covariate adjusted. The p-value is 2-sided and calculated using HC2 robust standard errors. For technical details refer to Appendix A.

Finally, the exploratory analysis estimated the same model specification as for the primary analysis, but the outcome variable was replaced with the 4 individual components of the training index: training completion, Net Promoter Score, engagement rating, and learning comprehension. The video had a higher score (that was both substantial and statistically significant) than the module for the 'completion' component (Table 6). The podcast had lower scores (that were both substantial and statistically significant) than the module for the 'promoter' and 'engagement' components. No adjustments were applied for multiple testing however these results would still hold even with such an adjustment.

Table 6: Effect of format on individual components of the 'effectiveness index'

		Grouped (video and podcast)	Video	Podcast
Completion	Marginal mean for module	0.591	0.591	0.591
	Treatment effect	0.039	0.135*	-0.069
	P-value	0.275	<0.001	0.101
	n	752	502	466
Promoter	Marginal mean for module	0.648	0.648	0.648
	Treatment effect	-0.104*	-0.079	-0.134*
	P-value	0.025	0.117	0.014
	n	579	396	335
Engagement	Marginal mean for module	0.751	0.751	0.751
	Treatment effect	-0.119*	-0.045	-0.213*
	P-value	0.005	0.329	<0.001
	n	579	396	335
Comprehension	Marginal mean for module	1.33	1.33	1.33
	Treatment effect	0.098	0.110	0.083
	P-value	0.143	0.124	0.292
	n	570	392	329

Note: * = p < 0.1. Treatment effects are covariate adjusted. The p-value is 2-sided and calculated using HC2 robust standard errors. For technical details refer to Appendix A.

Appendix C: Robustness checks

This appendix reports on 2 robustness checks that were conducted after the main analysis was completed. These checks were not pre-specified but were intended to test the sensitivity of the main results to various analytical decisions.

1. The first check tests the sensitivity of the results to imputation. For the main analysis, each missing value was replaced with an imputed value – the mean value for that outcome amongst all non-missing respondents. This check tested whether removing participants with imputed (missing) values) changed the overall result.

2. The second check tests the sensitivity of the results to the 'completion rate' component of the training effectiveness index. The video was clearly the best performer on completion rate, so this check tested whether removing the completion rate changed the overall result.

Both results were robust to both checks.

Check 1: Sensitivity to imputation of missing outcomes

The first robustness check ran an alternative analysis on the subset of the respondents who completed every question. This tested whether the inclusion of 182 imputed responses (where 'completion' was assigned to be zero, and the mean value was assigned for each of the other questions), was the main driver of the overall result. (That is, the sample size fell from 752 participants in the main analysis, to 570 participants.) The podcast result remained practically and statistically significant (in the negative) and the video result remained not statistically significant (Table 7).

Table 7: Robustness check – complete survey responses only

		Grouped (Video and Podcast)	Video	Podcast
Original results with imputation for missing outcomes	Marginal mean for module	3.31	3.31	3.31
	Treatment effect	-0.059	0.109	-0.263*
	P-value	0.515	0.280	0.013
	n	752	502	466
Robustness check with missing outcomes removed	Marginal mean for module	3.52	3.52	3.52
	Treatment effect	-0.123	0.090	-0.300*
	P-value	0.328	0.476	0.036
	n	570	392	329

Note: * = $p < 0.1$. Treatment effects are covariate adjusted. The p-value is 2-sided and calculated using robust standard errors. For technical details refer to Appendix A.

Check 2: Sensitivity to completion rate

In the second robustness check, the research team created an alternative *Training Effectiveness Index* out of 4, which was the same as the primary index except it excluded the completion question. This tested whether the completion component of the index was the main driver of the overall result.

As for the first robustness check, the podcast format still had a material, negative and statistically significant effect on the index compared to the click-through module (Table 8). The video format's effect on the index – compared to the click-through module – remained not statistically significant and the point estimate became slightly negative.

Table 8: Robustness check – alternative index excluding training completion

		Grouped (Video and Podcast)	Video	Podcast
Original results, including completions (0-5 score)	Marginal mean for module	3.31	3.31	3.31
	Treatment effect	-0.059	0.109	-0.263*
	P-value	0.515	0.280	0.013
	n	752	502	466
Robustness check excluding completions (0-4 score)	Marginal mean for module	2.72	2.72	2.72
	Treatment effect	-0.099	-0.025	-0.193*
	P-value	0.201	0.770	0.028
	n	752	502	466

Note: * = $p < 0.1$. Treatment effects are covariate adjusted. The p-value is 2-sided and calculated using robust standard errors. For technical details refer to Appendix A.

Appendix D: Additional descriptive results

The research team conducted additional descriptive analysis to better understand the trial data on the completion, promoter score and engagement questions. Figure 7 illustrates the completion rate responses grouped by the training format received. A notable observation is that the podcast format had a much higher proportion (16%) of participants who completed 'little or none' of the training compared to the module (7%) or video (4%).

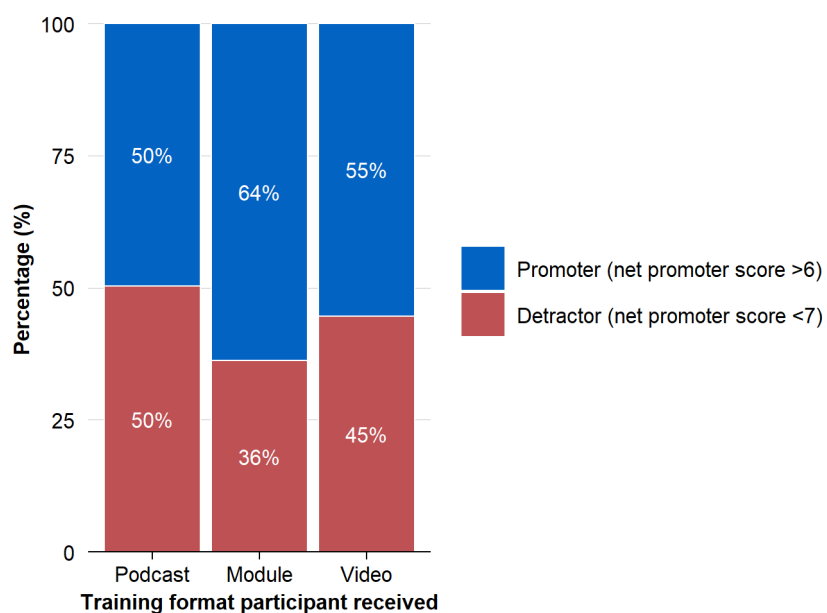Figure 7: Completion response grouped by training format received
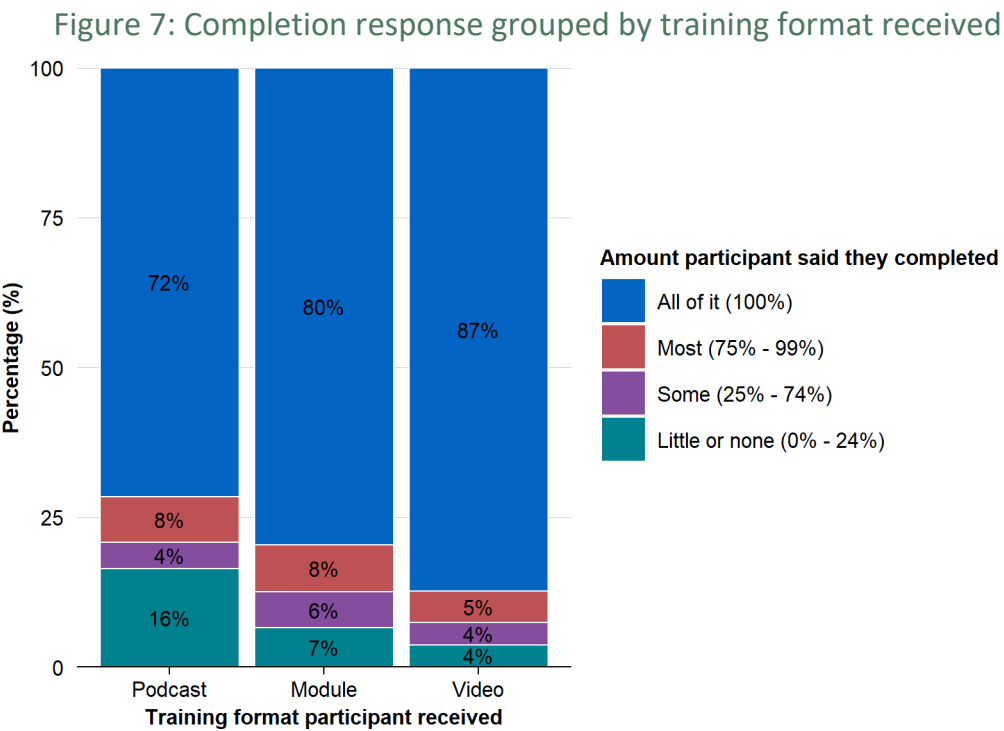
Note: Within the sample of 752 participations, 579 responded (77% response rate).

Figure 8 illustrates the spread of net promoter score ratings, grouped separately by the training format received. It shows consistently that the podcast was less likely to be recommended to a colleague or friend than the other formats.

Lastly, Figure 9 shows the spread of responses to the engagement question, grouped by the training format received. The engagement question asked, 'To what extent do you agree with the following statement: The training content was engaging'. The result shows that the negative effect of podcast format on engagement was mainly driven by less participants agreeing and more strongly disagreeing that the training's format was engaging.

Figure 8: Net promoter score grouped by training format received

Note: Within the sample of 752 participations, 579 responded (77% response rate).

Figure 9: Engagement response grouped by training format received
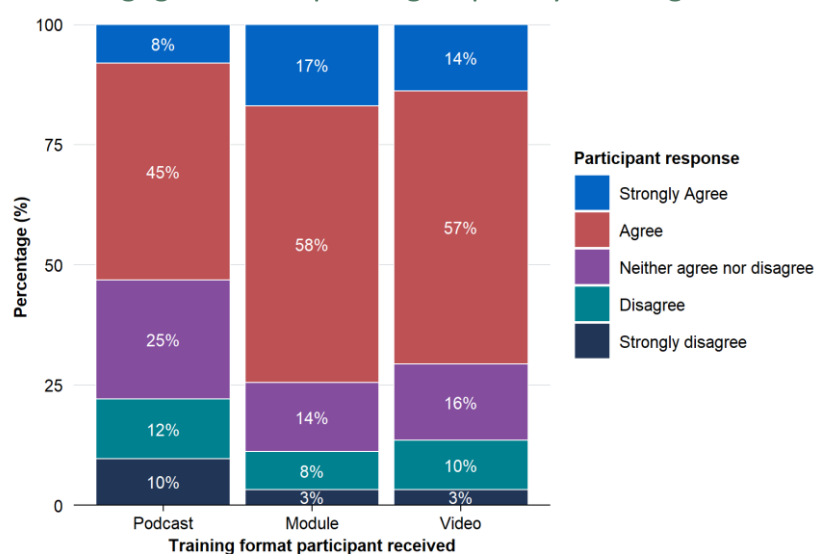
Note: Within the sample of 752 participations, 579 responded (77% response rate).

Appendix E: Focus group discussion guide

This discussion guide was used to prompt free-flowing discussions, and not all questions were asked at each focus group discussion. Follow-up or exploratory questions were asked when required to generate deeper insights.

Participants were provided with an overview of the evaluations methods and results which provided the rationale for discussion.

Focus Group Discussion Guide

1. **Experience of accessing e-Learning:**

    — How was your overall experience accessing and using the e-learning method you were assigned to?

    — Did you encounter any technical issues or barriers that affected your learning experience?

2. **Engagement with, and understanding of learning content from method you were assigned to:**

    — How engaging or interesting did you find the learning method you were assigned to (click-through training module, podcast, or micro-learning video)?

    — How easy or difficult was it to understand the content presented to you? Please provide examples.

3. **Challenges with the learning method assigned to:**

    — What were some aspects that you did not like for the learning method that you reviewed?

4. **Comparison of Video-Based Learning to Traditional Click-Through E-Learning:**

    — Based on your experience outside of this trial, how does video-based e-learning compare to traditional click-through e-learning modules in terms of engagement and effectiveness?

    — Do you think video-based learning offers any unique advantages or disadvantages compared to traditional (for example click-through e-learning)?

5. **Suggestions for Improvement:**

    — Based on your experience with micro-learning videos (within or outside of this trial), what are the main features of an effective and engaging micro-learning video?

References

ACE (Australian Centre for Evaluation) (2025a) *Impact Evaluation Podcast*, SoundCloud.

ACE (Australian Centre for Evaluation) (2025b) *What is the most effective format for delivering online training?*, AEA RCT Registry.

APS (Australian Public Service) Academy (2024) *Randomisation*, APS Academy.

J-PAL (Abdul Latif Jameel Poverty Action Lab) (2024) Why Randomize?, YouTube.

Miles M, Huberman M and Saldana J (2014) *Qualitative data analysis: a methods sourcebook*, 3rd edn, SAGE Publications, Arizona State University.