The impact of weighting by educational attainment and past vote on estimates of pre-election voting intentions: A case study using Australian polling data

D Pennay, S Misson and D Neiger

CSRM & SRC METHODS

PAPER NO. 2/2021

Series note

The ANU Centre for Social Research & Methods (CSRM) was established in 2015 to provide national leadership in the study of Australian society. CSRM has a strategic focus on:

- development of social research methods
- analysis of social issues and policy
- training in social science methods
- providing access to social scientific data.

CSRM publications are produced to enable widespread discussion and comment, and are available for free download from the CSRM website (https://csrm.cass.anu.edu.au/research/publications).

CSRM is located within the Research School of Social Sciences in the College of Arts & Social Sciences at the Australian National University (ANU). The centre is a joint initiative between the Social Research Centre and the ANU. Its expertise includes quantitative, qualitative and experimental research methodologies; public opinion and behaviour measurement; survey design; data collection and analysis; data archiving and management; and professional education in social research methods.

As with all CSRM publications, the views expressed in this Methods Paper are those of the authors and do not reflect any official CSRM position.

Professor Matthew Gray
Director, ANU Centre for Social Research & Methods
Research School of Social Sciences
College of Arts & Social Sciences
The Australian National University

October 2021

Darren Pennay is the Founder and former CEO of the Social Research Centre Pty Ltd. He is an Honorary Professor in the Practice of Survey Methodology at the ANU Centre for Social Research and Methods and an Adjunct Professor at the University of Queensland Institute for Social Science Research.

Sebastian Misson is a Senior Survey Statistician at the Social Research Centre.

Dr Dina Neiger is Chief Statistician at the Social Ressearch Centre and a Visitor at the ANU Centre for Social Research and Methods.

Abstract

The use of statistical adjustments ('weighting') to reduce bias in the estimates of voting intentions produced by pre-election opinion polls is an important issue for pollsters to consider. The weighting of pre-election polls was one of the main issues explored in the recent inquiry into the performance of the pre-election polls at the 2019 Australian federal election (initiated by the Association of Market and Social Research Organisations with the support of the Statistical Society of Australia).

This paper builds upon the work of that Inquiry and, in particular, looks at the efficacy of adding measures of educational attainment and past voting behaviour to the standard weighting solutions typically used by the commercial pollsters. The data used in this study are from a pre-election survey conducted by the ANU in April 2019.

We find that the amount of bias in estimates of voting intentions is routinely reduced, with only a minor impact on variance, when poststratification to educational attainment benchmarks is added to the standard weighting solution. We also find that poststratification to past vote benchmarks reduces bias when the measure of past vote that is used is one that is collected from respondents in close proximity to the previous election (a short-term recall measure of past vote). This reduction in bias is not observed when using a measure of past vote collected from respondents proximate to the next election (a long-term recall measure of past vote).

Working Paper No. 2/2021
An electronic publication downloaded from
<http://csrm.cass.anu.edu.au/research/publications

For a complete list of CSRM papers, see https://csrm.cass.anu.edu.au/research/publications.

ANU Centre for Social Research and Methods
Research School of Social Sciences
The Australian National University

Acknowledgments

The authors gratefully acknowledge the ANU Centre for Social Research and Methods for making their survey data available for this study and for their ongoing support of the social sciences by routinely lodging their survey data with the Australian Data Archive.

We would also like to thank the Social Research Centre for supporting this study and, in particular, Jack Barton for generating some of the initial weighting estimates and Columbia Winterton for her assistance in preparing this manuscript.

We are also grateful to Associate Professor Ben Edwards of the ANU Centre for Social Research and Methods for his helpful comments on a previous draft of this paper. A final note of thanks to the Life in Australia™ panellists for their ongoing participation in social science survey research of this nature.

The Social Research Centre is a wholly owned subsidiary of the ANU.

Acronyms

AAPOR	American Association for Public Opinion Research
ABS	Australian Bureau of Statistics
AEC	Australian Electoral Commission
AES	Australian Election Study
ALP	Australian Labor Party
ANU	Australian National University
CSES	Comparative Study of Election Systems
ERP	Estimated Residential Population
GRN	The Greens
LNP	Liberal-National Party coalition ('the Coalition')
PHON	Pauline Hanson's One Nation Party ('One Nation')
RMSE	Root mean square error
SRC	Social Research Centre
UAP	United Australia Party
UK	United Kingdom
US	United States
2PP	Two-party preferred

Contents

Tables

Figures

1 Introduction

It is generally agreed that the modern era of pre-election polling commenced in 1932 when George Gallup's first poll correctly predicted a local election in Iowa. A Gallup-affiliated company, Australian Public Opinion Polls, conducted the first Gallup poll in Australia in 1941 (Rhodes 2018). Pre-election polls have been the accepted means of estimating (predicting) election outcomes ever since. However, due to a spate of well-publicised polling misses around the world in recent years (Cornesse et al. 2020, 5–6), the accuracy of pre-election polling has increasingly been called into question with some critics having gone so far as to declare that 'polling is dead'.[1] Pre-election polling in Australia has not been immune from such criticism. According to one of the measures used by Pennay et al. (2020, 27) the performance of the Australian pre-election polls in 2019 was the least accurate since 1987.

These polling failures have led to comprehensive reviews of contemporary pre-election polling practices in the hope of identifying failings and improvements. Major reviews include the United Kingdom (UK) review of the 2015 general election (Sturgis et al. 2016) and the American Association for Public Opinion Research (AAPOR) Taskforce reviews of the 2020 and 2016 United States (US) Presidential elections (AAPOR 2017, 2021). The recent Australian contribution to this field, *The Inquiry into the Performance of the Opinion Polls at the 2019 Australian Federal Election* (Pennay et al. 2020), was instigated by the Association of Market and Social Research Organisations (now known as the Australian Data and Insights Association)[2] in conjunction with the Statistical Society of Australia.

Pennay et al. (2020, 73–74) found indirect evidence that the most likely reason for the failure of the Australian pre-election polls in 2019 was an uncorrected over-representation of more politically engaged and better educated voters and, as a result, an over-representation of Labor voters. Among the many recommendations arising from their report was that pollsters review their approach to sample balancing and/or weighting (2020, 76). Approaches to sample balancing/weighting that took into account the educational attainment and the most recent past vote of respondents were identified as *likely* to result in reduced bias (when bias is measured by comparing the vote choice estimates produced from a poll with the actual voting behaviour observed at the election). However, the evidence used by Pennay et al. (2020) in reaching this conclusion was indirect. They used data from two Australian *post*-election surveys, the 2019 Australian Election Study[3] and the 2019 Comparative Study of Election Systems[4] and relied on overseas studies of pre-election polling.

---

[1] A Google search of the phrase 'polling is dead' conducted on 30 May 2021 generated 12.3 million hits.

[2] https://dataandinsights.com.au/

[3] The AES is a post-election survey of a random sample of voters, conducted after each federal election since 1987. The data for this series is available from https://australianelectionstudy.org/.

[4] The Comparative Study of Electoral Systems survey (CSES Australia 2019) is a post-election survey of 2,000 Australian adults, conducted

In this paper, however, we are able to provide direct Australian evidence that incorporating educational attainment and/or past vote adjustments into the weighting solutions used for pre-election estimates of vote choice reduces the bias in these estimates. We are able to undertake this study because academics at the ANU Centre for Social Research and Methods made a dataset containing their pre-election measures of voting intentions publicly available via the Australian Data Archive.[5] It is hoped that, in a break from past practice, members of the newly formed Australian Polling Council follow this lead and in the future make unit recod data from their pre-election polls availabe for research purposes.[6]

via the Life in Australia™ probability panel and asked respondents about their level of interest in politics and for which party they voted at the 2016 and 2019 elections. This data set can be found at https://dataverse.ada.edu.au/dataverse.xhtml?alias=CSES

[5] Available from https://dataverse.ada.edu.au/dataset.xhtml?persistentId=doi:10.26193/GOVGBB

[6] In June 2021 the Australian Polling Council, formed by nine commercial pollsters following the 2019 federal election (Essential Research and Communications Group, Ipsos, JWS Research, Lonergan Research, Newgate Research, Omnipoll, Telereach, uCommunication and YouGov Galaxy), made a public commitment to transparency via the release of its Code of Conduct and Disclosure Statement. See https://www.linkedin.com/company/australian-polling-council/?originalSubdomain=au

2 Previous research

Survey data are weighted (statistically adjusted) to mitigate the effects of coverage error, account for any disproportionality in the sample design, adjust for non-response and, finally, to balance (calibrate) the sample to key population parameters known to be correlated with the measures of interest – voting intentions, in the case of pre-election polls. The ultimate goal of weighting is to reduce bias in the reported survey estimates but when weighting is used to correct for an inadequate sample it will generally come at the cost of increased variance (see, e.g., Little & Vartivarian 2005, 161).

Peytchev et al. (2018, 492) and others point out that the effectiveness of weighting 'depends on the properties of the variables (known as "auxiliary" variables) used to construct the weights and that the ideal auxiliary variables for nonresponse adjustment are strongly associated with both nonresponse and the survey variables of interest.'

2.1 Educational attainment, non-response and vote choice

Previous research shows that one of the common biases in survey research in Australia and elsewhere is that, seemingly regardless of the sample frames used, surveys attempting to represent the adult population almost invariably over-represent people with higher levels of educational attainment (Kohut et al. 2012, 10; Pennay et al. 2018, 11–12). The study by Pennay et al. was conducted in late 2015 and compared the representativeness of three probability samples and five non-probability samples, all conducted in Australia, and found that on average, across all samples, respondents had a higher level of educational attainment than the general population. At that time 23% of the Australian adult population had a Bachelor Degree or higher (Australian Bureau of Statistics (ABS) 2016). However, of samples examined by Pennay et al. (2018, 11–12) in the three probability samples the proportion of respondents with a Bachelor Degree or higher was 41–44% and in the five non-probability samples, 31–42%.

Previous research also shows educational attainment is correlated with vote choice. Pennay et al. (2020, 55–56) used data from the Australian component of the Comparative Study of Electoral Systems survey (CSES Australia 2019)[7] to show that 54% of those with a tertiary education said they had voted for Labor or the Greens in the most recent election and 33% for the Coalition. Of those with non-tertiary qualifications, the corresponding figures were 45% (Labor) and 42% (Coalition); and of those with no post-school qualifications, the corresponding figures were 40% and 48%.

[7] This post-election survey of 2,000 Australian adults, conducted via the Life in Australia™ probability panel, asked respondents about their level of interest in politics and for which party they voted at the 2019 election.

2.2 Past vote, non-response and vote choice

An analysis of the relationship between reported past vote at the previous election and reported current vote using data from the 2019 Australian Election Survey showed that '77% of those respondents who reported voting for the ALP, Liberal Party, National Party or Greens in 2016 reported voting the same way in 2019' (Pennay et al. 2020, 60). International studies (Durand et al. 2015, Sturgis et al. 2016, 2017, Wells 2019) also demonstrate that past voting behaviour is strongly correlated with current vote choice and at face-value is a desirable candidate to use in any weighting solution for surveys trying to produce an accurate estimate of voting intentions.

Improving voting intention estimates by balancing or weighting one's sample by 'previous vote' is frequent in electoral polls, particularly in Europe (Durand et al. 2015, 1) but seemingly seldom used in Australia (Pennay et al. 2020, 16). YouGov's UK Director of Political and Social Research, Anthony Wells, observed that 'almost all [pollsters in the UK] … use how people voted at the last election as a target when designing or weighting … polling samples' (Wells 2019, 2).

However apparently attractive the prospect of balancing or weighting an election poll sample so that it is representative of the population based on the voting distribution at the last election seems at face-value, such an approach is not without problems. Recall of past vote may be inaccurate for three reasons: (1) memory failure; (2) the tendency of voters to misreport a previous vote in order to reconcile it with how they currently wish to vote; and (3) social desirability (Durand et al. 2015, 3–12). In addition, some respondents will not have voted in the previous election.

Despite its widespread use the jury still seems to be out on whether weighting by past vote is an effective and reliable way of reducing bias in estimates of voting intentions. In an examination of 12 election results – Canadian (5), Quebec (4) and French (3) – Durand and her co-authors found that a past vote weighting correction resulted in 'little difference between corrected and uncorrected estimates of voting intentions' (Durand et al. 2015, 12). Similarly, the AAPOR Taskforce's evaluation of the United States (US) pre-election polls in 2020 found that 'using the self-reported 2016 vote to weight to the past vote does not fix the polling error [in the 2020 polls]' (AAPOR 2021, 66). By way of contrast, a 2013 report by Cooper on behalf of the British Polling Council found that 'past vote weighting can and often does make a significant difference [bias reduction] to voting intention numbers' (Cooper 2013, 2).

A common feature of the analyses of past vote weighting undertaken by Durand et al., the APPOR Taskforce and Cooper is that the measure of past vote that is being relied upon for these weighting adjustments is one that is provided by respondents at a point in time close to the upcoming election and within the same questionnaire as they are being asked about their voting intentions at the upcoming election. Relying on such a measure of past vote is likely to amplify the recall, reconciliation and social desirability errors identified by Durand et al. (2015).

One of the primary areas of interest for this paper is whether weighting solutions that use a past vote measure that is less prone to the types of errors identified above – that is, one that is taken very soon following the previous election – results in a greater bias reduction in the subsequent estimate of voting intentions than does using a past vote measure captured at the same time as when the prospective voting intentions questions are being asked.

The implications of weighting based on an inaccurate past vote measure can be substantial. Wells (2019, 2–3) conducted an experiment in which he re-weighted YouGov polling data for the 2017 Brexit election using reported past vote as collected immediately after that election (at which time 41% reported voting for Labour) and reported past vote collected from the same respondents in 2019, at which time only 33% reported having voted for Labour in 2017.

The difference these two readings made to the estimate of current voting intentions was substantial. When the 2017 measure of ‘past vote’ was used as the past vote weighting variable, estimated support for Labour was 21%; when the 2019 measure of ‘past vote’ was used, estimated support for Labour increased to 24%.

3 Data

The data used for this study is from the 29th ANUpoll conducted in April 2019. The purpose of the ANUpolls is to assess Australians' opinions on important and topical issues. These polls are typically conducted as omnibus surveys at least three times a year. Since 2017, ANUpolls have been conducted on samples drawn from Australia's only probability-based online panel, Life in Australia™. Life in Australia™ was developed and is maintained by the Social Research Centre Pty Ltd. Prior to this (2008 to 2016) the ANUpolls were undertaken by telephone and administered to samples generated by Dual-frame Random Digit Dialling.

The main focus of the 29th ANUpoll was to measure Australians' opinions on a broad range of issues regarding the role of universities. Given the omnibus nature of the ANUpoll, the questionnaire also included questions on gambling and, critically for this study, questions on voting intentions. Data collection was undertaken during the period 8–26 April 2019. The primary question for analysis in this paper is 'If a federal election for the House of Representatives was held today, which one of the following parties would you vote for (Liberal, Nationals, Labor, Greens, Liberal-National Party (QLD only) or some other party)?' The federal election was held on 18 May 2019 with pre-poll voting open from 29 April (Australian Electoral Commission (AEC) 2019a). The voting intentions estimates generated from this ANUpoll were published in a paper by Biddle (2019), *Predicting the unpredicted: what longitudinal data can tell us about the 2019 Australian federal election*.

To determine whether analyses of these data might support the drawing of more general inferences for commercial pollsters, the first question to be answered is, how do the pre-election voting intentions estimates produced by Biddle (2019) compare with those produced by the commercial polls at a similar point in the 2019 federal election cycle? Table 1 reproduces (and expands upon) the published estimates from the ANUpoll (Biddle 2019, 6) and compares these with estimates produced by the commercial pollsters from polls conducted at about the same time. This prima facie analysis shows that the ANUpoll yielded estimates that were broadly similar to those produced by the commercial pollsters. The weighted two-party preferred (2PP) estimates produced by the ANUpoll are identical to those produced by Ipsos, Newspoll and YouGov Galaxy and differed by 1 percentage point from the more accurate estimates produced by Essential and Roy Morgan.[8]

[8] Biddle (2019) did not report 2PP estimates. See Appendix for a description of how these were derived for this study.

Table 1 Voting intentions as measured by the ANUpoll and other pre-election polls compared with the federal election outcome, 2019

Poll		Primary vote (%)							2PP (%)	
	Dates	LNP	ALP	Greens	PHON	UAP	Other	DK	LNP	ALP
Election	18 May	41.4	33.3	10.4	3.1	3.4	8.3		51.5	48.5
ANUpoll	8–26 April	36	32	14	2	2	10	4	48	52
Essential	24–29 April	39	37	9	6	NS	9	NS	49	51
Ipsos	1–4 May	36	33	14	5.0	3	9	NS	48	52
Newspoll	11–14 April	39.0	39.0	9.0	4.0	3.0	6.0	5.0	48	52
Roy Morgan	20–21 April	39.0	35.5	9.5	4.5	2.0	9.5	3.5	49.0	51.0
YouGov Galaxy	23–25 April	37	37	9	4	4	9	NS	48	52

Key: ALP – Australian Labor Party; DK – Don't Know; LNP – Liberal-National Party coalition ('the Coalition'); NS – Not Stated; PHON – Pauline Hanson's One Nation; UAP – United Australia Party.
Source: Final election results as reported by the AEC (2019b). ANUpoll primary vote estimates as reported by Biddle (2019) with the 2PP calculated separately (see Appendix for further details). The estimates for the commercial pollsters are as reported in Goot (2021). Roy Morgan published their estimates to the nearsets 0.5% all others round to the nearest whole number.

3.1 The relationship between educational attainment, non-response and vote choice in the ANUpoll

3.1.1 Measuring educational attainment

Educational attainment was one of the profiling measures collected when members of the Life in Australia^{TM} panel were recruited. It is updated periodically. The various aspects of educational attainment measured are highest level of primary or secondary schooling completed, attainment or otherwise of any post-school qualifications and highest level of post-school educational attainment. Highest level of educational attainment is collected in five categories; (1) Postgraduate Degree – comprising of Master Degree, Doctoral Degree, other Postgraduate Degree, Graduate Diploma and/or Graduate Certificate; (2) Bachelor Degree; (3) Advanced Diploma and/or Diploma; (4) Certificate III and/or IV; and (5) Certificate I and/or II. For weighting purposes highest level of educational attainment was categorised as Bachelor Degree or higher (1, 2) and below Bachelor Degree (3, 4 and 5). See Appendix for benchmarks.

3.1.2 Educational attainment and non-response

The relationship between educational attainment and the dual weighting efficacy criteria of being associated with non-response and the variable of interest (vote choice) is shown below (Table 2). This analysis shows that for this sample, educational attainment is indeed associated with non-response. Forty-four per cent of respondents to the ANUpoll had a tertiary qualification compared with the population benchmark of 26%. At the other end of the scale the ANUpoll under-represents the population without a post-secondary qualification relative to the population (26% compared to 42%).

Table 2 ANUpoll estimates (unweighted) compared to educational attainment benchmarks

<table>
<tr><th>Highest educational qualification</th><th>Benchmark</th><th>ANUpoll (unweighted)</th></tr>
<tr><td>Tertiary qualifications</td><td>26</td><td>44</td></tr>
<tr><td>Postgraduate Degree level</td><td>6</td><td>15</td></tr>
<tr><td>Graduate Diploma and Graduate Certificate level</td><td>2</td><td>9</td></tr>
<tr><td>Bachelor Degree level</td><td>18</td><td>20</td></tr>
<tr><td>Non-tertiary qualifications</td><td>29</td><td>28</td></tr>
<tr><td>Advanced Diploma and Diploma level</td><td>10</td><td>12</td></tr>
<tr><td>Certificate III & IV level</td><td>18</td><td>15</td></tr>
<tr><td>Certificate I & II level</td><td>0.1</td><td>1</td></tr>
<tr><td>No qualifications</td><td>42</td><td>26</td></tr>
<tr><td>Secondary education – Year 12</td><td rowspan="2">34</td><td>12</td></tr>
<tr><td>Secondary education – Years 10 and 11</td><td>11</td></tr>
<tr><td>Secondary Education – Years 9 and below</td><td>8</td><td>3</td></tr>
</table>

Figures do not add to 100% as not stated and inadequately described responses have been excluded from the benchmark figures and the survey data.
Source: ABS (2016)

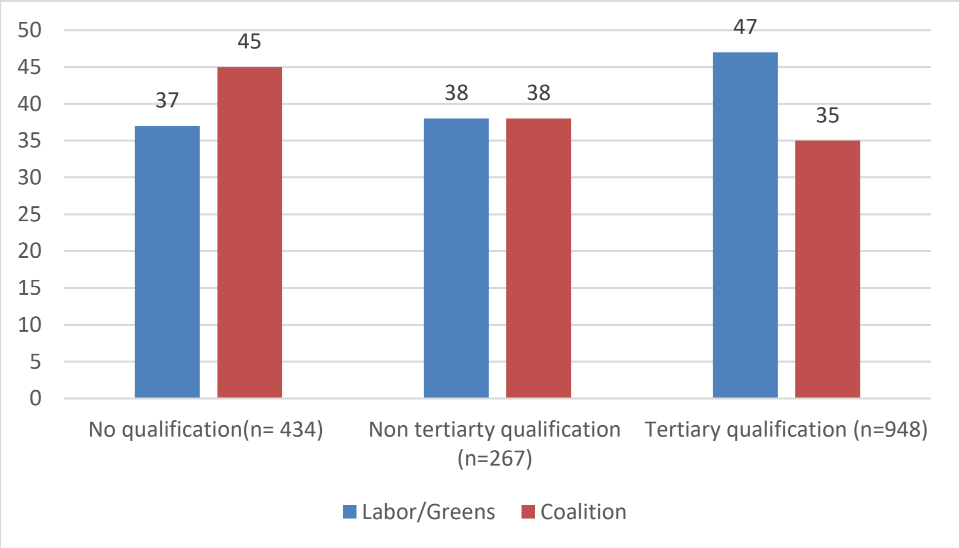
3.1.3 Educational attainment and voting intentions

A relationship between educational attainment and voting intentions (see Figure 1) is also evident in the ANUpoll data. Of those with a tertiary qualification, 47% said they intended to vote for Labor or the Greens and 33% for the Coalition; of those with non-tertiary qualifications 38% supported the Coalition with an identical proportion supporting Labor or the Greens. Among those with no post-secondary qualifications, 45% intended to vote for the Coalition and 37% Labor or the Greens. Given the relative size of the three educational attainment segments (no qualification (n=434), non-Tertiary qualification (n=267) and Tertiary qualification (n=948)), and the relative over-representation of those with tertiary qualifications, we can conclude that the educational attainment bias in this ANUpoll data, if not mitigated by weighting, would, clearly be a source of non-response bias.

As such, educational attainment meets the dual weighting efficacy criteria of being related to both non-response and the outcome measure of interest.

Our first hypothesis is that including a poststratification adjustment for educational attainment in the weighting solution will reduce bias and have an acceptable amount of variance compared to using only the more routine adjustments of age, sex and geography **(Hypothesis 1).**

Figure 1 The relationship between educational attainment and current voting intentions, unweighted ANUpoll data (%), 2019



Source: Authors' analysis

3.2 The relationship between past vote, non-response and vote choice in the ANUpoll

3.2.1 Measuring past vote

Given our interest is to also determine the impact that weighting by past vote has on the resultant estimates of voting intentions, it was necessary to identify and append appropriate 'past vote' variables to the ANUpoll dataset. The ability to append auxiliary variables (collected at other points in time) to respondents' answers is one of the very attractive features of panel surveys such as the ANUpoll and panel platforms such as Life in Australia^{TM}.

Two such measures from the Life in Australia^{TM} panel profile were appended to the ANUpoll data: (1) a measure of past vote based on short-term recall of past voting behaviour (the 'short-term' recall measure), and (2) a measure of past vote based on long-term recall of past voting behaviour (the 'long-term' recall measure). This broadly replicates the two past vote measures used by Wells (2019).

Short-term recall of past vote: The Life in Australia^{TM} panel was originally recruited by the Social Research Centre via a Dual-frame Random Digit Dialling sample frame in August/September 2016. The initial recruitment effort contacted 27,852 landline and mobile phone telephone numbers and resulted in 3,042 panel members, a recruitment rate of 21.6% (see Kaczmirek et al. 2019 for full details). These original panellists were asked about their vote choice at the preceding federal election held on 2 July 2016. The question wording and response options are as follows:

'Some people were unable to vote or chose not to vote in the last federal election. Did you vote in the federal election held on 2 July 2016?

1. Yes
2. No
3. (Don't know/can't recall)
4. (Refused)'

(If voted) 'Which party did you vote for first in the House of Representatives?

1. Liberal Party
2. Labor Party (ALP)
3. National (Country) Party
4. Greens
5. Other (please specify party), or
6. Voted informal
7. (Don't know/can't recall)
8. (Refused)'

The historic responses to this vote choice question were mapped to the ANUpoll responses and form our short-term recall measure of past vote.

Long-term recall of past vote: A long-term recall measure of past vote was included in the 2019 CSES conducted on the Life in Australia^{TM} panel in June 2019, that is, one month after the May 2019 federal election and some three years after the 2016 federal election. After a series of questions relating to the 2019 federal election, including vote choice at the 2019 election, respondents were asked:

'In the last Federal election in July 2016, when the Liberals were led by Malcolm Turnbull and Labor by Bill Shorten, which party got your first preference then in the House of Representatives election?

1. Liberal Party
2. National Party
3. Labor Party (ALP)
4. Greens
6. Liberal National Party (L-NP) (Queensland only)
96. Some other party/independent
97. Did not vote
95. Not eligible to vote
98. (Don't know)/Not sure
99. (Refused)/Prefer not to say'

The responses to this question were mapped to the ANUpoll data and form our long-term recall measure of past vote. See Appendix for past vote benchmarks and further details.

Apart from the differences in the timing of when the short-term and long-term recall measures of past vote were collected (some two-to-three months post-election for the short-term measure and almost three years post-election for the long-term measure) there are also some methodological differences in the framing of these questions. The question used to provide our short-term measure of past vote is preceded by a screening/filter question to identify whether respondents had voted in the 2016 election. The intention of this screening/filter question is to try to limit presenting the actual vote-choice question to respondents who cast a vote. The question which provides our long-term measure is not preceded by a screening/filter question and instead relies on respondents volunteering as part of their response whether they voted at the 2016 election. A further complication with our long-term past vote question is that it asks respondents who they voted for in the 2016 election following the subsequent election in 2019 and following a 2019 vote choice question. The relationship between the two variables is shown in Table 3. Eighty-six per cent of those who reported voting for the

Coalition at the July 2016 election, when asked in August/September 2016, reported voting for the Coalition when asked again in June 2019, following the subsequent May 2019 federal election.

The corresponding levels of concordance between these two measures of past vote for the other parties are 76% for Labor, 57% for the Greens and 41% for other parties.

Table 3 ANUpoll estimates (unweighted): short-term recall of past vote by long-term recall of past vote (row percentages)

	Long-term recall of party voted for at previous election				
Short-term recall of party voted for at previous election	Coalition %	Australian Labor Party %	Greens %	Other formal vote %	Informal – Did not vote – Not enrolled %
Coalition	86	6	1	5	2
Australian Labor Party	14	76	4	4	2
Greens	7	27	57	7	2
Other formal vote	33	15	7	41	5
Informal – Did not vote – Not enrolled	27	16	5	8	44

Source: Authors' analysis

Clearly, if this study was being designed from scratch rather than in retrospect then both the short-term and long-term past vote questions would have been administered in an identical fashion (the only difference being the timing of when the questions were asked) and the long-term recall measure would have been asked prior to the subsequent (2019) election. As it stands the different rendering of these questions is a limitation of this research; one of the implications being that the relationship between past vote and non-response is not as clear cut as it might otherwise have been due to differential measurement error.

3.2.2 Past vote and non-response

Table 4 illustrates that a relationship between past vote and non-response is present in the ANUpoll data but varies depending upon the past vote measure that is used. Based upon the short-term recall measure of past vote, the ANUpoll very accurately reflects the proportion of 2016 Coalition voters (42.7% compared with a benchmark of 42%), under-represents 2016 Labor voters (by 2.5 percentage points), over-represents 2016 Green voters by 2.9 percentage points and under-represents support for other parties by 2 percentage points. When using the long-term recall measure of past vote, the ANUpoll has a much higher proportion of 2016 Coalition supporters (47% compared to the benchmark 42%), under-represents support for Labor by 2.2

percentage points, very accurately reflects the level of support for the Greens at the 2016 election (within 0.6 percentage points) and underestimates support for other parties with an estimates level of support of 9.7% compared to a benchmark of 13%.

The higher proportion of respondents recalling a past vote for the Coalition when based on the long-term recall measure of past vote (47%) compared to the short-term measure of past vote (42.7%) is consistent with the reconciliation hypothesis put forward by Durand et al. (2015).

Subject to the interaction with other weighting variables, the differences between the two past vote measures are such that we would expect them to have a differential effect on the resultant weighted estimates of voting intentions. At face-value the disproportionate down-weighting of support for the Coalition required to adjust the long-term recall measure of past vote to benchmarks is likely to reduce the voting intentions estimate for the Coalition.

Table 4 ANUpoll primary vote estimates (unweighted) compared with past vote benchmarks

Vote choice	2016 election	ANUpoll Short-term recall of past vote	ANUpoll Long-term recall of past vote
	% of formal votes[a]	% formal votes[b]	% of formal votes[c]
Coalition	42.0	42.7	47.0
Labor	34.7	32.2	32.5
The Greens	10.2	14.1	10.8
Other formal vote	13.0	11.0	9.7

a AEC (2016).
b Excludes screen outs as well as don't knows, refused, informal votes and votes for no party.
c Excludes don't knows, refused, not eligible/did not vote, informal votes and votes for no party.
Source: Authors' analysis

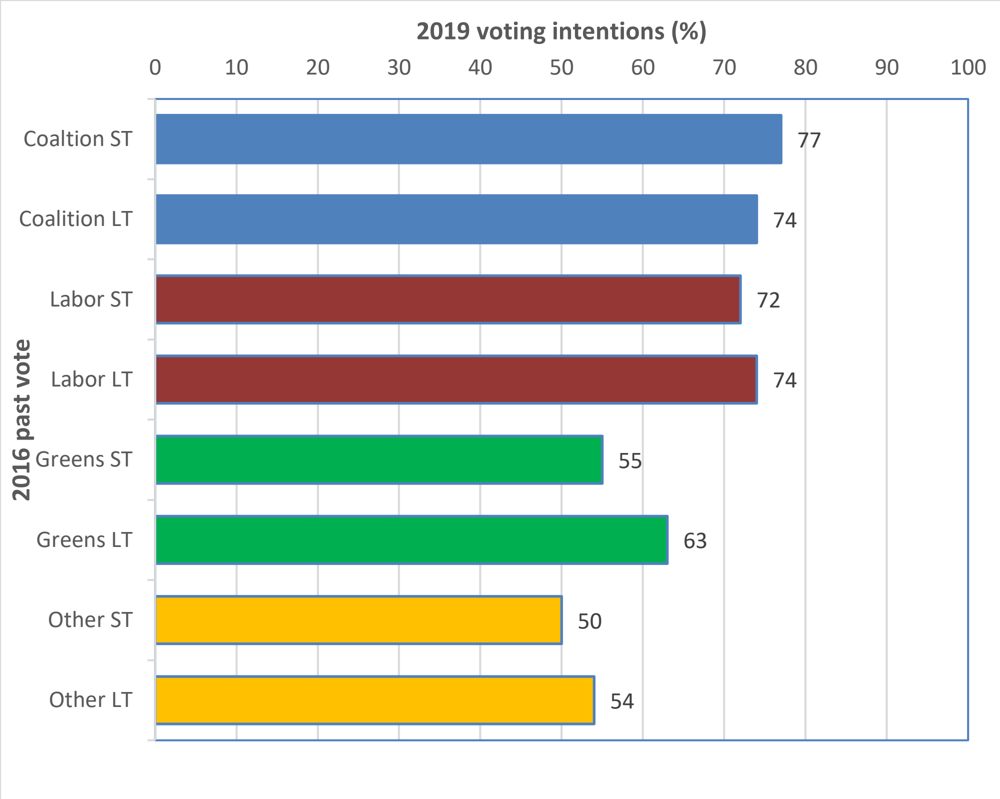
3.2.3 Past vote and current voting intentions

A correlation between past vote and current voting intentions is also evident in the ANUpoll data (see Figure 2). Based on the short-term (ST) recall measure of past vote), over three-quarters (77%) of those who reportedly voted for the Coalition in 2016 declared an intention to vote the same way in 2019. The corresponding result for Labor was 72%, 55% for the Greens and 50% for other parties. A similar pattern is evident when using the long-term (LT) measure of past vote. The correspondence between a reported past vote for Labor or the Coalition and a current intention to vote the same way was 74% for both Labor and the Coalition. Sixty-three percent of those who reported voting for the Greens in 2016 reported an intention to do so again in 2019 and 54% of those who reportedly cast their vote for another party in 2016 intended to do the same in 2019.

On this basis, both past vote measures meet the dual weighting efficacy criteria of being related to both non-response and the outcome measure of interest.

Our second hypothesis is that including a poststratification adjustment to past vote benchmarks, using a short-term recall measure of past vote, will reduce bias and have an acceptable amount of variance compared to weighting adjustments based on a long-term recall measure of past vote **(Hypothesis 2).**

Figure 2 Voting intentions in 2019 by reported past vote at the 2016 election



Key: LT – Long-term, ST – Short-term.
Source: Authors' calculations

4 Methods

4.1 ANUpoll subset

To enable the most robust possible comparison of the two past vote measures we limit our analysis to a subset of ANUpoll sample for which we have both past vote measures available. These data were available for 1,684 of the 2,205 respondents in the original ANUpoll data set. These 1,684 respondents form the data set for the remainder of our analysis.

4.2 Approach to weighting

To investigate the relative impact of educational attainment and past vote on estimates of voting intentions it was necessary to incorporate these measures into an appropriate weighting solution. To do this we stripped the data of all previous weights, including the design weights, and started from the unadjusted raw data. From this starting point we built up the weighting scenarios as follows:

- Weight 1: Age, sex and geography (state by capital city/rest of state). This is our baseline weight and main point of comparison for other weighting solutions.
- Weight 2: Age, sex, geography and educational attainment. The application of this weighting solution enables us to isolate the impact of adding educational attainment to the baseline weighting solution.
- Weight 3: Age by education, sex and geography. Same variables as above but for this weight, age is crossed with education to account for the age-related educational attainment gradient evident in the population (with educational attainment generally declining by age group).

The application of this weight will enable an assessment of whether this more granular approach to incorporating educational attainment is beneficial.

- Weights 4 to 6: The above (1-3) with the short-term measure of past vote added into the weights.
- Weights 7 to 9: The above (1-3) with the long-term measure of past vote added into the weights.
- Weights 10 to 11: Weighting just by short-term recall of past vote (weight 10) and long-term recall of past vote (weight 11).
- Weights 12 to 14: Weights 1-3 combined with a blended (short-term/long-term) measure of past vote.
- Weight 15: The blended measure of past vote on its own.

Each weight was calculated using the rake procedure from the survey package in R (Lumley 2020, 2010, 2004).

The population benchmarks for the demographic variables (age, sex, geography and educational attainment) were compiled from the ABS’s 2016 Census counts and the March 2019 Estimated Residential Population (ERP) figures. The voting benchmarks were released by the AEC and show the number (and proportion) of formal votes cast for each party, thereby excluding informal votes, unenrolled and ineligible members of the population and eligible non-voters. Given the absence of a reliable benchmark figure for non-voters and given the different way the two past vote measures accounted for non-voters, we decided that the best way to account for non-voting survey respondents in the weights was to treat them as missing.[9] This means that the non-voting proportion for each past vote recall measure remains unadjusted and support for each party is re-apportioned so as to match the proportional distribution of the AEC formal vote benchmarks. The benchmarks used for these adjustments are provided in the Appendix, Tables A2 and A3.

4.3 Error metrics

4.3.1 Measures of bias

Several measures of bias are commonly used to evaluate the performance of pre-election polls, all of which have the common feature of measuring bias by comparing the voting intention estimates/predictions of the polls against the actual election outcome. The measures used in recent reviews of the polls (Sturgis et al. 2016; AAPOR 2017; Pennay et al. 2020) include average absolute error on the primary vote, weighted average absolute error on the primary vote and absolute error on the primary vote margin.

The absolute errors of the original ANUpoll estimates (Table 5) show an error of 5.8 percentage points for the Coalition (LNP), 2.5 percentage points for the Greens, 1.6 percentage points for the United Australia Party (UAP) and 1.2 percentage points for both Labor (ALP) and Pauline Hanson's One Nation (PHON). The average absolute error using this metric is 2.5 percentage points (5.8+1.2+2.5+1.2+1.6)/5) =2.5.

[9] The authors were unable to directly source or satisfactorily derive eligible voter benchmarks from publicly available data released by the ABS, the AEC or the Australian Government Department of Home Affairs.

Table 5 Absolute errors of original ANUpoll estimates, 2019

Poll/Dates		Primary vote (%)							2PP vote (%)	
	Dates	LNP	ALP	Greens	PHON	UAP	Other	DK	LNP	ALP
Election	18 May	41.4	33.3	10.4	3.1	3.4	8.3		51.5	48.5
ANUpoll	8–26 April	35.6	32.1	13.9	2.3	1.8	10.2	4.1	48.0	52.0
Absolute error		5.8	1.2	2.5	1.2	1.6			3.5	3.5

Key: ALP – Australian Labor Party; DK – ; LNP – Liberal-National Party coalition ('the Coalition'); PHON – UAP – United Australia Party.
Source: Australian Election Commission and authors' analysis

The average absolute error metric gives equal weight to the estimate for each of the parties. However, given preferential voting in Australia, the relative accuracy of the poll estimates for the main parties (Labor and the Coalition) is more consequential in terms of the 2PP vote than the relative accuracy of the estimates for the other parties. As such, a better measure of the overall performance of the polls is a measure of bias that takes into account the average absolute error on the primary vote weighted by vote share. This is calculated by multiplying the primary vote share by the primary vote error for each party. The weighted average absolute error on the primary vote for the ANUpoll (using re-based vote share distribution which excludes other minor parties and adds to 100%) is LNP (0.452 x 5.8) + ALP (0.364 x 1.2) + GRN (0.113 x 2.5) = PHON (0.034 x 1.2) + UAP (0.037 x 1.6) = 2.9 percentage points.[10] This is our preferred measure of primary vote bias.

In addition to assessing the accuracy of the polls based on the weighted average absolute error of their primary vote, it is also important to have a measure which captures the bias in the polls based on the estimate of the two party-preferred vote (2PP). Since the early 1990s, when all the Australian election pollsters began presenting their results in terms of a 2PP vote, this has become the principal way of evaluating their performance. The average absolute error of the ANUpoll 2PP estimate is 3.5 percentage points.

4.3.2 Measures of variance

The variance introduced by the weights is measured using the design effect calculated by Taylor series linearisation by the svymean procedure in the survey package in R (Lumley 2020). The design effect reflects the ratio between the standard error calculated using appropriate adjustments for survey design over what the standard error would be if a simple random sample was assumed. Higher design effect indicates more

[10] The proportions used as the weights in this calculation are the benchmarks proportions prorated to adjust for the 'other' proportion not being included (see Table 5). So .452 = .414 / (1 - .083)

variability in the weights relevant to the statistic being calculated.

4.3.3 Overall measure

Mean square error (MSE) (Korn & Graubard 1999) is a measure which combines bias and variances to assess the impact of weighting on the total survey error as follows:

$$MSE = B^2 + V$$

where B is the primary measure of bias (in this case the 2PP bias) and V is a measure of variance estimated from the data set. Korn & Graubard (1999) estimate the design effect using the variance of the weights. However, given the assessment of accuracy in this paper is focused on a single measure (2PP vote) the Taylor series linearised design effect is used. This paper uses root mean squared error (RMSE) so that the result is on the original scale of the percentages.

4.4 Simulations

To obtain estimates of the degree to which the different voting intentions results produced by the different weighting solutions were due to sampling variation, 10,000 samples were obtained by random re-sampling with replacement of the original data to the same sample size. Each weighting scheme was calculated for each re-sample to obtain estimates for all weighting options. The reported standard errors represent the 95% confidence intervals of the re-samples, i.e.:

$$CI = \ t^* \pm 1.96\,.\,se^*$$

where CI is the confidence interval, t^* is the average esitmate and se^* is the standard deviation of the 10,000 resamples.

Probabilities represent the proprtion that one weighting scheme produces superior results to another weighting scheme adjusted to be two-tailed probabilities, i.e.:

$$p = \frac{\begin{cases} p^* \leq \ .5, \ p \\ p^* > \ .5, \ (1-p) \end{cases}}{2}$$

where p^* is the proporton of re-samples where one weighting scheme is superior to the other. Probabilities were adjusted for multiple comparisons using the technique described by Benjamini & Hochberg (1995) using the p.adjust function in R.

5 Results

Table 6 shows the each error metric calculated based on the original data, as well as the average for each metric from the simulations, alongside their 95% confidence intervals (with the simulated results in brackets). Table A4 (appended) shows the probabilities associated with the null hypothesis. In this case the RMSE of the 2PP vote (i.e. the final column of Table 6) for one weighting solution is equal to the RSME of a comparative weighting solution.

In interpreting the probability testing results it is important to note that '… a label of statistical significance does not mean or imply that an association or effect is highly probable, real, true, or important. Nor does a label of statistical nonsignificance lead to the association or effect being improbable, absent, false, or unimportant' (Wasserstein et al. 2019, 2).

Taken together, these results (in Table 6 and Table A4) show that the unweighted data have very little bias and provide the best overall estimate of voting intentions. While this suggests a strong prima facie case for using a high-quality probability-based online panel as a sampling frame for pre-election polling, this does need to be balanced by the knowledge that the original weighted estimates of voting intentions produced by the ANUpoll were not superior to those of the commercial pollsters. Of course, it is only with the benefit of hindsight that we can see that the unweighted ANUpoll estimates were less biased than the weighted estimates. It also needs to be borne in mind that the ANUpoll operates as an omnibus survey and relying on unadjusted estimates for all the other measures included in the survey is unlikely to be the optimal approach.

Weighting solution 1 (weighting by age, sex and geography) is our baseline weight for this analysis and is the type of weighting solution that, at the time of the 2019 federal election, was fairly typically used by commercial pollsters (see Pennay et al. 2020, 16). In terms of the measures of bias associated with weight 1, the weighted average absolute error on the primary vote is 3.24 percentage points and the average absolute 2PP error is 5.15 percentage points. This weighting design increases the standard error by a factor of 1.27 and has a RMSE (which encapsulates both bias and variance) of 5.31. Adding education (weight 2) increases the design effect to 1.46 but due to a reduction in bias provides a better overall solution (as measured by a reduction in RMSE) than weight 1. Crossing education with age (weight 3) results in slightly lower design effect than weight 2 and a reasonable reduction in both measures of bias meaning that, at face-value, it is the best of these three approaches. However, reference to Table A4 shows that the beneficial impacts of adding an educational attainment weighting adjustment to the baseline weighting solution (weight 3 vs weight 1) does not meet the $p<.05$ threshold of statistical significance ($p=.075$). It is worth noting, however, that in 97% of the simulations we ran, weight 3 produced a less biased estimate than weight 1.

Based on these observations, Hypothesis 1 is partially supported. Weighting solutions that included an adjustment for educational attainment consistently produced less biased estimates of voting intentions with only a modest impact on variance compared to weighting adjustments using only the more routine adjustments of age, sex

and geography. However, based on the data available for this study, the level of improvement achieved was not statistically significant.

Weights 4 to 9 enable us to evaluate the impact of adding our two different measures of past vote (short-term and long-term) to the above weighting solutions. A comparison of weight 4 (age, sex, geography and short-term recall of past vote) and weight 5 (age, sex, geography and long-term recall of past vote) shows that the short-term measure of past vote produces vote choice estimates with considerably less bias than those achieved with the long-term recall measure. The weighted average absolute error on the primary vote is 1.80 percentage points when the short-term measure is used compared to 3.23 percentage points for the long-term measure. The short-term measure also out-performs the long-term measure in terms of the average absolute 2PP estimate with an error of 2.99 percentage points for the short-term recall measure compared with 4.90 percentage points for the long-term recall measure and results in a significant reduction in RMSE (see Table A4). It is also the case that weight 4 has a significantly lower RMSE than weight 1. When a past vote adjustment is added to the age, sex, geography and education weighting solution a similar picture emerges. In these cases (weight 7 versus weight 6 and weight 9 versus weight 8) the inclusion of the long-term recall measure of past vote results in a significantly larger RMSE in comparison to that resulting from the use of the short-term recall measure.

These comparisons show that adding a short-term past vote adjustment to the various weighting solutions results in less biased estimates than when using the long-term recall measure and also results in better estimates than solutions that do not include any past vote adjustment.

Another important comparison is that weight 9 (age and education by sex by geography by long-term recall), the best of the multi-factor solutions using the long-term measure of past vote, does not produce a statistically superior result than weight 1 (age, sex and geography). This means that incorporating a long-term recall measure of past vote in conjunction with educational attainment does not, on this occasion, produce a clearly better solution than just weighting by age, sex and geography. Based on a comparison of weights 1 to 9, the best overall weighting solution is weight 8 (age by education, sex, geography and the short-term recall measure of past vote).

On this basis our second hypothesis is also supported. Weighting solutions that calibrated to past vote benchmarks, based on a short-term recall measure of past vote, had reduced bias and less variance than weighting solutions based on a long-term recall measure of past vote.

We can extend the above to also state that weighting solutions that used a short-term measure of past vote performed better than weighting solutions that did not include any adjustment for past vote.

For the sake of completeness we also ran weights 10 and 11, these being single factor weighting solutions which adjusted the estimates just by the short-term measure of past vote (weight 10) and the long-term measure of past vote (weight 11).

Just adding the short-term recall weighting adjustment, on its own, results in the best overall weighting solution

(RMSE of 1.10) with weight 8 being the next best. Weight 11 (weighting by long-term recall of past vote on its own) results in the third-best weighting solution of the 11 compared so far. The strong performance of these single factor weighting solutions makes sense given the relatively unbiased nature of the unweighted sample and the very high correlation between reported past vote and current voting intentions.

A further reason why just weighting by past vote might act as a catchall solution is the high correlation between voting behaviour and other variables typically used to weight surveys/polls such as gender, age and location as well as other sociodemographic characteristics such as social class, income, educational attainment and asset ownership (Cameron & McAllister 2019, 17–22).

Table 6 Comparison of weighting solutions by specified error metrics

		Primary vote	2PP			
No.	Solutions	Weighted Avge absolute error	Avge Absolute error	Std. error	Design effect	Root Mean square error
	Unweighted estimate	0.63 (1.16, 0.14 – 2.18)	0.78 (1.12, -0.50 – 2.75)	1.15 (1.15, 1.13 – 1.16)	1.00 (1.00, 1.00 – 1.00)	1.39 (1.70, 0.53 – 2.88)
1	Weighted by age, sex, geography	3.24 (3.49, 1.89 – 5.10)	5.15 (5.15, 2.67 – 7.64)	1.27 (1.26, 1.18 – 1.35)	1.26 (1.26, 1.10 – 1.41)	5.31 (5.32, 2.93 – 7.71)
	Weighting estimates by education					
2	Age, sex, geography, education	3.00 (3.18, 1.38 – 4.98)	4.57 (4.55, 1.89 – 7.21)	1.37 (1.37, 1.27 – 1.47)	1.46 (1.47, 1.27 – 1.67)	4.77 (4.78, 2.28 – 7.27)
3	Age by education, sex, geography	2.58 (2.93, 1.30 – 4.56)	4.08 (4.08, 1.46 – 6.71)	1.34 (1.34, 1.25 – 1.43)	1.41 (1.42, 1.24 – 1.59)	4.29 (4.33, 1.90 – 6.76)
	Weighted estimates by past vote					
4	Age, sex, geography and short-term	1.80 (2.02, 0.58 – 3.46)	2.99 (2.96, 0.68 – 5.24)	1.18 (1.17, 1.03 – 1.32)	1.06 (1.06, 0.80 – 1.32)	3.21 (3.23, 1.20 – 5.26)
5	Age, sex, geography and long-term	3.23 (3.31, 1.88 – 4.75)	4.90 (4.90, 2.94 – 6.87)	1.00 (1.00, 0.87 – 1.13)	0.78 (0.79, 0.59 – 0.99)	5.00 (5.01, 3.09 – 6.93)
6	Age, sex, geography, education and short-term	1.70 (2.00, 0.52 – 3.48)	2.77 (2.74, 0.35 – 5.14)	1.27 (1.27, 1.10 – 1.45)	1.25 (1.26, 0.92 – 1.60)	3.05 (3.09, 1.03 – 5.15)
7	Age, sex, geography, education and long-term	3.25 (3.30, 1.69 – 4.92)	4.81 (4.79, 2.59 – 6.98)	1.14 (1.14, 0.97 – 1.30)	1.01 (1.02, 0.72 – 1.32)	4.94 (4.93, 2.82 – 7.04)
8	Age by education, sex, geography and short-term	1.41 (1.81, 0.47 – 3.15)	2.41 (2.39, 0.12 – 4.67)	1.22 (1.22, 1.07 – 1.36)	1.14 (1.14, 0.88 – 1.41)	2.70 (2.76, 0.84 – 4.68)

Table 6 Comparison of weighting solutions by specified error metrics (cont.)

		Primary vote	2PP			
No.	Solutions	Weighted Avge absolute error	Avge Absolute error	Std error	Design effect	Root Mean square error
9	Age by education, sex, geography and long term	2.95 (3.06, 1.59 – 4.53)	4.45 (4.44, 2.35 – 6.52)	1.07 (1.07, 0.94 – 1.20)	0.90 (0.90, 0.68 – 1.12)	4.57 (4.57, 2.57 – 6.58)
	Weighted estimates by past vote only					
10	Weighting by short-term past vote only	0.20 (0.82, 0.04 – 1.60)	0.64 (0.89, -0.37 – 2.15)	0.90 (0.90, 0.86 – 0.93)	0.61 (0.61, 0.56 – 0.65)	1.10 (1.34, 0.43 – 2.25)
11	Weighting by long-term past vote only	1.76 (1.85, 0.70 – 3.00)	2.80 (2.80, 1.19 – 4.42)	0.82 (0.82, 0.78 – 0.86)	0.52 (0.52, 0.47 – 0.57)	2.92 (2.93, 1.41 – 4.46)
	Weighted estimates using a blended estimate of past vote					
12	Age, sex, geography and blended past vote	2.84 (2.91, 1.32 – 4.51)	4.39 (4.37, 2.17 – 6.56)	1.12 (1.12, 0.98 – 1.26)	0.97 (0.97, 0.73 – 1.22)	4.53 (4.52, 2.42 – 6.62)
13	Age, sex, geography, education and blended estimates of past vote	2.90 (2.97, 1.33 – 4.62)	4.36 (4.33, 2.06 – 6.61)	1.18 (1.18, 1.02 – 1.34)	1.09 (1.09, 0.79 – 1.39)	4.52 (4.51, 2.34 – 6.67)
14	Age by education, sex, geography and blended estimate of past vote term	2.54 (2.68, 1.14 – 4.22)	3.92 (3.89, 1.64 – 6.15)	1.16 (1.16, 1.02 – 1.30)	1.04 (1.05, 0.79 – 1.30)	4.09 (4.08, 1.96 – 6.20)
15	Weighted by blended measure of past vote only	1.01 (1.17, 0.09 – 2.25)	1.84 (1.84, 0.19 – 3.49)	0.87 (0.87, 0.83 – 0.91)	0.58 (0.58, 0.53 – 0.62)	2.03 (2.08, 0.66 – 3.51)

Results outside brackets represent the observed estimate based on the original data, results in brackets represent the avergae estimate from the simulated samples and the upper and lower confidene intervals.
Source: Australian Election Commission and authors' analysis

One argument against the generalisability of these findings is the high level of attrition and low response rates experienced by commercial online panels (Tourangeau et al. 2013, 42) which means that, even if the panel operators routinely collect a short-term measure of past vote from all panel members, it is unlikely that many of these panel members will be responding to survey requests when the next pre-election polling cycle is underway in earnest. This means that the only available measure of past vote likely to be available to pollsters using non-probability online panels is one collected at different times over the election cycle. This poses the question, what impact would using a hybrid short-term/long-term recall measure of past vote in weighting solutions have on the resultant estimates of voting intentions?

While we cannot replicate this situation exactly, we can approximate it by creating a measure of past vote using a combination of our short-term and long-term measures. To do this we created a 'blended' past vote variable whereby a random half of respondents were allocated their short-term measure of past vote and the remaining half their long-term measure of past vote. This blended past vote measure was then added to our three primary weighting solutions (see Table 6, weights 12 to 14) and run on its own (weight 15).

A comparison of our baseline weighting solution (age, sex, geography – weight 1) with weight 12 (age, sex, geography and the blended measure of past vote) shows that the addition of the blended measure of past vote produces prima facie (but not statistically significant) reductions in bias for the weighted average absolute primary vote, the absolute average 2PP vote and the resultant RMSE. Adding the blended past vote measure to the age, sex, geography and education weight (weight 13 compared with weight 2) also shows prima facie but not statistically significant improvements. The same is true when comparing weight 3 (age by education, sex and geography) with weight 14 (age by education, sex, geography and blended past vote). Using the blended measure of past vote on its own (weight 15) results in the second best overall weighting solution of all 15 options (with a RMSE of 2.03), second only to weighting only by short-term recall of past vote (weight 10 – RMSE 1.10).

Table 7 shows the actual 2PP vote choice estimates produced by selected weighting solutions. This provides another view on the impact of the various weighting scenarios. The most basic of the weighting solutions (weight 1) produces a 2PP estimate of 46.4% for the Coalition and 53.6% for Labor – some 5.1 percentage points adrift of the election outcome. Compare this with our best composite weighting solution (weight 8 – age and education by sex by geography by short-term past vote) which produces a 2PP estimate of 49.3% for the Coalition and 50.7% for Labor, an improvement of 2.9 percentage points and now 2.2 percentage points shy of the election outcome. The weighting solution that uses just the blended measure of past-vote recall on its own (perhaps the only measure of past vote that will be available to pollsters using non-probability online panels) produces the second best overall estimate of voting intentions. Finally, our best overall weighting solution (weight 10 – short-term recall measure only) has the Coalition (50.9%) ahead of Labor (49.1%). This is the only solution to have the Coalition ahead and is just 0.6 percentage points away from the actual election outcome.

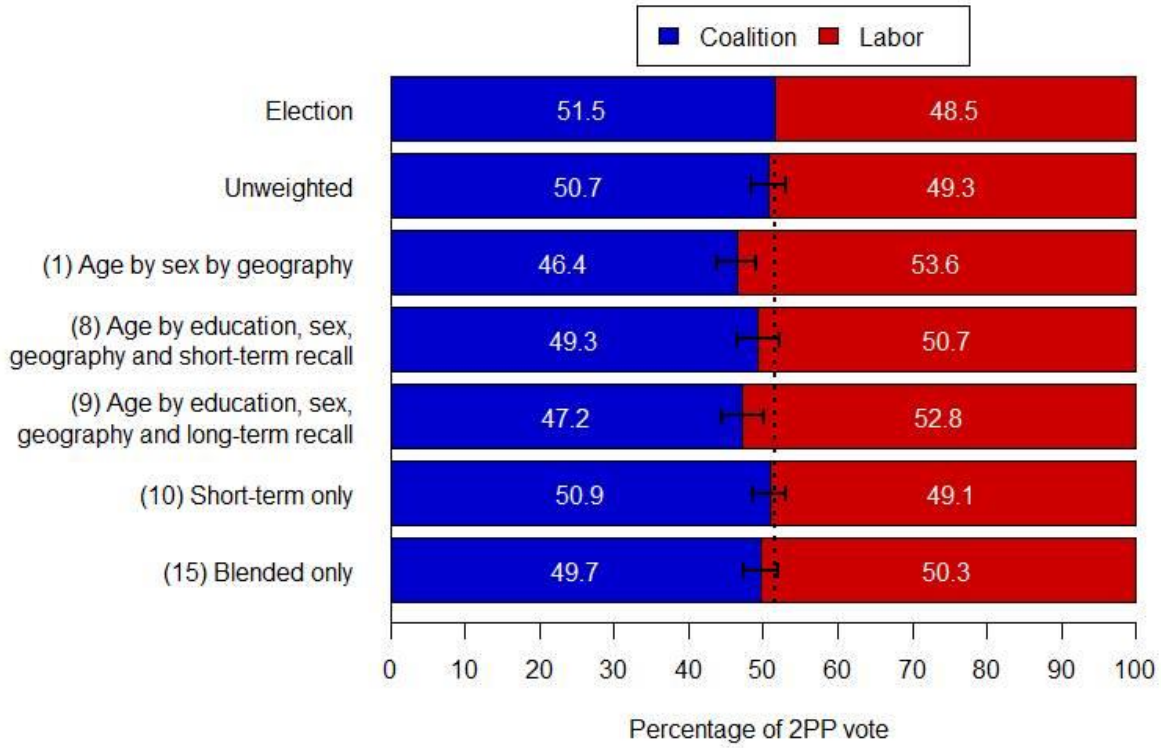
Table 7 Comparison of selected weighting solutions by estimated voted intentions

No.	Selected weighting solution	Primary vote (%)						2PP (%)	
		LNP	ALP	Greens	PHON	UAP	Other	LNP	ALP
	Election	41.4	33.3	10.4	3.1	3.4	8.2	51.5	48.5
	Unweighted	41.2	32.5	11.9	3.2	1.3	9.8	50.7	49.3
1	Age, sex and geography	36.3	33.9	15.5	2.9	1.3	10.1	46.4	53.6
8	Age by education, sex, geography and short-term recall	39.3	33.2	12.1	3.6	1.4	10.4	49.3	50.7
9	Age by education, sex, geography and long-term recall	36.8	34.2	13.3	3.6	1.6	10.4	47.2	52.8
10	Short-term only	41.4	33.5	10.2	3.4	1.4	10.2	50.9	49.1
15	Blended only	40.1	34.0	11.0	3.3	1.3	10.1	49.7	50.3

Source: Australian Election Commission and authors' analysis

Figure 3 shows the 2PP vote estimates for selected weighting solutions, with confidence intervals, enabling a comparison to the final election result. Of the weighting solutions shown, all but two were within their margin of error. Weight 1 (age, sex and geography) and weight 9 (age by education, sex, geography and long-term recall) being the exceptions.

Figure 3 Confidence intervals for the 2PP vote for selected weighting solutions, 2019



Source: Australian Election Commission and authors' analysis

6 Discussion and implications

This study shows that bias can be reduced, with only a minor increase in variance, when pre-election vote choice estimates are adjusted to match independent population benchmarks for educational attainment and past vote.

What, if any, are the implications of these findings for Australia's pollsters?

The positive impact that including an adjustment for educational attainment had on the resultant estimates of voting intentions produced from the ANUpoll, raises the question as to whether similar improvements would be seen if applied by commercial pollsters who, in the main, use non-probability online panels. These results suggest that pollsters would do well to at least consider some experimentation along these lines. Further experimentation could also be undertaken to see if any further improvement can be gained by disaggregating the educational attainment targets. The approach taken in this paper treats educational attainment as a binary variable (university graduate or not). Splitting the graduate population into graduates and post-graduates is one avenue that could be explored, as could splitting non-university graduates into those with and without non-tertiary post-secondary qualifications and/or those who did/did not successfully complete Year 12 of secondary school.

Panel proprieters have a role to play in ensuring that pollsters can avail themselves of a measure of past vote. This can be achieved if they routinely collect voting behaviour at the previous federal election as a profiling variable when recruiting new panellists and update this measure for all panellists following each election. If this is done then a measure of past vote would be available for all panel members, even though the time lag between collecting these data and the previous election may vary considerably.

While the findings in this paper are mainly applicable to pre-election polls undertaken on online panels, they may have broader applicability. This is especially the case if those pollsters not using online panels for their pre-election polls start to routinely collect a past vote measure from their respondents. If they are able to do so using a better measure of long-term recall than was available for this research, preferably one that utilises a screening question and is asked prior to the subsequent election, this may result in reduced bias in their resultant estimates of voting intentions. Some experimentation would be required.

Another possible argument against the generalisability of these findings, is that the various weighting solutions were applied to a sample (drawn from a probability-based online panel) that had very little bias to begin with in respect of the outcome measure of interest and therefore is not applicable to the real-world situation faced by commercial pollsters. However, the fact that the original voting intentions estimates produced by the ANUpoll were broadly similar to those produced by the commercial pollsters (refer back to Table 1) goes some way to refuting this argument.

One of the issues not addressed in this study is the possible impact that weighting by past vote might have on other survey estimates, in particular when these estimates could be expected to be strongly aligned with vote choice (attitudes to immigration and the environment come to mind). Peytchev and colleagues looked at this issue using 15 variables from the 2012 US General Social Survey. They found that the changes were ‘generally small but three of the fifteen estimates are significantly different. The largest change is 4%.’ (Peytchev et al. 2019, 499–500). The possible impact of weighting by past vote on other estimates being generated from Australian social surveys, presents as an interesting opportunity for further research.

7 Concluding remarks

As far as we know, this is the first Australian study to show that including an educational attainment adjustment in the weights improves the accuracy of the resultant estimates of voting intentions (while leading to only a modest increase in variance). Given that this comes on the back of strong international evidence showing similar results, it seems that weighting by educational attainment should be strongly considered by Australian pollsters. A slight caution remains, however, as the reductions in bias that we observed in this study, whilst uniform, did not meet the p<.05 threshold of statistical significance.

With regard to weighting by past vote our findings show that incorporating a short-term past vote adjustment into standard weighting solutions results in less biased estimates of vote choice than using a long-term recall measure of past vote. Including a short-term measure of past voting behaviour also results in better estimates than solutions that do not include any past vote adjustment at all. However, adding a blended or a long-term measure of past vote does not consistently produce a better outcome than weighting by age, sex, geography and education.

The results of our study reflect those of Wells (2019) in that 'how or when the (past vote) data was collected makes a difference'. In Wells' UK study the difference in the intended vote for the Labour Party was 3 percentage points (21% to 24%) depending upon which measure of past vote was used (short-term or long-term). In our study the improvement in the accuracy of the estimates resulting from the use of the short-term recall measure of past vote over the long-term measure of past vote is 2.1 percentage points (refer back to Table 7, weights 8 and 9). This is a sizeable difference in the context of pre-election polling.

Weighting by past vote is not a panacea but under the right conditions can result in substantial bias reduction. Wells observes that 'how to deal with false recall (of past-vote) used to be one of the big methodological debates within British polling. Ipsos MORI still don't use past vote weighting at all because of their concerns over false recall. In more recent years, recalled vote seemed to be closer to reality, and it has become less of an issue. But with the recent major shifts in party support it may once again become a major concern' (Wells 2019, 3).

This observation by Wells encapsulates the conventional wisdom that best approach to weighting pre-election polls is constantly changing. As noted in the 2021 AAPOR Taskforce Report 'polling error is not easily corrected using standard demographic adjustments' (AAPOR 2021, 59). On this basis, our contention is that, given the possible benefits, the appropriate use of educational attainment and past vote weighting adjustments should be an area of further consideration and exploration by Australian pollsters. Further, if such adjustments are then used to generate public domain estimates of voting intentions they should be transparently disclosed.

References

AAPOR (American Association for Public Opinion Research) (2017). AAPOR Taskforce Report, 2017. An Evaluation of 2016 Election Polls in The United States, https://www.aapor.org/Education-Resources/Reports/An-Evaluation-of-2016-Election-Polls-in-the-U-S.aspx

—— (2021). 2020 Pre-Election Polling: An Evaluation of the 2020 General Election Polls, https://www.aapor.org/Education-Resources/Reports/2020-Pre-Election-Polling-An-Evaluation-of-the-202.aspx

ABS (Australian Bureau of Statistics) (2016). 2016 Census – Employment, Income and Education, TableBuilder, HEAP – 1 Digit Level by AGEP Age, Counting: Persons Place of Usual Residence (filtered to persons aged 18 years and over).

AEC (Australian Electoral Commission) (2019a). Federal Election Timetable, https://www.aec.gov.au/Elections/Federal_Elections/2019/timetable.htm

—— (2019b). First preferences by party and Two-party preferred by State, https://results.aec.gov.au/24310/Website/HouseDownloadsMenu-24310-Csv.htm

—— (2016). 2016 Federal Election House of Representatives First Preferences By Party, https://results.aec.gov.au/20499/Website/HouseStateFirstPrefsByParty-20499-NAT.htm

Benjamini Y & Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* 57:289–300, http://www.jstor.org/stable/2346101

Biddle N (2019). *Predicting the unpredicted: what longitudinal data can tell us about the 2019 Australian federal election. Centre for Research and Methods*, CSRM Working Paper No. 6/2019, Canberra, https://csrm.cass.anu.edu.au/sites/default/files/docs/2019/9/19-081-WP-ELECTION-Publish-02_0.pdf

Cameron S & McAllister I (2019a). *The 2019 Australian federal election: results from the Australian Election Study*, Australian National University, Canberra, https://australianelectionstudy.org/wp-content/uploads/The-2019-Australian-FederalElection-Results-from-the-Australian-Election-Study.pdf

Cooper A (2013). Poll Methodology, Weighting And Adjustment System, Britsh Polling Council, http://www.britishpollingcouncil.org/wordpress/wp-content/uploads/2013/11/andrew-cooper.pdf Retrieved on 13 August, 2021

Cornesse C, Blom A, Dutwin D, Krosnick J, De Leeuw E, Legleye S, Pasek J, Pennay D, Phillips B, Sakshaug J & Struminskaya B (2020). A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research. *Journal of Survey Statistics and Methodology* 8(1): 4–36, https://doi.org/10.1093/jssam/smz041

Durand C, Deslauriers M & Valois I (2015). Should recall of previous votes be used to adjust estimates of voting intention? *Survey Insights: Methods from the Field, Weighting: Practical Issues and 'How to' Approach*, https://surveyinsights.org/?p=3543

Goot M (2021). How good are the polls? Australian election predictions, 1993–2019. *Australian Journal of Political Science* 56(1):35–55, https://doi.org/10.1080/10361146.2020.1825616

Kaczmirek L, Phillips B, Pennay D W, Lavrakas P J & Neiger D (2019). *Building a probability-based online panel: Life in Australia™*, CSRM and SRC Methods Paper No. 2/2019, Canberra, https://csrm.cass.anu.edu.au/research/publications/building-probability-based-online-panel-life-australia

Kohut A, Keeter S, Doherty C, Dimock M & Christian L (2012). Assessing the representativeness of public opinion surveys, Pew Research Center, Washington, DC, https://www.pewresearch.org/politics/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/

Korn EL & Graubard, B I (1999). *Analysis of health surveys*, John Wiley & Sons Inc, New York.

Little RJ & Vartivarian S (2005). Does weighting for nonresponse increase the variance of survey means? *Survey Methodology* 31(2):161, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.470.7842&rep=rep1&type=pdf#page=57

Lumley T (2020). Survey: analysis of complex survey samples. R package version 4.0

—— (2010). Complex Surveys: A Guide to Analysis Using R. John Wiley and Sons.

—— (2004). Analysis of complex survey samples. *Journal of Statistical Software* 9(1):1–19.

Pennay D, Goot M, Neiger D, Trewin D, Lavrakas PJ, Stirton J, Hughes P, Sheppard J & McAllister I (2020). *Report of the Inquiry into the Performance of the Opinion Polls at the 2019 Australian Federal Election,* Association of Market and Social Research Organisations and the Statistical Society of Australia, https://dataandinsights.com.au/amsro-polling-inquiry-final-report/

Pennay D, Neiger D, Lavrakas PJ & Borg K (2018). *The Online Panels Benchmarking Study: a Total Survey Error comparison of findings from probability surveys and nonprobability online panel surveys in Australia*, Centre for Research and Methods, CSRM Methods Paper No. 2/2018, Canberra, https://csrm.cass.anu.edu.au/research/publications/online-panels-benchmarking-study-total-survey-error-comparison-findings

Peytchev A, Presser S & Zhang M (2018). Improving traditional nonresponse bias adjustments: combining statistical properties with social theory. *Journal of Survey Statistics and Methodolog* 6(4):491–515, https://doi.org/10.1093/jssam/smx035

Rhodes C (2018). A Brief history of opinion polls, Museum of Australian Democracy, https://www.moadoph.gov.au/blog/a-brief-history-of-opinion-polls/

Sturgis P, Kuha J, Baker N, Callegaro M, Fisher S, Green J, Jennings W, Lauderdale B & Smith P (2017). An assessment of the causes of the errors in the 2015 UK general election opinion polls. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181(3): 757–781, https://doi.org/10.1111/rssa.12329

Sturgis P, Baker N, Callegaro M, Fisher S, Green J, Jennings W, Kuha J, Lauderdale, B & Smith P (2016). Report of the Inquiry into the 2015 British general election opinion polls, National Centre for Research Methods, University of Nottingham, Southampton, https://research.google/pubs/pub45171/

Tourangeau R, Conrad FG & Couper MP (2013). *The Science of Web Surveys*, Oxford University Press, Oxford.

Wasserstein RL, Schirm AL & Lazar NA (2019). Moving to a world beyond '$p < 0.05$'. *The American Statisticia* 73:sup1: 119, https://doi.org/10.1080/00031305.2019.1583913

Wells A (2019). False recall, and how it affects polling, YouGov, https://yougov.co.uk/topics/politics/articles-reports/2019/07/17/false-recall-and-how-it-affects-polling Retrieved on 11 January, 2021.

Appendix

Creating a 2PP estimate for the ANUpoll

As reported by Pennay et al. (2020, 38), commercial pollsters appear to use three different approaches to derive their estimates of the 2PP vote: 1) Benchmark data from previous elections; 2) stated preferences; or 3) a combination of these approaches.

Given that the ANUpoll did not ask a second preference vote choice question, the 2PP measure created for this study was based on the preference flow from the previous (2016) federal election.

Table A1 shows the source data used for these calculations and compares the preference flows from the 2016 election with the actual preference flows at the 2019 election.

Table A1 Preference flows from the previous (2016) election compared with the actual preference flows at the 2019 election

Party	Previous election		Actual preferences	
	ALP	LNP	ALP	LNP
Greens	0.819	0.181	0.822	0.178
One Nation	0.495	0.505	0.348	0.652
United Australia Party[a]	0.463	0.537	0.349	0.651
Other	0.492	0.508	0.517	0.483
Don't know[b]	0.485	0.515	0.449	0.551

a The UAP did not field candidates in the 2016 election but did so in the 2013 election so the election flow from 2013 is used for the UAP. 2016 preferences are used for all others.

b Allocated in accordance with the overall flow of preferences to each major party.

Key: ALP – Australian Labor Party; LNP – Liberal-National Party coalition.

Source: AEC (2019b, 2016)

Weighting benchmarks

Table A2 Demographic benchmarks used for weighting

Benchmark	Population %	Source
Age		
18–24 years	12.2	ABS Estimated Residential Population (ERP), March 2019 adjustment
25–34 years	19.3	
35–44 years	17.1	
45–54 years	16.5	
55–64 years	14.9	
65 or more years	20.1	
Gender		
Female	50.9	ABS ERP, March 2019 adjustment
Male	49.1	
Education		
Bachelor and above	25.5	ABS Census 2016 with ERP March 2019 adjustment
Below Bachelor	74.5	
Age by Education		
18–24	12.2	ABS Census 2016 with ERP March 2019 adjustment
25–34	7.4	
25–34 Below Bachelor	11.8	
35–44 Bachelor and above	6.2	
35–44 Below Bachelor	10.9	
45–54 Bachelor and above	4.3	
45–54 Below Bachelor	12.2	
55–64 Bachelor and above	3.3	
55–64 Below Bachelor	11.6	
65+ Bachelor and above	2.7	
65+ Below Bachelor	17.4	
Geography		
Greater Sydney	20.7	ABS Census 2016 with ERP March 2019 adjustment
Rest of NSW	11.3	
Greater Melbourne	19.8	
Rest of VIC	6.3	
Greater Brisbane	9.6	
Rest of QLD	10.2	

Table A2 Demographic benchmarks used for weighting (cont.)

Benchmark	Population %	Source
Greater Adelaide	5.5	
Rest of SA	1.6	
Greater Perth	8.1	
Rest of WA	2.2	
Greater Hobart	0.9	
Rest of TAS	1.2	
Greater Darwin	0.6	
Rest of NT	0.4	
Australian Capital Territory	1.7	

Key: NT – Northern Territory; SA – South Australia; TAS – Tasmania; WA – Western Australia

Table A3 Voting benchmarks from the 2019 election and prorated weighting targets

Primary vote at the 2016 election	AEC[a]	Prorated short-term recall target	Prorated long-term recall target	Prorated blended recall target
Coalition	42.0	33.5	38.7	36.1
Australian Labor Party	34.7	27.7	31.9	29.8
The Greens	10.2	8.2	9.4	8.8
Other	13.0	10.4	11.9	11.2
Not enrolled, did note vote, voted informally	–	20.3	8.1	14.1

Source: (AEC 2016)

Table A4 Probabilities from compairsons of weighting schemes for root mean sqaured error of 2PP vote

Weighting solution	0 – Uwtd	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 – A,S,G	**↑0.031**														
2 – A,S,G,E	**↑0.048**	0.392													
3 – AxE,S,G	0.075	0.075	0.111												
4 – A,S,G, ST	0.173	**↓0.027**	0.126	0.263											
5 – A,S,G, LT	**↑0.022**	0.831	0.864	0.546	**↑0.047**										
6 - A,S,G,E, ST	0.208	**↓0.031**	0.067	0.183	0.831	**↓0.043**									
7 - A,S,G,E, LT	**↑0.027**	0.798	0.878	0.604	0.086	0.864	**↑0.039**								
8 - AxE,S,G, ST	0.311	**↓0.022**	**↓0.031**	0.075	0.235	**↓0.027**	0.183	**↓0.027**							
9 - AxE,S,G, LT	**↑0.031**	0.510	0.864	0.848	0.164	0.274	0.092	0.144	**↑0.037**						
10 – ST only	0.433	**↓0.022**	**↓0.027**	**↓0.031**	0.087	**↓0.017**	0.084	**↓0.019**	0.106	**↓0.026**					
11 – LT only	0.161	**↓0.039**	0.133	0.241	0.850	**↓0.014**	0.924	**↓0.027**	0.864	**↓0.037**	0.084				
12 – A,S,G, Bl	**↑0.037**	0.409	0.864	0.864	**↑0.037**	0.460	0.085	0.659	**↑0.027**	0.942	**↑0.027**	0.094			
13 – A,S,G,E, Bl	**↑0.037**	0.460	0.848	0.864	0.098	0.460	**↑0.039**	0.460	**↑0.027**	0.906	**↑0.027**	0.118	0.963		
14 – AxE,S,G, Bl	0.060	0.215	0.496	0.848	0.235	0.183	0.161	0.215	**↑0.042**	0.433	**↑0.033**	0.241	0.285	0.161	
15 – Bl only	0.604	**↓0.012**	**↓0.027**	**↓0.039**	0.183	**↓0.000**	0.285	**↓0.008**	0.460	**↓0.008**	0.241	0.084	**↓0.017**	**↓0.019**	**↓0.027**

Key: A – Age, S – Sex, G – Geography, E – Education, ST – Short-term, LT – Long-term, Bl – Blended, Uwtd – Unwieghted. ↑ Row is significantly higher than column, ↓ Row is significantly lower than column.
Source: Authors' analysis