The effect of survey administration mode on youth mental health measures: Social desirability bias and sensitive questions

Debra J. Rickwood [*], Cassandra L. Coleman-Rose

Faculty of Health, University of Canberra, Australia

ABSTRACT

Aim: Research on trends in youth mental health is used to inform government policy and service funding decisions. It often uses interviewer-administered surveys, which may be affected by mode effects related to social desirability bias. This study sought to determine the impact of survey administration mode on mental health measures, comparing mode effects for sensitive mental health measures (psychological distress and wellbeing) and non-sensitive (physical activity) measures.

Methods: Data were from two large national community samples of young Australians aged 12–25 years conducted in 2020 (N = 6238) and 2022 (N = 4122), which used both interviewer-administered and self-report modes of data collection.

Results: Results showed participants reported lower psychological distress and higher wellbeing in the interviewer-assisted compared with the self-report mode. No mode effects were found for the non-sensitive physical activity measures. No interaction between mode and gender was found, but an age group by mode interaction revealed that those in the 18–21 and 22–25-year age groups were more strongly affected than younger adolescents.

Conclusions: These findings suggest underestimates of mental health issues from interview survey formats, particularly for young adults. The results show how even a weak mode effect can have a large impact on mental health prevalence indicators. Researchers and policy makers need to be aware of the impact social desirability bias can have on mental health measures and consider taking steps to mitigate this effect.

1. Introduction

There is a critical need for accurate population estimates of mental health status. Most researchers and practitioners working in the mental health field argue that mental health problems are on the rise, particularly for young people; adolescent mental health is claimed to be in crisis [1]. Prevalence data are used to advocate for increased funding and service provision in areas of identified need. Given the limited health funding available, and the many health conditions that warrant greater expenditure, a solid evidence base to guide targeted initiatives is essential for policy and funding decisions.

The World Health Organization reports that the incidence of mental health problems is growing worldwide, particularly for young people [2]. Defined as those aged between 12 and 25 years, and encompassing the life stages of early adolescence and emerging adulthood, young people are a critical focus for greater mental healthcare investment in Australia and internationally [3]. Young

Australians are the most impacted by mental disorders, with almost two in five people aged 16–24 years experiencing symptoms of a mental disorder in the 12 months prior to a 2020–21 survey, more than any other age bracket [4]. Given the high personal and community cost of mental ill-health, especially for youth, there is a clear priority for governments to monitor the psychological distress and mental wellbeing of young Australians.

Funded by the Australian Government, the Australian Bureau of Statistics collects national mental health data to provide an overview of Australia's mental health to inform health policies and allocation of resources [5]. The national mental health surveys are collected through face-to-face interviews [6]. Yet, research increasingly shows that interviewer-administered modes of collecting sensitive data may underestimate the real prevalence of mental ill-health due to mode effects attributed to social desirability bias [7–9].

For example, when examining psychological distress levels Klein et al. [7] found that adult respondents in an online self-report survey were over four times more likely to be in a high or very high psychological distress category than respondents in National Health Surveys conducted by face-to-face interviews, despite both groups responding to the same questions. The authors argue that the level of psychological distress in Australians may have been underestimated in previous national studies relying on interviewer-administered modes.

1.1. Social desirability bias and sensitive questions

Research indicates that people completing interviewer-administered questionnaires are more likely to provide socially desirable responses than those completing self-administered questionnaires [8,10–12]. Social desirability bias is the tendency to under-report socially undesirable attitudes and behaviours and over-report more desirable attributes [13]. This may happen for two reasons: first, for impression management, which is the deliberate presentation of self to conform to an audience's normative expectations; and second, due to self-deception, which is based on motivation to maintain a positive self-concept that may be unconscious. A major review concluded that social desirability bias is generally motivated by impression management, particularly the desire to avoid embarrassment and the consequences from disclosing sensitive information [14].

Consequently, social desirability bias has a greater impact on sensitive questions, particularly those that are related to, or perceived to be related to, social norms [15]. Sensitive questions include mental health measures, as mental health is a socially relevant construct with strong normative expectations and where deviation from these norms incurs social stigma [16]. Consequently, the presence of an interviewer when these questions are asked may have a significant impact, whereby participants in an interviewer-administered survey mode will be motivated to comply more closely with social norms, which include appearing to be mentally healthy. Prevalence rates of both mental ill-health [7] and mental wellbeing [9,17] have been shown to be affected by mode of administration in this way. For example, when comparing telephone interview responses with online self-report for questions relating to the mental health of young Australians, Milton et al. [8] found that 18.4% of online self-report respondents reported having thoughts of suicidal ideation in the past 12 months, compared to 11.1% of telephone interview respondents.

A widely used indicator of population mental health status is the Kessler Psychological Distress Scale, or K10. This is the key mental health measure for Australia's national mental health and wellbeing surveys and national health surveys and is regularly reported in government reports on health and mental health (e.g., *Australia's Health 2022* [18]). Research by Klein et al. [7] shows that in interviews respondents are more likely to provide socially desirable responses to the K10, with online self-administered surveys showing much higher rates of psychological distress than respondents in face-to-face interviews. The authors maintain that their research was the first time that the K10 had been administered anonymously online to a national sample in Australia, suggesting that previous research relying on the K10 measure in interviewer-administered modes may have significantly underestimated levels of psychological distress in the population. However, this study had important confounds, with the online and face-to-face responses collected at different time periods and by different organisations: online responses were from 2018 while face-to-face responses were from 2014 to 2015 and 2017–2018, thereby confounding time and mode.

Positive mental health measures, where higher scores indicate greater wellbeing, also appear to be impacted by social desirability bias. A study relating to mode differences in a health survey of adults in Germany found that respondents in the interviewer-administered telephone survey were more likely to report better mental and psychosocial health than those who responded in self-administrated modes who were more likely to report poor mental health [17]. Similarly, Zager Kocjan et al. [9] examined mode effects on the short form of the Mental Health Continuum (MHC-SF) national health data in Slovenia, examining emotional, social and psychological wellbeing. They found that participants were more likely to report better wellbeing when completing face-to-face interviews than when completing online self-report surveys. Both Klein et al. [7] and Zager Kocjan et al. [9] suggest that previous research relying on the K10 and MHC-SF measures in interviewer-administered modes may have significantly underestimated levels of psychological distress and overestimated psychological wellbeing in the population.

1.2. Non-sensitive questions: physical activity

Non-sensitive questions that are not affected by social desirability bias are those that do not have a clear normative element. Questions regarding factual behavioural information with simple categorical answers are thought to be least likely to be influenced by social desirability bias and mode effects [19]. Although there is little research on precisely what types of questions are not impacted, some evidence shows that social desirability does not affect questions relating to physical activity [13,20,21].

Crutzen and Goeritz [13] conducted a longitudinal study on a representative sample of the Dutch adult population examining whether social desirability bias was associated with self-reported physical activity levels. They found no relationship between scores

on the Marlow-Crowne Social Desirability Scale and the short form of the International Physical Activity Questionnaire (IPAQ). Chu et al. [20] found comparable results for physical activity measured by self and interviewer-administration, against externally validated physical activity measurement data in a sample of adults in Singapore. Similar findings came from a study examining physical activity measurement in fifth grade students in the United States, with both interviewer and self-administered reports showing comparable correlations with data from a measurement device [21].

1.3. Age and gender effects

Whether mode effects are influenced by age or gender is currently mostly unknown, although there is a small body of emerging research showing that young adults and males may be most likely to be affected. Klein et al. [7] found a significant interaction between age and mode, with those in the 18–24 year age bracket more likely to report poorer mental health in anonymous online self-report surveys compared with face-to-face surveys than those in older age brackets. This study did not include people aged under 18 years, however, and had only a small number of participants in the 18–24 year age bracket. A study of health status and emotional distress in Swedish adolescents found a small mode effect between self-report postal surveys and telephone surveys [22]. While the older age groups reported higher anxiety and depression in the postal survey compared with over the telephone, the opposite was true for the 13–15 year-olds, who reported lower anxiety and depression in the postal survey compared to the telephone. The mode effect was reversed in early adolescence. Consequently, the conclusion that younger adults are likely to be more affected is tentative.

Younger adult males were shown to be more likely to provide socially desirable responses to questions when there was an interviewer present than when not [23]. In this study of French adults, males aged between 18 and 29 years reported higher levels of drug use for most drugs in a self-administered survey than were reported in an interviewer-administered survey. Adolescent males were found to be more prone to social desirability effects than adolescent females in a comparison of audio-assisted self-administrated interviewers and face-to-face interviews regarding reporting of sexual behaviours in rural Malawi [24]. Males were more likely to report higher sexual activity, but less likely to report stigmatised sexual relationships, in face-to-face interviews than in the self-administered mode. In comparison, there was no significant difference in reported general sexual behaviours between the two modes for females, although they were also slightly more likely to report stigmatised sexual relationships in the self-administered mode.

1.4. The current study

The current study sought to determine the effect of survey mode, specifically computer assisted phone interview versus online self-report, on mental health measures for young people in Australia using two large, recent, national community surveys. This study furthers knowledge in four important ways. First, it has a specific focus on young people in the teen and early adult years, which is a life-stage of critical interest in the mental health field and for which accurate prevalence data are essential. Second, we address limitations of previous studies, such as by Klein et al. [7], in that our data are collected at the same time period and by the same organisation for both modes. Third, we include a non-sensitive comparator question, namely physical activity, which should not show mode effects. Finally, we replicate our findings through two large community surveys conducted two years apart.

We hypothesised that:

1) survey mode would affect measures according to the sensitivity of the measure—sensitive measures (mental health measures) would demonstrate survey mode differences, whereby the interviewer-administered measures would differ from the self-report measures, whereas the non-sensitive measures (physical activity) would not show a mode effect;
2) for the sensitive mental health measures, the interviewer-administered mode was expected to show lower mental ill-health and higher mental wellbeing scores compared with the self-report mode;

Table 1
Number of participants in each mode by gender and age group, for study one and two.

Mode	Gender	Age Group				
		12–14	15–17	18–21	22–25	Total
Study One (2020)						
CATI	Females	120	121	134	132	507
	Males	136	134	120	122	512
	Total	256	255	254	254	1019
Online	Females	623	624	656	672	2575
	Males	652	644	640	648	2584
	Total	1275	1268	1296	1320	5159
Study Two (2022)						
CATI	Females	124	115	127	125	491
	Males	128	118	131	128	505
	Total	252	233	258	253	996
Online	Females	364	365	373	370	1472
	Males	388	392	392	405	1577
	Total	752	757	765	775	3049

3) mode effects for the sensitive mental health measures were expected to be most pronounced for males and those in the young adult age group compared with other age/gender groups; and

4) findings would be consistent over the two survey collections (2020 and 2022).

2. Method

2.1. Samples and participants

Participants were from national community samples of Australian young people aged 12–25 years. Study one, conducted in May–June 2020, had 5203 participants complete the survey in an on-line self-report mode and 1035 participants through a Computer Assisted Telephone Interview (CATI) mode. Study two, conducted in August–September 2022, had 3107 participants complete the survey online and 1015 through a CATI. Demographics are provided in Table 1.

Study one participants comprised 3096 males (49.6%) and 3082 females (49.4%) with 60 people (1.0%) identifying another gender or choosing not to respond. Participants were evenly spread across the age groupings: 12–14 ($n = 1540$), 15–17 ($n = 1546$), 18–21 ($n = 1564$), and 22–25 ($n = 1588$). Due to the stratified sample, gender and age group were equivalent across the CATI and online survey groups. There was also equivalent representation across the CATI and online survey groups for participants who were First Nations (n = 65, 6.3%; n = 307; 6.0%) and who lived in non-metropolitan areas (n = 278, 26.9%; n = 1359, 26.1%); but not for those not born in Australia, where there was lower representation in the CATI sample than the online survey sample (n = 82, 7.9%; n = 1170, 22.5%, respectively).

Study two participants were 2075 males (50.3%) and 1970 females (47.8%) with 77 (1.9%) identifying another gender or choosing not to respond. Again, participants were evenly spread across the age groupings: 12–14 ($n = 1025$), 15–17 ($n = 1001$), 18–21 ($n = 1050$), and 22–25 ($n = 1046$). Again, due to the stratified sample, gender and age group were equivalent across the CATI and online survey groups. There was also equivalent representation across the CATI and online survey groups for participants who were First Nations (n = 43, 4.3%; n = 129; 4.2%); who lived in non-metropolitan areas (n = 263, 25.9%; n = 884, 28.5%); and who were not born in Australia (n = 83, 8.2%; n = 241, 8.2%).

2.2. Procedure

Data came from the National Youth Mental Health Surveys commissioned by headspace Australia's National Youth Mental Health Foundation and conducted by research organisation Kantar. Participants were recruited by Kantar using quota sampling, with quotas set for age, gender and state/territory. CATI participants comprised 60% recruited from an existing market research panel and 40% through random digit dialling. Online participants were recruited through an online research panel of over 400,000 Australians aged over 18 years. Participants under 18 were recruited through their parent panel members. Importantly, participants were not randomly assigned to the CATI or online survey administration conditions, although the same panels were used, but the CATI also entailed random digit dialling. Participants provided informed consent, and consent was obtained from parents for those under 18 years. Participants received a very small compensation for their time.

The CATI participants were contacted by phone. Interviews were conducted by an interviewer over the phone, guided by a questionnaire displayed on a computer screen. The interviewer records answers using the pre-coded responses displayed on the screen. Any questionnaire routing or complex survey logic is handled by the CATI program. Interviews took an average of 30 min in 2020 and 10 min in 2021 (there were substantially fewer questions asked in 2022). Online participants were contacted online via email or text, and in both 2020 and 2022 the average completion time was 30 min.

2.3. Measures

Both surveys comprised a large number of questions related to young people's mental health and wellbeing. Of interest to the current study, psychological distress and wellbeing were assessed in both surveys for both administration modes and are categorised as sensitive question measures. Physical activity is the non-sensitive measure type, which was also assessed in both surveys for both modes, but by different questionnaires in each survey. Demographics were also obtained.

Psychological distress was measured using the Kessler Psychological Distress Scale (K10) [25], which comprises 10 questions asking respondents how often they felt particular ways, including "worthless" and "hopeless", in the last four weeks. Total scores can range from 10 to 50; scores of 10–15 indicate low, 16–21 moderate, 22–29 high, and 30–50 very high distress [26].

Wellbeing was measured using the Mental Health Continuum - Short Form (MHC-SF [27]; which has 14 items measuring three facets of wellbeing: emotional, social, and psychological. Participants were asked to indicate how often they had experienced a particular feeling in the last month (e.g., happy, satisfied with life). Total scores could range between 0 and 70, with higher scores indicating greater wellbeing.

Physical activity was measured in study one by two items asking about physical activity in the context of being impacted by COVID-19. Participants were asked whether COVID-19 had a positive or negative impact on (1) exercise and or physical activity, and (2) participation in sport. Scores for the two items were summed and the total could range from 2 to 10, with scores above the midpoint indicating a positive impact. In study two, the physical activity measure was adapted from the International Physical Activity Questionnaire (IPAQ) short form [28] included the items measuring (1) vigorous activity, (2) walking, and (3) sitting over the last seven days. Higher scores indicate more of the relative activity. The IPAQ short form has been found to have high reliability and

validity [28], including for adolescents [29].

2.4. Data analysis

IBM SPSS Version 28 was used for all analyses. Data were weighted by gender and state/territory within age groups. The 60 and 77 participants in studies one and two, respectively, who were non-binary or did not report gender were not included in analyses involving gender due to insufficient sample size.

To analyse psychological distress and wellbeing, two $2 \times 2 \times 2 \times 4$ factorial analysis of variance (ANOVA) were conducted to compare the main effects of study (2022, 2020), survey mode (CATI, online), gender (male, female), and age group (12–14, 15–17, 18–21, 22–25 years) and the interactions between these variables. Because different physical activity measures were used in each study, these were analysed separately. For study one, a $2 \times 2 \times 4$ factorial ANOVA was conducted for the physical activity measure, with the factors survey mode, gender, and age group. For study two, a multivariate ANOVA was conducted with the same factors, but with three dependent variables for the three physical activity questions (vigorous activity, walking, sitting).

The K10 scores were also analysed categorically because this is how psychological distress is most often reported [30]. Pearson's chi-squared test of contingencies was used to examine whether the mode of administration was related to the K10 psychological distress categories, separately for each study year. Separate analyses by year were undertaken and gender and age group were not included to simplify the results (given these effects were revealed through the ANOVAs) and demonstrate the impact of mode on percentages in each K10 category.

Due to the very high power, a stringent alpha value of 0.001 was set, with a cut-off for effect sizes of less than 0.001 to exclude trivial effects.

3. Results

Means and standard deviations for each measure by the factorial design are presented in Tables 2 (2020 survey) and 3 (2022 survey).

3.1. Psychological distress (K10)

The ANOVA of K10 scores revealed that all four main effects and the study by mode and mode by age group two-way interaction effects were significant (Table 4). The interaction of study year and survey mode showed a very weak effect, whereby the online scores were significantly higher than the CATI scores in 2020 but, while the difference was in the same direction, it did not attain significance in 2022. The mode by age group interaction was stronger, with post-hoc tests showing that the mode effect was not significant for the 12–14 and 15–17 year age groups, but was significant for the 18–21 and 22–25 year age groups, with online scores being higher than CATI scores as shown in Fig. 1.

3.2. Psychological distress categories

For the categorical K10 scores, in study one, the chi-squared test was statistically significant, $\chi^2 (3, N = 6239) = 75.89, p < .001,$

Table 2
Study year 2020 - Mean scores for psychological distress, wellbeing, and physical activity by study year, survey mode, gender and age group.

Measure	Mode	Gender	Age Group				
			12–14	15–17	18–21	22–25	Total
			M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Psychological Distress (K10)	CATI	Females	19.5 (6.6)	21.8 (7.6)	21.5 (6.9)	21.0 (6.9)	21.0 (7.0)
		Males	17.7 (5.2)	19.0 (6.2)	18.3 (6.3)	20.3 (6.6)	18.8 (6.1)
		Total	18.6 (5.9)	20.3 (7.0)	19.9 (6.8)	20.6 (6.7)	19.9 (6.7)
	Online	Females	19.6 (7.0)	20.3 (8.1)	25.2 (7.8)	24.2 (7.9)	22.4 (8.2)
		Males	19.3 (7.3)	18.9 (7.6)	23.9 (8.0)	24.0 (7.6)	21.5 (8.0)
		Total	19.4 (7.2)	19.5 (7.9)	24.5 (7.9)	24.1 (7.8)	21.9 (8.1)
Wellbeing (MHC-SF)	CATI	Females	53.2 (10.8)	50.5 (12.7)	47.6 (13.5)	48.9 (10.7)	50.1 (12.1)
		Males	56.1 (8.8)	52.2 (12.1)	51.2 (12.3)	47.5 (13.4)	51.8 (12.1)
		Total	54.7 (9.9)	51.4 (12.4)	49.4 (13.0)	48.2 (12.1)	50.9 (12.1)
	Online	Females	47.5 (13.6)	45.2 (14.4)	38.9 (14.0)	39.3 (15.0)	42.7 (14.7)
		Males	48.5 (12.8)	47.4 (13.4)	42.5 (14.4)	41.6 (14.2)	45.0 (14.0)
		Total	48.0 (13.2)	46.3 (13.9)	40.8 (14.3)	40.4 (14.6)	43.9 (14.4)
Physical Activity	CATI	Females	5.3 (2.2)	5.2 (2.0)	5.0 (2.0)	4.8 (2.0)	5.1 (2.0)
		Males	5.6 (2.0)	5.1 (2.2)	4.8 (2.0)	4.8 (2.1)	5.1 (2.1)
		Total	5.5 (2.1)	5.1 (2.1)	4.9 (2.0)	4.8 (2.0)	5.1 (2.1)
	Online	Females	4.9 (2.1)	4.9 (1.9)	5.4 (2.0)	5.1 (2.0)	5.1 (2.0)
		Males	4.9 (2.0)	4.8 (2.0)	5.5 (2.1)	5.4 (2.3)	5.2 (2.1)
		Total	4.9 (2.0)	4.9 (2.0)	5.5 (2.1)	5.3 (2.1)	5.1 (2.1)

Table 3
Study year 2022 - Mean scores for psychological distress, wellbeing, and physical activity measures by survey mode, survey mode, gender and age group.

Measure	Mode	Gender	Age group				
			12–14	15–17	18–21	22–25	Total
			M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Psychological distress (K10)	CATI	Females	22.2 (7.7)	23.4 (7.5)	24.4 (8.0)	22.5 (6.9)	23.1 (7.6)
		Males	18.8 (5.1)	19.3 (5.9)	20.1 (6.3)	21.0 (6.9)	19.8 (6.1)
		Total	20.4 (6.7)	21.3 (7.0)	22.2 (7.5)	21.7 (6.9)	21.4 (7.1)
	Online	Females	20.2 (7.6)	20.0 (7.5)	26.7 (8.0)	24.1 (7.7)	22.8 (8.2)
		Males	18.5 (6.8)	18.3 (7.0)	24.4 (8.4)	22.9 (8.2)	21.0 (8.1)
		Total	19.3 (7.2)	19.1 (7.3)	25.5 (8.2)	23.5 (8.0)	21.9 (8.2)
Wellbeing (MHC-SF)	CATI	Females	48.2 (12.4)	46.9 (11.4)	44.4 (12.9)	46.7 (12.3)	46.5 (12.3)
		Males	55.7 (9.1)	50.4 (10.8)	50.9 (11.2)	46.7 (13.5)	51.0 (11.7)
		Total	52.1 (11.7)	48.7 (11.2)	47.7 (12.5)	46.7 (12.9)	48.8 (12.2)
	Online	Females	47.1 (14.2)	45.3 (14.7)	38.0 (15.1)	40.1 (14.8)	42.6 (15.1)
		Males	49.5 (12.5)	45.8 (13.9)	41.7 (14.3)	42.8 (14.4)	44.9 (14.1)
		Total	48.3 (13.4)	45.6 (14.3)	39.9 (14.8)	41.5 (14.7)	43.8 (14.7)
Physical Activity – Vigorous	CATI	Females	239.9 (234.4)	256.0 (248.2)	247.4 (271.7)	234.0 (264.5)	244.3 (253.3)
		Males	305.7 (265.1)	327.5 (312.4)	469.3 (389.8)	322.4 (323.2)	356.1 (330.3)
		Total	275.1 (252.2)	293.4 (284.5)	365.9 (356.3)	282.0 (300.1)	304.0 (301.5)
	Online	Females	225.2 (274.8)	204.1 (284.0)	256.3 (334.8)	202.8 (282.3)	222.0 (295.6)
		Males	268.5 (306.5)	281.4 (330.3)	370.7 (358.0)	324.7 (346.0)	311.5 (337.9)
		Total	247.6 (292.3)	244.4 (311.7)	314.8 (351.3)	266.4 (323.6)	268.5 (321.7)
Physical Activity – Walking	CATI	Females	332.6 (377.3)	354.5 (335.5)	456.0 (413.7)	356.6 (387.2)	374.6 (381.2)
		Males	356.4 (352.5)	385.6 (375.5)	506.1 (451.6)	482.9 (464.6)	431.0 (416.0)
		Total	345.3 (363.8)	370.8 (356.0)	482.7 (433.9)	425.2 (434.2)	404.7 (400.7)
	Online	Females	305.0 (337.0)	316.6 (360.3)	431.0 (427.3)	385.3 (411.7)	360.1 (389.2)
		Males	328.5 (381.1)	317.5 (350.7)	478.2 (451.8)	421.6 (416.6)	386.8 (407.1)
		Total	317.1 (360.4)	317.1 (355.2)	455.2 (440.3)	403.8 (414.4)	373.8 (398.)
Physical Activity – Sitting	CATI	Females	371.3 (279.7)	478.6 (288.9)	428.6 (285.0)	393.5 (189.0)	416.7 (267.8)
		Males	398.8 (295.0)	407.7 (274.2)	366.2 (249.6)	428.8 (239.7)	399.7 (266.5)
		Total	386.0 (287.6)	441.5 (282.9)	395.3 (268.0)	412.7 (217.9)	407.6 (267.1)
	Online	Females	367.0 (231.5)	433.5 (269.1)	426.5 (293.6)	435.2 (276.3)	415.8 (270.0)
		Males	391.3 (233.8)	404.4 (240.0)	373.0 (248.5)	365.6 (281.5)	383.5 (252.2)
		Total	379.5 (232.9)	418.5 (254.7)	399.0 (272.6)	399.7 (281.0)	399.2 (261.4)

Table 4
Main and interaction effects for study, survey mode, gender, and age group on K10 scores.

Variable	F	df	p	η_p^2
Study	**18.18**	**1, 10191**	**<.001**	**.002**
Survey Mode	**43.35**	**1, 10191**	**<.001**	**.004**
Gender	**113.49**	**1, 10191**	**<.001**	**.011**
Age Group	**89.05**	**3, 10191**	**<.001**	**.026**
Study x Mode	**17.58**	**1, 10191**	**<.001**	**.002**
Study x Gender	7.61	1, 10191	.006	.001
Study x Age Group	3.00	3, 10191	.029	.001
Mode x Age Group	**45.27**	**3, 10191**	**<.001**	**.013**
Gender x Age Group	4.97	3, 10191	.002	.001
Study x Mode x Gender	0.10	1, 10191	.755	.000
Study x Mode x Age Group	0.18	3, 10191	.912	.000
Study x Gender x Age Group	0.16	3, 10191	.925	.000
Mode x Gender x Age Group	1.05	3, 10191	.368	.000
Study x Mode x Gender x Age Group	0.07	3, 10191	.975	.000

Note. Significant effects where $p < .001$ and $\eta_p^2 > 0.001$ are in bold.

although the association was weak, with Cramer's $V = 0.11$. Participants in the online condition were more likely to have high levels of distress. CATI respondents reported higher percentages of low (28.9%) and moderate (37.3%) psychological distress compared to online respondents (24.6% and 28.4%, respectively). Online respondents reported higher percentages of high (28.0%) and very high (19.0%) psychological distress than CATI respondents (24.1% and 9.8%, respectively).

In study two, the chi-squared test was also statistically significant $\chi2$ (3, $N = 4122$) = 50.67, $p < .001$ with a weak association (Cramer's $V = 0.11$). Those in the online condition were more likely to have very high (19.4%) and low (25.7%) levels of psychological distress in comparison to those in the CATI condition (14.8% and 19.6%, respectively). Those in the CATI condition were more likely to have moderate (38.3%) levels of psychological distress than those in the online condition (27.4%). Respondents in the online and CATI conditions reported similar levels of high psychological distress (27.6% and 27.3%, respectively).
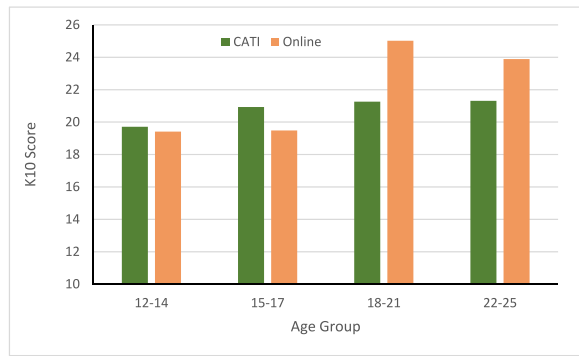
Fig. 1. Mean K10 scores by age group and survey mode.

The mode effect was particularly strong for the older age-groups, with a mode by age group interaction (see Table 4). Fig. 2 illustrates the nature of this effect, showing the percentages only for those aged 18–25. This clearly reveals the consistent pattern over both 2020 and 2022 for the young adults whereby psychological distress scores were higher in the online self-report mode.

3.3. Wellbeing (MHC-SF)

The ANOVA of MHC-SF scores revealed that all four main effects and the mode by age group two-way interaction were significant (Table 5). Post-hoc tests for the mode by age group interaction showed that the mode effect was not significant for the 12–14 and 15–17 year age groups, but was significant for the 18–21 and 22–25 year age groups, although the difference was in the same direction for all age groups with CATI scores higher than online scores, as shown in Fig. 3.

3.4. Physical activity

The factorial ANOVA for physical activity in study one indicated no significant main or interaction effects (see Table 6). In study two, the multivariate ANOVA for physical activity revealed main effects for gender and age group, but no other significant effects (Table 7). Of note, none of the survey mode effects was significant.

4. Discussion

Overall, our results demonstrate a clear mode effect for young adults for sensitive mental health-related measures, such that lower mental ill-health and higher mental wellbeing are reported in a computer assisted interview than an online self-report survey.
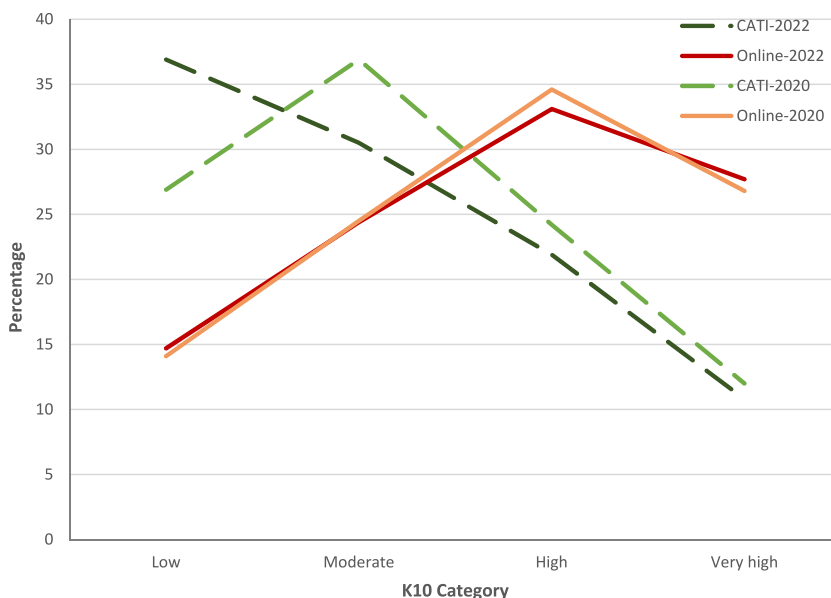

Fig. 2. of respondents aged 18–25 Years in each K10 category by mode and study.

Table 5

Main and interaction effects for study, survey mode, gender, and age group on MHC-SF scores.

Variable	F	df	p	η_p^2
Study	**11.11**	**1, 10191**	**<.001**	**.001**
Survey Mode	**309.27**	**1, 10191**	**<.001**	**.029**
Gender	**59.51**	**1, 10191**	**<.001**	**.006**
Age Group	**83.25**	**3, 10191**	**<.001**	**.024**
Study x Mode	8.97	1, 10191	.003	.001
Study x Gender	4.41	1, 10191	.036	.000
Study x Age Group	0.71	3, 10191	.548	.000
Mode x Gender	1.00	1, 10191	.318	.000
Mode x Age Group	**7.11**	**3, 10191**	**<.001**	**.002**
Gender x Age Group	5.26	3, 10191	.001	.002
Study x Mode x Gender	3.83	1, 10191	.050	.000
Study x Mode x Age Group	0.50	3, 10191	.682	.000
Study x Gender x Age Group	0.74	3, 10191	.526	.000
Mode x Gender x Age Group	4.81	3, 10191	.002	.001
Study x Mode x Gender x Age Group	0.10	3, 10191	.958	.000

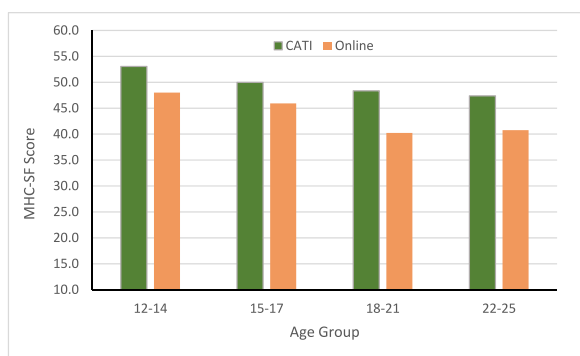*Note.* Significant effects where $p < .001$ and $\eta_p^2 > 0.001$ are in bold.

Fig. 3. Mean MHC-SF scores by age group and survey mode.

Table 6

Main and interaction effects for survey mode, gender, and age group on physical activity in study one (2020).

Variable	F	df	p	η_p^2
Survey Mode	0.30	1, 5671	.586	.000
Gender	1.12	1, 5671	.291	.000
Age Group	2.18	3, 5671	.089	.001
Survey Mode*Gender	0.44	1, 5671	.505	.000
Survey Mode * Age Group	13.02	3, 5671	<.001	.007
Gender*Age Group	0.93	3, 5671	.427	.000
Survey Mode*Gender*Age Group	1.32	3, 5671	.265	.001

Note. Significant effects where $p < .001$ and $\eta_p^2 > 0.01$ are in bold.

Table 7

Main and interaction effects for survey mode, gender, and age group on physical activity measures in study two (2022).

Variable	F	df	p	η_p^2
Survey Mode	2.92	3, 3844	.033	.002
Gender	**14.06**	**3, 3844**	**<.001**	**.011**
Age Group	7.03	9, 11538	<.001	.005
Survey Mode x Gender	0.55	3, 3844	.649	.000
Survey Mode x Age Group	0.63	9, 11538	.769	.000
Gender x Age Group	2.11	9, 11538	.025	.002
Survey Mode x Gender x Age Group	1.29	9, 11538	.235	.001

Note. Significant effects where $p < .001$ and $\eta_p^2 > 0.01$ are in bold.

Although both types of survey administration were anonymous, in the interview mode the participant is disclosing their responses to another person, which is likely to invoke social desirability biases. The online survey does not have this social disclosure element and, so, is not as prone to social desirability bias, arguably allowing young people to more truthfully disclose their mental health status. Prior research with young people has shown that in general they prefer to 'type than talk' [31] and that using an online format leads to greater disclosure of sensitive personal information, such as sexuality and alcohol and other drug use [32]. Our results suggest that this effect is also evident in population surveys.

The mode effect was moderated by age, and only evident for the two age groups over 18 years. The effect was in the same direction, but not significant, for the adolescents. No prior research has considered such effects for younger adolescents to our knowledge. Studies of adult populations by Klein et al. [7] and Beck et al. [23], reported that their youngest adult age brackets (aged 18–24 years and aged 18–29 years, respectively) were the most impacted by mode effects. Our findings, including younger age groups, suggest there may be something unique about the young adult years that maximises mode effects. That young adults have the worst mental health, evidenced in the current study by their highest mental ill-health and lowest wellbeing scores, and as generally found in population studies (e.g., Ref. [4]), suggests that poorer mental health status may make such questions particularly sensitive for participants at this age and, thereby, exacerbate mode differences related to questionnaire sensitivity.

Nevertheless, mode effects were not moderated by gender, and females had poorer mental health than males. However, it could be argued that females may find mental health questionnaires less sensitive than males. This was not investigated in the current study, but has been demonstrated in research showing that females have less stigmatising attitudes towards mental illness [33], and are more comfortable disclosing their mental health and more emotionally expressive [34]. Consequently, it is likely that mental health questionnaires are less sensitive for females than for males. Nevertheless, males show more stigma towards mental illness than females [35], and could be argued to be more motivated to appear mentally healthy. It may be that the interaction of prevalence and stigma cancel each other to negate gender differences, a question for future research to address.

Although the effects were quite weak statistically when analysing the variance around mean scores, the considerable impact of such small effects on population health estimates is evident in Fig. 1. Analysis of the categorical K10 scores revealed substantial differences in the proportion of young people within each category according to survey mode. Of note, in the very high distress range, there was a 10% difference between the online (19.0%) and CATI (9.9%) results for 2020 and a five percent difference for 2022 (19.4% and 14.8%, respectively). Such large proportional differences may have serious implications for policies based on prevalence estimates.

4.1. Strengths and limitations

The strengths of the current study include the large community samples involved and coverage of the age range from 12 to 25 years. In particular, the younger ages have not been previously considered in research into mode effects. Further, data for both modes were collected at the same time, eliminating the confounding effects inherent in the study by Klein et al. [7]. Nevertheless, participants were not randomly assigned to the CATI or online survey groups and, although similar panels were used for recruitment, the addition of random digit dialling for the CATI recruitment means that there were some methodological differences in recruitment between the survey types that may impact the results.

This research was also the first to include a non-sensitive, physical activity measures as a point of comparison with the sensitive mental health measures. The finding that there were no significant mode effects for the non-sensitive measures helps eliminate other potential explanations, such as differences in lengths of the CATI and online surveys, and supports the theory that mode differences for sensitive mental health measures are due to social desirability bias. Note, however, that different measures of physical activity were used in 2020 and 2022, with the 2020 questions comprising just two items asked in the context of physical activity being impacted by COVID. The 2022 surveys used a well-accepted standardised physical activity measure.

In study two, the length of the CATI was significantly shorter than the online questionnaire, which may have reduced response biases due to fatigue and satisficing in the CATI. However, the inclusion of comparative non-sensitive physical activity measures somewhat mitigated the potential impact of this, with no mode effects found for these measures. Interestingly, the mode effect was weaker in the 2022 data compared with the 2020 data, where the CATI was of equivalent length to the online survey.

In our study, interviews were carried out over the phone using computer assisted technology. This is a common population surveillance method, although many population surveys are undertaken in-person, like the ABS surveys. It is unclear what the mode effects would be if measures were administered in-person, but our reasoning suggests that this format would be even more prone to social desirability effects for sensitive questions such as mental health measures, as suggested in research among adult samples where in-person and phone surveys appear to be more prone to these effects than impersonal formats [36,37]. Importantly, interviewer-administered surveys are much more expensive and resource intensive than online self-report questionnaires, yet may not provide as valid a measure of sensitive constructs.

We did not measure social desirability bias directly in our study; rather, we inferred that this phenomenon was responsible for the mode effects we found. Social desirability measures have been used in psychological and consumer research for decades [38] with the most commonly used instrument being the Marlowe-Crowne Social Desirability Scale, which defines social desirability as the need people have to respond in a way that is culturally approved [39]. It may be possible to use such a measure to determine whether social desirability is more pronounced for interview versus self-report respondents, although a recent meta-analysis seriously questions the validity of such scales with the implication that they not be used to determine social desirability bias [40].

5. Conclusion

This research provides support, using a highly powered, robust research design, for the few other studies showing that mode of administration affects reporting of mental health status in survey research, and that this varies according to age [7,22,23]. Like these studies, we suggest that our results show that interviewer-administered formats may underestimate mental ill-health and overestimate mental wellbeing compared with an online self-report format. Theoretical explanations support that the effect is in the direction of interviewer-administered surveys underestimating mental ill-health, rather than self-report modes over-estimating it, due to the motivation for positive impression management when another person is present.

Acknowledging that mode of administration and social desirability bias can impact responses to sensitive mental health questions is essential for policy-relevant research. This study shows clearly how even a weak effect can have a large impact on mental health prevalence rates. The results suggest that previous Australian population data using only interviewer-administered K10 questions may have significantly underestimated psychological distress in young Australians. Researchers and policy makers should be aware of the potential impact of social desirability bias on answers to sensitive questions on mental health.

Author contributions

Debra Rickwood: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Writing - original draft; Writing - review & editing.

Cassandra Coleman-Rose: Conceptualization; Formal analysis; Writing - original draft; Writing - review & editing.

Role of funding source

Data collection was funded by headspace National Youth Mental Health Foundation. The funder had no role in the analysis and interpretation of data; in the writing of this article; or in the decision to submit it for publication.

Ethics declaration

This study was reviewed and approved by Bellberry Limited Human Research Ethics Committee (approval numbers: 2020-04-395 and 2022-05-526).

All participants provided informed consent to participate in the studies.

Data availability statement

The data that has been used is confidential.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e20131.

References

[1] D. Gunnell, J. Kidger, H. Elvidge, Adolescent mental health in crisis, BMJ 361 (2018) k2608.
[2] World Health Organization (WHO), Mental Health - Burden, 2021. Available from: https://www.who.int/health-topics/mental-health#tab=tab_2.
[3] A. Malla, et al., Youth mental health should Be a top priority for health care in Canada, Can. J. Psychiatr. 63 (4) (2018) 216–222.
[4] Australian Bureau of Statistics (ABS), National Study of Mental Health and Wellbeing 2020-21, 2022. Available from: https://www.abs.gov.au/statistics/health/mental-health/national-study-mental-health-and-wellbeing/latest-release.
[5] Department of Health and Aged Care, Intergenerational Health and Mental Health Study, Department of Health and Aged Care, 2022. Available from: https://www.health.gov.au/initiatives-and-programs/intergenerational-health-and-mental-health-study.
[6] Australian Bureau of Statistics (ABS), National Study of Mental Health and Wellbeing Methodology, 2020, p. 21. Available from: https://www.abs.gov.au/methodologies/national-study-mental-health-and-wellbeing-methodology/2020-21.
[7] J.W. Klein, G. Tyler-Parker, B. Bastian, Measuring psychological distress among Australians using an online survey, Aust. J. Psychol. 72 (3) (2020) 276–282.
[8] A.C. Milton, et al., Comparison of self-reported telephone interviewing and web-based survey responses: findings from the second Australian young and well national survey, JMIR Ment. Health 4 (3) (2017).
[9] G. Zager Kocjan, D. Lavtar, G. Sočan, The effects of survey mode on self-reported psychological functioning: measurement invariance and latent mean comparison across face-to-face and web modes, Behav. Res. Methods 55 (3) (2022) 1226–1243.
[10] W.S. Aquilino, Effects of interview mode on measuring depression in younger adults, J. Off. Stat. 14 (1) (1998) 15–29.

[11] S. Burkill, et al., Using the web to collect data on sensitive behaviours: a study looking at mode effects on the British national survey of sexual attitudes and lifestyles, PLoS One 11 (2) (2016).
[12] M.A.C. D'Ancona, Measuring xenophobia: Social desirability and survey mode effects, Migr. Stud. 2 (2) (2014) 255–280.
[13] R. Crutzen, A.S. Goeritz, Does social desirability compromise self-reports of physical activity in web-based research? Int. J. Behav. Nutr. Phys. Activ. 8 (1) (2011).
[14] R. Tourangeau, T. Yan, Sensitive questions in surveys, Psychol. Bull. 133 (5) (2007) 859.
[15] I. Krumpal, Determinants of social desirability bias in sensitive surveys: a literature review, Qual. Quant. 47 (4) (2011) 2025–2047.
[16] R.M. Norman, et al., The role of perceived norms in the stigmatization of mental illness, Soc. Psychiatr. Psychiatr. Epidemiol. 43 (11) (2008) 851–859.
[17] J. Hoebel, et al., Mode differences in a mixed-mode health interview survey among adults, Arch. Publ. Health 72 (1) (2014).
[18] Australian Institute of Health and Welfare (AIHW), Australia's Health 2022, 2022. Available from: https://www.aihw.gov.au/reports-data/australias-health#:~:text=In%202022%2C%20no%20health%20issue,numbers%20in%20Australia%20remained%20low.
[19] A.I. Christensen, et al., Effect of survey mode on response patterns: comparison of face-to-face and self-administered modes in health surveys, Eur. J. Publ. Health 24 (2) (2014) 327–332.
[20] A.H.Y. Chu, et al., Reliability and validity of the self- and interviewer-administered versions of the global physical activity questionnaire (GPAQ), PLoS One 10 (9) (2015).
[21] J.F. Sallis, et al., Validation of interviewer- and self- administered physical activity checklists for fifth grade students, Med. Sci. Sports Exerc. 28 (7) (1996) 840–851.
[22] L. Wettergren, E. Mattsson, L. von Essen, Mode of administration only has a small effect on data quality and self-reported health status and emotional distress among Swedish adolescents and young adults, J. Clin. Nurs. 20 (2011) 1568–1577.
[23] F. Beck, R. Guignard, S. Legleye, Does computer survey technology improve reports on alcohol and illicit drug use in the general population? A comparison between two surveys with different data collection modes in France, PLoS One 9 (1) (2014).
[24] C.A. Kelly, et al., Social desirability bias in sexual behavior reporting: evidence from an interview mode experiment in rural Malawi, Int. Perspect. Sex. Reprod. Health. 39 (1) (2013) 14–21.
[25] R. Kessler, et al., Short screening scales to monitor population prevalences and trends in non-specific psychological distress, Psychol. Med. 32 (2002) 959–976.
[26] G. Andrews, T. Slade, Interpreting scores on the Kessler psychological distress scale (K10), Aust. N. Z. J. Publ. Health 25 (2001) 494–497.
[27] C.L.M. Keyes, The mental health continuum: from languishing to flourishing in life, J. Health Soc. Behav. 43 (2) (2002) 207–222.
[28] C.L. Craig, et al., International physical activity questionnaire: 12-country reliability and validity, Med. Sci. Sports Exerc. 35 (8) (2003) 1381–1395.
[29] M. Hagströmer, et al., Concurrent validity of a modified version of the International Physical Activity Questionnaire (IPAQ-A) in European adolescents: the HELENA study, Int. J. Obes. 32 (S5) (2008) S42–S48.
[30] Australian Bureau of Statistics (ABS), First Insights from the National Study of Mental Health and Wellbeing, 2020-21, 2021. Available from: https://www.abs.gov.au/articles/first-insights-national-study-mental-health-and-wellbeing-2020-21.
[31] S. Bradford, D.J. Rickwood, Young people's views on electronic mental health assessments: prefer to type than talk? J. Child Fam. Stud. 24 (5) (2014) 1213–1221.
[32] S. Bradford, D.J. Rickwood, Acceptability and utility of an electronic psychosocial assessment (myAssessment) to increase self-disclosure in youth mental healthcare: a quasi-experimental study, BMC Psychiatr 15 (2015), 305.
[33] L. Townsend, et al., Gender differences in depression literacy and stigma after a randomized controlled evaluation of a universal depression education program, J. Adolesc. Health 64 (4) (2019) 472–477.
[34] D. McDuff, et al., A large-scale analysis of sex differences in facial expressions, PLoS One 12 (4) (2017), e0173942.
[35] C. Henderson, L. Potts, E.J. Robinson, Mental illness stigma after a decade of Time to Change England: inequalities as targets for further improvement, Eur. J. Publ. Health 30 (3) (2019) 497–503.
[36] D. Heerwegh, Mode differences between face-to-face and web surveys: an experimental investigation of data quality and social desirability effects, Int. J. Publ. Opin. Res. 21 (1) (2009) 111–121.
[37] X. Zhang, et al., Survey method matters: online/offline questionnaires and face-to-face or telephone interviews differ, Comput. Hum. Behav. 71 (2017) 172–180.
[38] S.N. Beretvas, J.L. Meyers, W.L. Leite, A reliability generalization study of the marlowe-crowne social desirability scale, Educ. Psychol. Meas. 62 (4) (2002) 570–589.
[39] D.P. Crowne, D. Marlowe, A new scale of social desirability independent of psychopathology, J. Consult. Psychol. 24 (4) (1960) 349–354.
[40] L. Lanz, I. Thielmann, F.H. Gerpott, Are social desirability scales desirable? A meta-analytic test of the validity of social desirability scales in the context of prosocial behavior, J. Pers. 90 (2) (2022) 203–221.