Smith, C., Cross, M., Duncan, S., Grimes, A., Howden-Chapman, P., Le Gros, L., Logan, A., MacKay, L., McPhee, J., Murphy, K., Narayanan, A., Penny, G., Pierse, N., & Stewart, T. (2026). Measuring experienced wellbeing: Two methods for smartphone-based Ecological Momentary Assessment. *International Journal of Wellbeing*, *16*(1), 4693, 1-27. https://doi.org/10.5502/ijw.v16i1.4693

ARTICLE

Measuring experienced wellbeing: Two methods for smartphone-based Ecological Momentary Assessment

Conal Smith · Matt Cross · Scott Duncan · Arthur Grimes
Philippa Howden-Chapman · Lydia Le Gros · Amber Logan · Lisa MacKay
Julia McPhee · Kate Murphy · Anantha Narayanan · Guy Penny · Nevil Pierse
Tom Stewart

Abstract: Since 2010 the widespread availability of smartphones has lowered barriers to collecting experienced wellbeing (EWB) data through ecological momentary assessment (EMA). This paper reports on two New Zealand studies that explore the impact of social context, activity, and the diurnal cycle on the EWB of social housing tenants. The two studies adopt different approaches to implementing EMA via a smartphone. The first uses a smartphone app to collect information on EWB, while the second adopts an instant messaging approach.

Despite differences in the EMA approach used and sample characteristics, both studies produce very similar results. Social contact is associated with higher experienced wellbeing, as is travelling when compared to other activities. Both studies also effectively replicate findings from the wider literature on the relationship between the diurnal cycle and experienced happiness.

Because the studies both took place in New Zealand in the second half of 2023 and focus on similar target populations, a comparison between them has useful lessons for the impact of the different approaches to EMA on data collection. We find that the instant messaging approach achieved better compliance with the EMA protocol, suggesting that the familiarity of instant messaging may be beneficial for data quality for some respondents. However, the instant messaging approach requires more researcher time per interview and cannot collect GPS information on the respondent's location, sacrificing a major strength of EMA measures.

Keywords: Well-being, Happiness, Ecological Momentary Assessment, EMA, Smartphone

1. Introduction

Over the last decade, measures of subjective wellbeing have been increasingly used to monitor social progress, to inform public policy, and to help understand what matters for a good life. A foundational concept in the measurement of subjective wellbeing is the distinction between evaluative measures of wellbeing and measures of experienced wellbeing (Kahneman, 1999; OECD, 2013). Where evaluative measures[1] capture a person's reflective assessment of their life

[1] The standard evaluative measure of wellbeing is a survey measure of life satisfaction such as that proposed by the OECD in their Guidelines for Measuring Subjective Wellbeing (OECD, 2013):
The following question asks how satisfied you feel, on a scale from 0 to 10. Zero means you feel "not at all satisfied" and 10 means you feel "completely satisfied". Overall, how satisfied are you with life as a whole these days?

Conal Smith
Kōtātā Insight Limited, New Zealand
conal.smith@kotatainsight.co.nz

(or some aspect of it), measures of experienced wellbeing capture a person's feelings at a particular point in time[2]. Although the two types of measure are related to each other – in that they both provide information about a person's subjective mental state – they capture different information and have different drivers (Boarini et al., 2013). An evaluative assessment as to whether a person is "happy with their life as a whole" is not the same thing as an experiential measure of whether a person feels "happy" now.

Many papers instrumental in establishing the relevance and validity of measures of subjective wellbeing focus on experienced wellbeing. The influential textbook *Wellbeing: Foundations of hedonic psychology* (Kahneman, Diener, and Schwarz, 1999) focuses largely on measures of experienced wellbeing. So too do many of the most cited papers on subjective wellbeing published from the late 1990s through to the middle of the next decade (Kahneman, Wakker, and Sarin, 1997; Kahneman et al., 2004a, 2004b; Kahneman and Krueger, 2006). However, despite the strong focus on experienced wellbeing in the late 1990s and early 2000s, most papers on subjective wellbeing now focus on evaluative measures – particularly measures of overall life satisfaction.

One major factor driving research towards evaluative measures of subjective wellbeing, as opposed to experienced wellbeing, has been the cost of collecting experienced wellbeing data. Where evaluative wellbeing measures can be collected at relatively little cost through standard household surveys, experienced wellbeing data requires either the use of a device (such as a pager or smartphone) for ecological momentary assessment (EMA; Shiffman, Stone, and Hufford, 2008) or a full time-use diary for the day reconstruction method (Kahneman et al., 2004a). This makes experienced wellbeing data more complex and resource intensive to collect and imposes a higher burden on respondents.

Despite these costs, experienced wellbeing data is valuable because it can capture the impact of life circumstances on wellbeing, where evaluative wellbeing measures fail. Kahneman and Sugden (2005) identify two broad types of case where experienced wellbeing measures are valuable. The first are circumstances where exposure to an activity is an endogenous choice driven by the participant's preferences (such as attending a cultural event). Here experienced wellbeing measures – which can capture how within person variations in experienced wellbeing change in different contexts – can better identify the impact of a cultural event on wellbeing than evaluative measures. The second area where experienced wellbeing measures are valuable is where spatial and temporal matching are important. Because experienced wellbeing measures capture how a person feels at a particular point in time, they can be more sensitive to where the respondent is when responding than evaluative measures if location is captured as part of data collection (MacKerron and Mourato, 2013).

Although the day reconstruction method remains an important technique for collecting experienced wellbeing data, the widespread ownership of smartphones has allowed for the use of EMA on a large scale (Killingsworth and Gilbert, 2010; MacKerron and Mourato, 2013). While early EMA studies relied on a dedicated device such as a pager, which would need to be provided by the researcher, this is not the case when use of a suitable device – such as a smartphone – is

[2] There is less standardisation of experienced wellbeing measures than is the case for evaluative measures. A typical question is that used in the Day Reconstruction Method (Kahneman et al., 2004a). Respondents are asked about the feelings during particular episodes during the previous day with questions covering 11 different feelings in the initial study. The question item for happiness is given below:
Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt: Happy?

already widespread in the population. In addition to availability, the use of a smartphone app for EMA has several other advantages. An app allows for a dedicated and customised interface for data collection, can collect a wide range of data beyond experienced wellbeing (e.g. GPS data, ambient sound levels, accelerometery), and can be customised to provide feedback to the user to encourage respondent uptake. However, such an approach also requires the respondent to download and set up the app with the appropriate permissions. Both downloading and use of the app may require the use of mobile data, which could have costs for the respondent.

An alternative approach to implementing EMA on a smartphone is used by Kitsaras et al. (2020) and is based on Short Message Service (SMS) text messages. In this case a chatbot implements the EMA protocol by sending prompts and receiving responses via SMS text message. This has the advantage that the respondent does not need to download an app and the interface is one most people are familiar with. It is also relatively easy to set up so that any costs associated with messaging are met by the researcher rather than the respondent. However, unlike an app, the SMS based approach cannot capture and store data while offline and is thus less suitable where cellphone coverage is incomplete.

This study describes two EMA measurement projects targeting broadly similar population groups (social housing tenants and their neighbours) in the same country over the same time period, but using different methods. The first study – Te Hotonga Hapori (THH) – collects experienced wellbeing data via a custom smartphone app, while the second – Public Health and Urban Regeneration: Maximising Wellbeing (PHUR) – uses a chatbot and SMS text messaging to collect comparable data.

The objectives of the paper are twofold. First, the paper aims to present the substantive results of the two studies with respect to the interactions between time use, activity, and social context for social housing tenants in a developed country. However, because the focus of the two studies is so similar, it is also valuable to explore the lessons learned in implementing the EMA protocol via SMS text messaging as opposed to via dedicated EMA app. To address both topics we first describe the data collection approach adopted in each study. The paper then explores the survey sampling properties associated with each study and then examines patterns of experienced wellbeing in both studies. We then discuss the main findings to emerge from the two studies with a particular emphasis on lessons for data collection and research design.

One reason for focusing particularly on the lessons for data collection in this paper is that, despite the potential importance of methodological factors in research design, the literature on method effects in EMA is relatively thin (Csikszentmihalyi and Larson, 2014; Liao et al., 2016; Jones et al., 2019; and Wang et al., 2025, are notable exceptions). This article contributes to this literature by systematically comparing the THH and PHUR EMA studies. The impact of the two data collection approaches is of particular interest as groups with a low socio-economic status – such as social housing tenants – are often more challenging to obtain responses from but their experiences are often of high interest to public policy.

1.1 Two EMA methods

The last decade has seen a strong focus on making better use of data to inform public policy in New Zealand under the ambit of "social investment" (centre-right National Party 2015–2017, 2023-) and "wellbeing budgeting" (centre-left Labour Party 2017–2023). Although the two approaches are framed somewhat differently, in principle they share much in common including a clear focus on empirically evaluating the outcomes achieved through public spending. In this context, two independent research programmes were commissioned in 2020. Both projects are

multi-disciplinary multi-year studies addressing the relationship between social housing tenant wellbeing and housing characteristics from a range of perspectives.

The *THH* study focuses primarily on the impact of urban redevelopment on wellbeing and is centred on four social housing redevelopments in Auckland undertaken by Kāinga Ora (New Zealand's largest public housing provider). The PHUR study looks at the relationship between urban redevelopment and wellbeing across six providers of social housing with different governance models (including Kāinga Ora). Both studies include an EMA component intended to support analysis of the relationship between characteristics of place and experienced wellbeing.

1.2 Te Hotonga Hapori

The THH study includes qualitative components, a national survey of tenant wellbeing collected in three waves (2022, 2023, and 2024), analysis of administrative data, and a neighbourhood study with two waves of data collection (2023 and 2024) involving both a traditional wellbeing survey and EMA. The 2023 wave of EMA data is examined in this paper.

EMA study participants were recruited through a two-stage process. First, residents of four Auckland neighbourhoods undergoing urban redevelopment (Aorere, Oranga, Waikōwhai, and Wesley) were recruited into THH by interviewers going door to door. Interviewing took place over a year from April 2023 to May 2024. Successful contacts were asked to complete a wellbeing survey which provides information on evaluative wellbeing (life satisfaction), domain-specific wellbeing outcomes based on the New Zealand Treasury's Living Standards Framework (Smith, 2018; Hughes, 2021), demographic information, and neighbourhood characteristics. A total of 1997 addresses were identified within the four neighbourhoods of which 1667 were not vacant. These were approached up to 4 times each from which 478 useable responses[3] were achieved for a response rate of 28.7%.

In addition to completing the wellbeing survey respondents were asked to download the THH experienced wellbeing smartphone app[4], with 188 providing useable data (response received and consistent with protocol requirements – see below). This app was complemented by a web application which handled data management, questionnaire setup, and data downloads. The web app also allowed researchers to see real-time updates of responses, serving as the backend that communicates with the mobile app. All data collected are securely stored in a cloud server.

Participants were asked a fixed set of five questions at each prompt to assess their experienced wellbeing and social interaction. Responses were recorded using a slider to indicate values ranging from 0 (not at all) to 10 (completely). The experienced wellbeing questions were:

- How happy do you feel right now?
- How anxious do you feel right now?
- How tired do you feel right now?

On a second screen, respondents identify their current activity from a list of options such as eating, housework, working, studying, caring for others, traveling, exercising, or other activities. Participants select the one option that best describes their activity at the time of the prompt. Similarly, on a third screen respondents are asked who they are with at that moment, with

[3] Valid responses were those with no missing data on core demographic and wellbeing measures.

[4] As part of the wider THH research programme respondents were also asked to wear a wrist-mounted accelerometer (Axivity AX3) and hip-mounted GPS device (QStarz BT-Q1000XT) for the next 7 days which provided detailed physical activity and location data.

options including "My children", "Other family", "Other people I know / friends", "People I don't know / other people". Respondents can select all applicable options. This structured approach allows for detailed and nuanced data collection about participants' daily experiences and interactions. The full question instrument is attached as Appendix 1.

The EMA protocol used with the app sends 3 prompts daily over a period of 7 consecutive days, implying 21 measurements per respondent with full compliance. Prompt times are randomly allocated within three broad time slots (morning 8am-12pm, afternoon 12pm-5pm, and evening 5pm-9pm). On receiving a prompt, the respondent then had 60 minutes to respond to the first question with the prompt lapsing and no data being collected if they did not respond within this time. Once a person has responded to a prompt, they must then complete all questions within 30 minutes of starting their response. In addition to the substantive responses (or lack thereof), the app records the exact time at which the prompt was sent and at which time a response was received[5].

Alongside information gathered directly from the responses, the app collects contextual information associated with each response. The most important element here is the GPS coordinates of the respondent at the time that they completed the question instrument. This is taken directly by the app from the respondent's smartphone.

1.2 Public Housing and Urban Redevelopment: Maximising wellbeing

The PHUR study was designed to examine the interaction between the governance models of social housing providers, physical housing characteristics, and the wellbeing of social housing tenants. Data collection includes qualitative components, a tenant wellbeing survey collected in 2022, analysis of administrative data, and an EMA study.

Participants for the EMA study were recruited from respondents to the tenant wellbeing survey who indicated that they were willing to be recontacted for further research. The tenant wellbeing survey was broadly similar in content to the THH tenant wellbeing survey. Both questionnaires include comparable demographic data and the same evaluative wellbeing question on overall life satisfaction.

The sample frame was based on an address list provided by social housing providers with respondents receiving a pre-notification postcard about the research branded with both the provider and the research team, followed by up to three survey packs which included letters from the housing provider, the research team, and the survey itself. The survey could be completed on paper and returned by post or on-line. 2,580 surveys were posted to addresses on the register of which 148 were returned with "gone no address" or a similar response. There were 575 completed surveys received, for a response rate of 23.6%. Surveying took place in July and August 2022.

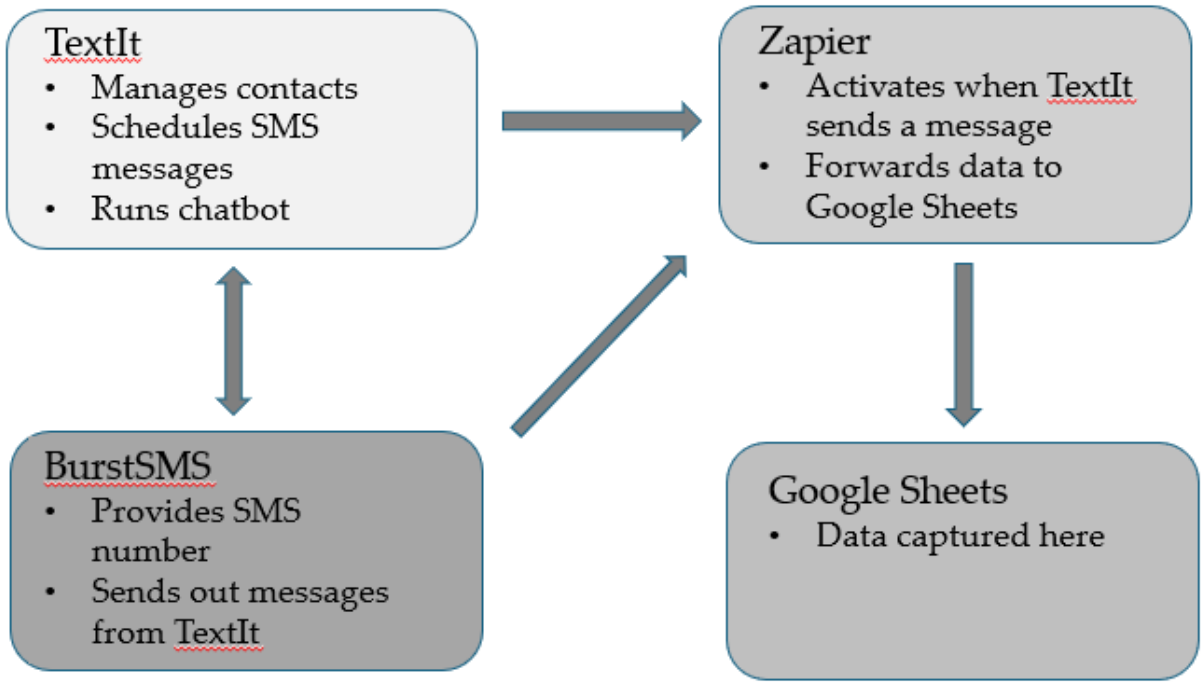
Of the respondents who completed the 2022 PHUR tenant wellbeing survey 67.1% gave permission to be recontacted for future research. A sample of 84 respondents was drawn in 2023 for the EMA study of whom 42 agreed to participate. The EMA sample was then split into two waves, one of which ran from 30 July to 5 August 2023 and the other from 27 August to 2 September 2023.

Following Kitsaras et al. (2020) the EMA protocol was implemented through a chatbot and SMS text messages using three tools. A simple chatbot management tool – TextIt – was used to

[5] Timestamps are stored on the app and relate to the respondent's activities. When data is stored on the app and sent later due to a lack of mobile coverage the timestamps relate to the respondent's activities, not the time of the download to the server.

build the instrument interface, manage the contact list for the respondents, schedule texts, and collect responses. A phone number for the project was provided by a commercial SMS provider – Burst SMS. Text messages in the EMA – those sent via the chatbot and responses to the phone number – were paid for by PHUR making participation in the EMA study costless to respondents. Finally, another online tool – Zapier – was used to link the Burst SMS account to TextIt and to direct responses to the TextIt chatbot to a password protected spreadsheet in Google Sheets[6]. Immediately after each wave was completed data was deleted from Google Sheets and moved to a secure data storage site. This data collection process is illustrated in Figure 1 below.

Figure 1. PHUR EMA data collection



From the respondent's point of view the EMA instrument simply involved responding to a series of SMS text messages on their mobile phone through the standard interface. Responses involved texting a single number indicating the response from among those specified in the question text. Some respondents provided more complex responses, writing out their answers in full. These were coded to the instrument response categories during data cleaning. The PHUR question instrument asked respondents the following seven questions:

- On a scale of 0 to 10, where 0 is completely unhappy and 10 is completely happy, how happy do you feel right now?
- Where are you right now?
- What are you doing at the moment?
- Who is with you at the moment?
- Right now do you feel (very cold… very warm)
- Do you feel respected where you are right now?
- Do you feel at ease where you are right now?

[6] The data forwarded to Google Sheets included a unique respondent identifier code and answers to the questions in the EMA instrument. However, no identifying information was stored in this dataset at any point.

The questions on where the respondent is, what they are doing, and who they are with each have four response options while subjective temperature is captured on a five-point labelled Likert scale. Although asked of all respondents, the final two questions (feeling respected and at ease) have yes/no responses and were included specifically to capture aspects of experienced wellbeing that may be of salience for indigenous Māori tenants. These questions were developed in a joint process involving both the PHUR wellbeing team and Māori researchers from the broader PHUR project. A full version of the EMA instrument for PHUR is attached as Appendix 2.

The EMA protocol used in PHUR was very similar to that used in THH. Respondents were prompted three times per day over the course of seven days (Sunday to Saturday) for a maximum of 21 responses per person with full compliance. When the prompt was sent respondents were given 30 minutes to respond to the first question and 5 minutes between responses after this. Each question was sent to the respondent after receipt of an answer to the preceding question. If the response time elapsed with no response a termination message was sent to the respondent.

Prompts were sent out at 30 minutes past the hour between 8 am and 8 pm each day. A pre-determined schedule was used to ensure an even spread of prompts across the day and across respondents. This schedule was also used to ensure that respondents who had indicated that they had non-standard hours were only contacted during the period that they had specified[7].

Although the SMS-based nature of the PHUR study precludes gathering other information such as GPS coordinates directly from the respondent's smartphone, the question on "where are you right now" was designed to enable linkage between the respondent's experienced wellbeing responses, their subjective temperature assessments, and temperature monitors installed in the respondent's home as part of the broader PHUR research programme.

The two studies discussed here were not intended as a methodological comparison of approaches to EMA. In fact, the choice of an SMS methodology for the PHUR study was heavily influenced by the fact that looking at experienced wellbeing represents only a small part of the broader PHUR work programme. Nonetheless, the two studies provide a strong basis for examining the impact of different EMA assessment methods and, more generally, illustrating some of the relative strengths and weaknesses of each approach. Table 1 below summarises the key characteristics of each study.

2. Survey sampling properties

It is evident from Table 1 that, although the two studies are similar, there are important differences between them that extend beyond methodology. In particular, the sample recruitment strategy is a little different as is the geographic and temporal scope. Before exploring the relationship between experienced wellbeing and the respondent's activity and social context, it is therefore useful to explore the degree to which differences in EMA outcomes are affected by features of study design. To do so we adopt a simple framework that breaks the impact of study design on EMA outcomes into two broad categories. Notably, we consider how the design of each study affects:

1. Selection into the study sample
2. Compliance with the study protocol given selection into the sample

This provides a simple structure for capturing the impact of study design on EMA measures and can be thought of as capturing "who participates" and "whether they respond".

[7] This affected 1 respondent out of 42.

Table 1. Summary of study EMA characteristics

	THH	PHUR
EMA mode	Smartphone app	Chatbot and SMS text
Population	Residents from four Auckland neighbourhoods undergoing urban redevelopment by Kainga Ora	Social housing tenants from 6 social housing providers, one of which was Kainga Ora
Region	Auckland	Auckland, Porirua, Wellington, Christchurch
Survey response rate	28.7%	23.6%
Recruitment rate to EMA study	39.3% (188 EMA responses out of 478 survey responses)	33.6% (67.1% permission to recontact x 50% agreement to EMA)
Net EMA response rate	11.3%	8.3%
Number of respondents	188	42
Number of responses	1,573	539
Field dates	20/4/23 to 27/5/24	30/7/23 to 2/9/23.
Protocol length	3 prompts per day over 7 days	3 prompts per day over 7 days
Protocol prompt times	Random time within three time slots (morning, afternoon, evening)	Prompts at 30 minutes past the hour on a fixed schedule ensuring equal sampling across people and hours over the week.
Standdowns	Valid responses are those where the respondent commenced responding within 60 minutes of initial prompt and completed responding to all questions within 30 minutes.	Valid responses are those received within 30 minutes of initial prompt and within 5 minutes of responding to the previous question.
Headline wellbeing measure	How happy do you feel right now? From 0 (not at all) to 10 (completely).	On a scale of 0 to 10, where 0 is completely unhappy and 10 is completely happy, how happy do you feel right now?

2.1 Selection into the study sample

Table 1 suggests that, despite differences in study design, both THH and PHUR achieve very similar rates of recruitment into their EMA studies (11.8% for THH vs 8.3% for PHUR). The slightly higher recruitment rate for THH may be linked to the fact that THH recruited respondents into the EMA study as part of the same process as initial recruitment for the wellbeing survey and the EMA study followed immediately. In the case of PHUR, however, the EMA study took place a year after the initial wellbeing survey and involved recontacting respondents who had agreed to participate in further research.

Table 2 explores the demographic profile of respondents and compares this with the characteristics of the larger samples from which the EMA samples were drawn. Two features are immediately apparent from Table 2. First, despite both studies including social housing as a core study element it is evident that the respective populations of the two wellbeing surveys differ in important ways. Second, both the experienced wellbeing samples are relatively good reflections of the overall survey population that they draw from (i.e. there are relatively few demographic characteristics where there are large differences between either of the survey samples and their corresponding EMA sub-sample).

The sample frame for the PHUR wellbeing survey consists of tenants from six different social housing sites. In contrast, the sample for the THH wellbeing survey is people residing in one of

four Auckland neighbourhoods where a significant social housing redevelopment is taking place. This means that while the focus of THH as a study is very much on social housing and wellbeing, its sample includes social housing tenants, private sector tenants, and homeowners.

Table 2. Demographic profile of study respondents

	THH WB survey sample N=478		**THH EMA sample N=188**		**PHUR WB survey sample N=591**		**PHUR EMA sample N=37**	
	Prop	**SE**	**Prop**	**SE**	**Prop**	**SE**	**Prop**	**SE**
Age								
Under 35	0.35	0.02	0.39	0.04	0.12	0.01	0.05	0.04
35 to 64	0.51	0.02	0.52	0.04	0.58	0.02	0.62	0.08
65 plus	0.14	0.02	0.10	0.02	0.30	0.02	0.32	0.08
Tenancy								
Social Housing	**0.25**	**0.02**	**0.14**	**0.03**	1.00	-	1.00	-
Sex								
Male	0.30	0.02	0.30	0.03	0.39	0.02	0.54	0.08
Ethnicity								
NZ European	0.32	0.02	0.41	0.04	0.50	0.02	0.62	0.08
Maori	0.16	0.02	0.13	0.02	0.27	0.02	0.24	0.07
Pacific	**0.36**	**0.02**	**0.24**	**0.03**	0.23	0.02	0.08	0.04
Asian	0.17	0.02	0.19	0.03	0.03	0.01	0.05	0.04
Qualifications								
No qualifications	0.25	0.02	0.18	0.03	0.37	0.02	0.57	0.08
HH Size								
1	0.08	0.01	0.05	0.02	**0.49**	**0.02**	**0.73**	**0.07**
2	0.20	0.02	0.20	0.03	0.19	0.02	0.14	0.06
3	0.18	0.02	0.22	0.03	0.11	0.01	0.05	0.04
4	0.20	0.02	0.23	0.03	**0.09**	**0.01**	**0.00**	**0.00**
5	0.14	0.02	0.13	0.02	0.06	0.01	0.03	0.03
6	0.06	0.01	0.06	0.02	0.04	0.01	0.05	0.04
7+	0.14	0.02	0.11	0.02	**0.03**	**0.01**	**0.00**	**0**
HH income								
$20,000 or less	0.12	0.01	0.06	0.02	0.28	0.02	0.26	0.08
$20,001 to $40,000	**0.23**	**0.02**	**0.12**	**0.02**	0.34	0.02	0.50	0.09
$40,001 to $70,000	0.19	0.02	0.21	0.03	0.15	0.01	0.21	0.07
$71,000 to $100,000	0.12	0.01	0.15	0.03	0.05	0.01	0.00	0.00
$100,001 or more	0.35	0.02	0.45	0.04	**0.18**	**0.02**	**0.03**	**0.03**

Note. WB – Wellbeing, EMA – Ecological Momentary Assessment; Prop – proportion, SE – Standard Error. Values in bold indicate a statistically significant within-study difference between the composition of the EMA sample and the survey sample it was drawn from at the 5% level in a chi-square test. All distributions (age, tenancy, sex, ethnicity, qualifications, HH size, and HH income) are significantly different between the PHUR and THH studies at the 5% level in a chi-square test.

This difference is reflected in the composition of the samples, with THH having a younger sample with fewer males, fewer New Zealand European or Māori respondents, and more Pacific and Asian respondents. The THH sample is also less likely to have no qualifications, to live alone,

and to have a household income below $40,000[8]. The main factor responsible for the differences between the two survey samples is that PHUR draws solely on social housing tenants while THH does not. The fact that THH focuses entirely on Auckland, which has a much higher concentration of people of Pacific and Asian ethnicity than the rest of New Zealand is also likely to be a factor in differences between the two populations.

While there are differences between the survey populations for the two research programmes, the EMA studies are both a relatively good reflection of the survey sample from which they are drawn. There are only three differences between the THH EMA sample and the survey sample from which it is drawn that are statistically significant at the 5 percent level (highlighted in bold in Table 2). These are the proportion of the EMA population which identifies as Pacific, the proportion with an annual household income of between $20,000 and $40,000, and the proportion of respondents who are social housing tenants.

The situation is similar when we compare the PHUR EMA study with the wellbeing survey population from which it was drawn. There are only a handful of statistically significant differences between the two populations; notably the proportion of the EMA population living alone is higher than for the PHUR wellbeing survey and the proportion with an annual household income of more than $100,000 is lower. In addition, there are a few other differences that appear large – such as the proportion of the population that is Pacific – but which are not statistically significant due to the small sample size of the PHUR EMA study.

2.2 Compliance with the study protocol

Compliance with the EMA study protocol is important for data quality. Unlike the wellbeing surveys, each of the EMA studies involves multiple responses from each respondent over the course of a week. With 3 responses per day over a 7-day period, full compliance with the EMA protocol involves 21 responses per individual respondent. The achieved number of responses can be lower than this maximum for two reasons. Either the respondent drops out of the study before the full week passes – perhaps due to finding the survey too intrusive or burdensome – or the respondent remains in the EMA study, but fails to respond to some prompts.

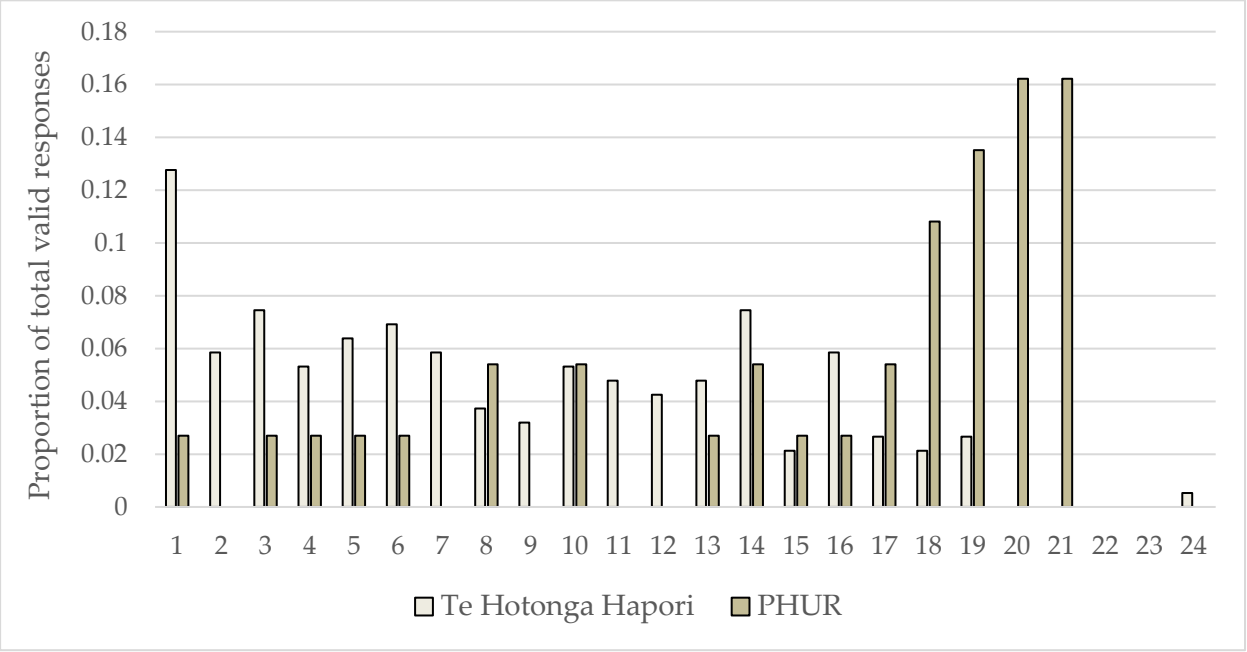
Figure 2 shows the proportion of respondents in each study who responded to any given number of prompts over the study period. For example, approximately 13% of the THH sample responded to only one prompt (and consequently provided little useful data). In contrast, just under 3% of the PHUR sample responded to only one prompt during the study period.

In contrast to recruitment into the study – where both programmes produced similar results – there is a clear difference in compliance with the study protocols between the two programmes. For the app-based study (THH) there was a fairly even distribution of levels of compliance up to about 16 responses (of 21 potential responses). Over a third of the sample (37.8%) responded 5 times or less. Very few respondents within the THH sample (7.9%) achieved high compliance (17 or more responses) and none responded to the full 21 prompts.

The SMS-based approach to EMA adopted by the PHUR study, on the other hand, achieved very high rates of compliance. More than 60% of the PHUR sample complied with 17 or more of the prompts while less than 11 percent responded 5 or fewer times. The impact of these distributions is reflected in a large difference in the mean responses per person for the two studies (8.4 for THH as opposed to 15.5 for the PHUR study).

[8] All differences are significant at $p<0.05$ in a chi-square test.

Figure 2. Number of prompts for which a valid response was received



Potential factors influencing the difference in compliance rates between the two studies include the different demographic composition of the two samples (table 2), differences in respondent burden[9], and mode effects reflecting the impact of app-based data collection as opposed to SMS. Although the two studies do not allow for the latter two effects to be disentangled, it is possible to explore the degree to which demographic differences account for differences in protocol compliance.

Table 3 reports the results of linear regression models on a combined dataset including both studies with the number of responses as the dependent variable. Column 1 includes only a dummy variable identifying respondents in the PHUR study as an explanatory variable. In column 2 this dummy is replaced by a series of demographic controls and in column 3 both the demographic controls and the dummy for the PHUR study are included. The results of column 2 indicate that respondent ethnicity, labour force status and, to a lesser degree, age are significantly associated with EMA protocol compliance[10]. However, comparing the magnitude of the coefficient on the PHUR study in columns 1 and 3 suggests that these differences explain little of the mean difference in compliance rates between the two studies. Being a social housing tenant is associated with just over three fewer responses on average (column 3). However, since the PHUR study is entirely social housing tenants the net effect of this is to increase the size of the study design features as reflected in the dummy variable for being in the PHUR study which increases from 7.15 in column 1 to 9.45 in column 3.

While the number of responses is one dimension of protocol compliance, it is also of interest whether the two modes are associated with differences in when people respond. Table 4 presents the results of a series of ordered probit regressions on the number of responses within the

[9] Respondents for THH were asked to wear a wrist-mounted accelerometer over the same period as the EMA study and completed the wellbeing survey immediately before starting the EMA protocol, neither of which was part of the protocol for the PHUR study.

[10] Controls for educational qualifications, household income, and household size were included in initial regressions but were dropped from the models presented in Table 3 as none were significant. Their exclusion does not materially affect the results presented in Table 3.

morning, afternoon, and evening response windows respectively on the same pooled EMA dataset used in Table 3, column 3.

Table 3. Number of responses as a function of demographic characteristics

	(1) Survey dummy only	(2) Demographic controls only	(3) Survey and demographic controls
VARIABLES	*Responses*	*Responses*	*Responses*
PHUR study	7.15***		9.45***
	(1.05)		(1.32)
Social housing tenant		2.20*	-3.13**
		(1.02)	(0.97)
Male		-0.26	-0.92
		(0.94)	(0.87)
Māori		-2.57*	-2.08+
		(1.25)	(1.11)
Pacific		-4.06***	-2.69**
		(1.02)	(0.89)
Asian		-2.30+	-2.11+
		(1.19)	(1.14)
		Age (reference = 18 to 24)	
25 to 34		1.15	0.84
		(1.44)	(1.44)
35 to 44		0.81	0.40
		(1.48)	(1.47)
45 to 54		2.12	1.38
		(1.58)	(1.51)
55 to 64		3.27+	1.71
		(1.80)	(1.73)
65 to 74		0.04	-1.14
		(1.89)	(1.85)
75 plus		-2.31	-3.67*
		(1.69)	(1.82)
		Labour force status (reference = not in labour force)	
Employed		-0.92	-0.48
		(0.97)	(0.86)
Unemployed		-4.67***	-2.34+
		(1.38)	(1.35)
Constant	8.37***	10.38***	10.37***
	(0.41)	(1.53)	(1.53)
Observations	225	225	225
R-squared	0.18	0.18	0.31

Note. Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10. OLS regression estimate with the number of EMA prompts responded to as the dependent variable. Gender and ethnic variables are binary dummies with a value of 1 if "yes". All others independent variables are categorical variables with coefficient capturing the difference from the reference category.

There is some evidence that demographic factors affect when people respond, with Māori respondents significantly less likely to respond in the morning or evening, Pacific respondents significantly less likely to respond in the afternoon or evening, and social housing tenants less likely to respond in the morning or evening. However, these effects are not large. In contrast, being in the PHUR study was associated with an increased probability of responding at all times of day, but this effect was larger in the morning (β=1.78) and particularly the afternoon (β=1.97) compared to the evening (β=1.11)[11].

Table 4. Probability of responding to EMA prompts by hour of day, pooled sample

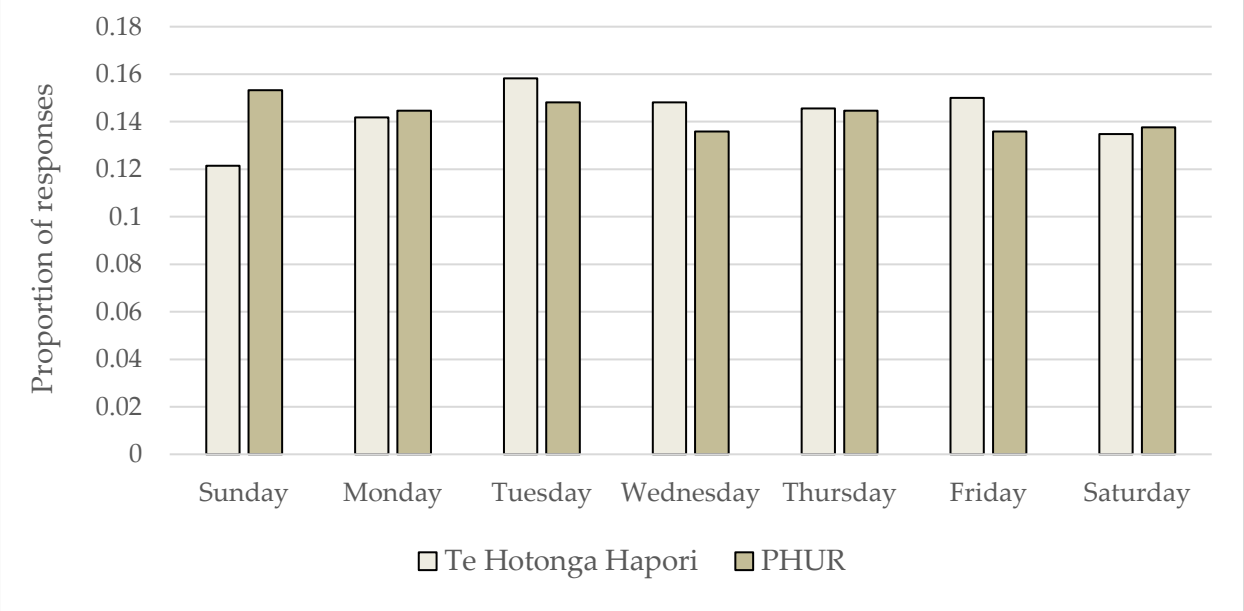
VARIABLES	(1) Morning response (From 8am until 12pm)	(2) Afternoon response (From 12pm until 4pm)	(3) Evening response (After 4pm)
PHUR study	1.78***	1.97***	1.11***
	(0.32)	(0.34)	(0.24)
Social housing tenant	-0.66**	-0.29	-0.59**
	(0.22)	(0.24)	(0.20)
Male	-0.14	-0.09	-0.25
	(0.16)	(0.16)	(0.17)
Māori	-0.56*	-0.17	-0.40+
	(0.22)	(0.23)	(0.23)
Pacific	-0.30+	-0.58**	-0.47**
	(0.18)	(0.18)	(0.17)
Asian	-0.30	-0.54*	-0.30
	(0.22)	(0.21)	(0.22)
Age (reference = 18 to 24)			
25 to 34	0.11	0.34	-0.11
	(0.26)	(0.30)	(0.29)
35 to 44	0.19	0.05	-0.21
	(0.27)	(0.30)	(0.31)
45 to 54	0.18	0.24	0.07
	(0.29)	(0.31)	(0.31)
55 to 64	0.52+	0.08	0.09
	(0.31)	(0.33)	(0.35)
65 to 74	0.15	-0.72+	-0.28
	(0.35)	(0.38)	(0.36)
75 plus	-0.77	-0.25	-0.74
	(0.47)	(0.38)	(0.45)
Labour force status (reference = not in labour force)			
Employed	-0.09	-0.09	0.02
	(0.18)	(0.18)	(0.16)
Unemployed	-0.53	-0.20	-0.39
	(0.34)	(0.26)	(0.28)
Observations	225	225	225

Note. Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10. Ordered Probit regression estimates with the number of EMA prompts responded to within the response window as the dependent variable. Gender and ethnic variables are binary dummies with a value of 1 if "yes".

[11] The difference between the PHUR coefficients for morning and afternoon or evening is not statistically significant at the $p < 0.05$ level but the difference between the afternoon and evening PHUR coefficient is significant.

There is relatively little evidence of differences across the two EMA modes in the distribution of responses across the days of the week. Figure 3 shows the proportion of responses for each EMA study by day of the week. With the exception of a slightly lower response rate for THH on Sunday and an elevated response rate on the same day for the PHUR study, there is relatively little difference in response rates across the week. One key difference between the two studies that may explain the difference in rates on Sunday is that the PHUR study data collection ran from Sunday to Saturday for all respondents, so the first day of data collection for each respondent was the Sunday. In contrast, THH ran for a full week from the day after the respondent was interviewed in the wellbeing survey, which had the potential to be any day of the week.

Figure 3. Number of prompts responded to by day of week



3. Patterns of experienced wellbeing across EMA methods

The analysis in the previous section shows clear differences in the sample characteristics of the two studies and also in compliance with EMA prompts. However, the former does not explain much of the variance in the latter. This section explores what the two studies show about the relationship between experienced wellbeing, social context, and activities for social housing tenants. Because the two surveys use slightly different categories for social context and activity[12] we provide an initial descriptive overview of mean happiness by social context and activity for the PHUR and THH studies separately (table 5) before exploring how experienced wellbeing varies with changes in social context and activity in a more nuanced way using pooled data and a fixed effects regression (table 6).

The two studies appear to produce a very different pattern of mean happiness by social context (Table 5). For THH, mean happiness appears highest in contexts where family are present and is particularly high where both family and friends are potentially present. The lowest level

[12] Although the EMA measurement instruments used in THH and PHUR are generally well-aligned, there is one area where they differ quite substantially. This is where data is collected on the respondent's social context at the time they are prompted. The app used in THH allows for multiple responses and this is reflected in the "who are you with at the moment" question. In contrast, the SMS chatbot used for the PHUR study was not designed to accommodate multiple responses so four mutually exclusive categories were developed. Where THH focuses on the respondent's relationship to the people they are with (family, children, friends, strangers), PHUR focuses on whether the respondent lives with the people they are with when responding.

of experienced happiness is reported when alone. The pattern is, however, very different for the PHUR study where the highest level of experienced happiness is reported when the respondent is with people they do not live with, while the lowest is reported in a situation combining people they live with and people they do not.

The two studies both show relatively high mean experienced wellbeing for paid and unpaid work and for travel, with other activities associated with lower levels of happiness. This is counter-intuitive given a standard finding in analysis of experienced wellbeing data is that work and travel are associated with lower levels of happiness (Kahneman and Krueger, 2006; MacKerron and Mourato, 2013). A final observation is that life satisfaction from the wellbeing survey and mean experienced happiness from the EMA are almost identical in PHUR whereas there is a large difference between the two for THH. For THH mean experienced happiness is 1.54 points (on a 0 to 10 scale) below life satisfaction.

Table 5. Mean experienced wellbeing by social context and activity

	Te Hotonga Hapori		**PHUR**	
	Mean	**SE**	**Mean**	**SE**
With				
Alone	6.82	0.1	7.02	0.14
With (PHUR)				
People I live with			6.85	0.31
People I don't live with			7.49	0.23
Both	8.28	0.37	6.3	0.67
With (THH)				
Family and/or children	7.11	0.08		
Friends and/or strangers	6.85	0.13		
Activity				
Paid or unpaid work/housework	7.09	0.09	7.85	0.15
Travel	7.03	0.26	8.05	0.36
Other	6.94	0.08	6.72	0.14
Summary Measures				
Experienced Happiness	6.99	0.13	7.32	0.34
Life satisfaction	8.44	0.13	7.38	0.32
Life satisfaction-happiness gap	-1.45	0.14	-0.06	0.28

Some of the counter-intuitive happiness levels for social context and especially for activity reported in Table 5 may reflect the fact that a simple analysis of means fails to control for important confounding factors. Perhaps most importantly, respondents with paid work are likely to have higher incomes than others without paid work, raising their experienced happiness compared to respondents without paid work. However, the differences in sample characteristics discussed in section 2 represent another potential source of concern. To control for unobserved confounding factors at the individual level we estimate the following regression model with individual fixed effects for each dataset individually and for both combined:

$$W_{i,d,h} = \beta_0 + \beta_1 A_i + \beta_2 S_i + \beta_3 h_h + \beta_4 d_d + \gamma_i + \epsilon$$

where $W_{i,d,h}$ is the experienced happiness of individual i at hour h on day d; A_i is the activity at the time happiness is reported, S_i is who the respondent is with when happiness is reported, and γ_i is an individual fixed effect.

The results of this model are reported in Table 6 and differ substantially from the picture in Table 5 confirming the importance of controlling for confounding factors. For THH lack of social contact is associated with significantly lower experienced happiness. Being alone is associated with a decrease of 0.28 points of happiness ($p < 0.05$) relative to being with others. The results for the PHUR study also show the impact of social contact where being alone is also associated with a decrease in happiness of 0.28 points although a smaller sample size means that this is only marginally significant ($p < 0.10$).

The results in Table 6 also show much less evidence of an impact of activity on experienced wellbeing. For THH, which has a larger sample, there is some evidence that travelling is associated with higher experienced wellbeing than work ($p < 0.1$) while there are no significant coefficients in the PHUR sample. However, in both studies all coefficients on activities and social context are of the same sign and are broadly similar in magnitude. Pooling both samples (table 5, column 3) largely replicates the same qualitative results as regressions on the surveys individually (columns 1 and 2) but with higher levels of statistical significance (largely due to the greater statistical power provided by a larger sample). This is much more consistent with the wider literature on work and experienced wellbeing than the results In Table 5.

Table 6. Impact of social context and activity on experienced wellbeing

VARIABLES	(1) PHUR Survey	(2) THH Survey	(3) Pooled surveys
With whom (reference = other people)			
Alone	-0.28+	-0.28*	-0.27**
	(0.15)	(0.11)	(0.09)
Activity (reference = paid or unpaid work)			
Travel	0.31	0.37+	0.40*
	(0.35)	(0.22)	(0.18)
Other	0.14	0.10	0.11
	(0.15)	(0.10)	(0.08)
Constant	6.59***	6.92***	6.81***
	(0.33)	(0.23)	(0.19)
Day of week controls	Yes	Yes	Yes
Hour of day controls	Yes	Yes	Yes
Observations	547	1,572	2,119
R-squared	0.06	0.03	0.03
Number of IDvar	37	188	225

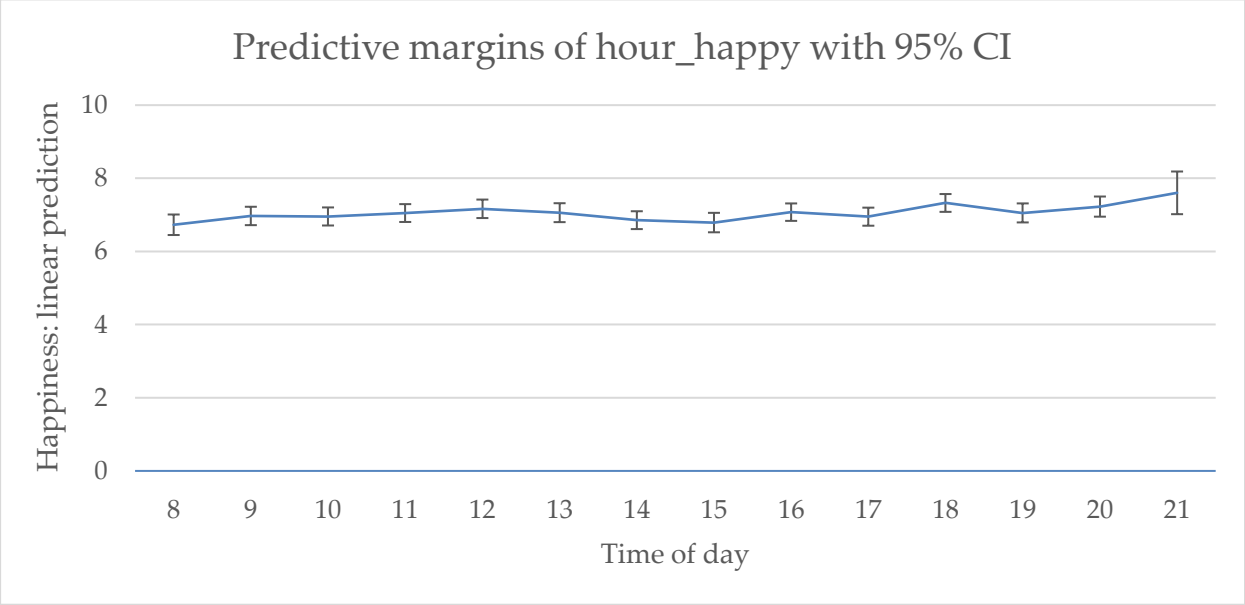
Note. Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10. Ordinary least squares regression with fixed effects and robust standard errors. Coefficients capture the change in points of life satisfaction (0-10) associated with the activity or social context described relative to the reference.

Another approach to examining the data collected through each of the two studies is to look at the extent to which each study reproduces the diurnal rhythms of affect (Kahneman et al., 2004a). For measures of positive affect such as experienced happiness the literature typically shows happiness increasing in strength over the morning and rising to a peak about midday. Happiness decreases immediately after midday before gradually rising in the evening.

Using the hour of day controls from the regressions presented in Table 6, it is possible to reconstruct the diurnal rhythm of happiness for the pooled THH and PHUR datasets. To do this we calculate the expected value of happiness for each hour of the day, holding the day of the

week, activity and social context constant. The resulting profile of expected happiness is presented in Figure 4.

Figure 4. Diurnal rhythm



Note. Predicted experienced happiness based on hour of day plus 95% confidence interval from pooled PHUR and THH data.

The predicted happiness values in Figure 4 largely replicates the diurnal rhythm of happiness from Stone et al. (2006). Happiness starts low, rises until midday, falls during the afternoon and then rises again in the evening. Although the predicted happiness scores in Figure 4 appear relatively flat, the changes in response to time of day are consistent with other studies. For example, with a difference between 8am and 9pm of 0.9 points and a difference between 3pm and 9pm of 0.8 points the level of variation over the course of the day is broadly comparable with results from Stone et al. (2006) which have an equivalent gap of 1.3 points between 8am and 9pm and 0.9 points between 3pm and 9pm[13].

4. Lessons and discussion

The results presented here provide information on the experienced wellbeing of social housing tenants. Although the two studies discussed here use slightly different instruments to collect information on social context and activities at the time of EMA prompts, they present a remarkably consistent picture. Social context is important for the wellbeing of social housing tenants, with the experienced wellbeing of social housing tenants about 0.27 points lower (on a 0-10 scale) when people are alone than when they have company. This effect is strong enough that it is marginally significant ($p<0.10$) in the smaller PHUR sample and is significant at a higher level in the THH ($p<0.05$) and pooled ($p<0.01$) samples.

A second, and unexpected, result is the positive association between time spent travelling and experienced wellbeing. While the fact that travel is associated with higher experienced

[13] Stone et al. (2006) use a 0-6 scale as opposed to the 0-10 scale used in the PHUR and THH studies. The values reported here for changes in happiness over the course of the day have been made comparable by multiplying the Stone et al. results by 11/7. The two profiles of happiness over the course of the day (Figure 4 and Stone et al.) have a correlation coefficient of 0.82 ($p<0.001$).

wellbeing than work (paid or unpaid) is consistent with the wider literature on experienced wellbeing (e.g. Krueger et al., 2009), the fact that travel is more positive than other activities is unexpected. This may suggest that getting out of the house is of greater importance for social housing tenants than for the population at large.

The relatively small sample sizes and the sample selection effect documented in section 2 represent an important limitation with respect to both the studies reported here. However, the fact that the main empirical results are not sensitive to the non-trivial differences in sample composition between the PHUR and THH studies provides some reassurance as to the robustness of these results. A second reason for some confidence in the results is that, where the THH and PHUR studies can be tested against results from the wider literature on experienced wellbeing (such as looking at the relationship between the diurnal cycle and happiness), both studies largely reproduce the expected results.

4.1 Lessons for data collection

In addition to the substantive findings, the studies discussed here also provide important lessons on the applicability of the different approaches to EMA adopted each study. Given that EMA studies are costly to undertake in terms of both budget and respondent burden, any information that can improve study design is potentially valuable. Information of this sort is particularly valuable when – as in the case of the PHUR and THH studies – it helps identify the relative strengths of different data collection approaches with groups that are difficult to engage with but of high policy interest such as social housing tenants.

4.1.1 Cost

Cost is unavoidably a major factor in research design. Rarely, if ever, are research funds sufficient to allow methodological criteria to be the sole factor informing decisions. The two studies discussed here have very different cost profiles. These are presented in Table 7.

Table 7. Cost of App and SMS based data collection (THH and PHUR)

Study	Approach	Fixed costs		Data costs		Research time per respondent	
THH	App-based	High	c$170,000[14] to build and test the app across Apple and Android platforms	Low	Data costs covered by housing provider	Low	No management of respondents required
PHUR	SMS-based	Low	c$500 for subscriptions to TextIt, Zapier, and Burst SMS	High	c$40 per respondent in text costs at $0.10 per message	High	c15 minutes per respondent

[14] I.e. approximately US$100,000. Note that this included significant additional functionality beyond simply implementation of the EMA to include the ability to capture photos, audio recordings and to trigger prompts through geo-fencing.

In the case of THH most of the cost of the project is front-loaded. Development costs for the app were considerable reflecting the need for a custom app for THH's specific information needs. However, this represents a one-off cost that does not scale with the number of times that the app is used and places little burden on the research staff. The project leveraged the "Zero Data Initiative" in collaboration with Kāinga Ora, the social housing provider for the area in which the study took place which met data costs associated with the project. However, even had the data costs been fully borne by the research programme, they were low relative to the per-text costs of the SMS approach.

In contrast, development costs for the SMS-based approach adopted by the PHUR project were much lower. The development of the EMA system was undertaken by one of the research team with minimal direct financial outlay albiet at the expense of some weeks of researcher time. Costs per respondent, however, were significantly greater than for the app-based approach. Each complete response requires 16 SMS messages to administer contributing to a significant per-response cost. In addition, the chatbot developed for the PHUR project required researcher intervention to schedule the prompts, involving approximately two minutes per day over 7 days per respondent (roughly fifteen minutes of researcher time per respondent).

This difference in cost structures suggests that, where an app needs to be developed from scratch for a piece of research, the breakeven point on cost relative to an SMS-based approach is relatively high (approximately 1000 respondents for one week based on the costs encountered by THH and the PHUR study). This reflects the higher upfront costs for app development but greater ease of scalability for an app-based study (where the time researcher time requirements do not increase linearly with the number of respondents). In contrast, if an app is available "off the shelf" the lower cost per respondent associated with an app-based approach makes the app significantly less expensive.

4.1.2 Flexibility

One important difference between the two approaches is flexibility in testing changes to the question instrument. Changes to an app potentially require both programming time to implement and for respondents to update the app if this has not been set up to happen automatically. In some cases – such as changes to the questions or timing – implementation is relatively straight forward, but other aspects are more complicated (particularly if changes would require respondents to update or re-download the app). A chatbot, on the other hand, allows for changes to the question instrument to be handled centrally without respondent interaction. Further, making changes to the chatbot requires very little in the way of programming skills (TextIt – the chatbot interface used in the PHUR study has a click and drag flowchart interface for setting up the chatbot). The net impact is that an SMS/chatbot approach to collecting EMA data is flexible in terms of making changes to the question instrument or customising the questions asked when compared to an app.

4.1.3 Integrating with other data

While the SMS approach to collecting EMA data is flexible, the use of SMS as opposed to an app to deliver questions means that there is no ability to access data from the respondent's mobile device other than the responses to questions asked by the chatbot. This is a significant limitation since smartphones can provide information on the respondent's GPS coordinates at the time of response, basic accelerometery and ambient sound levels, among other data. While accessing such information requires informed consent, collecting GPS data is routine in app-based EMA studies (e.g. Killingsworth and Gilbert, 2010; MacKerron and Mourato, 2013).

The use of GPS-linked EMA data to explore the relationship between experienced wellbeing and place is one of the strongest reasons for collecting EMA data (Kahneman and Sugden, 2005) and has been used to explore the impact of green space on wellbeing (MacKerron and Mourato, 2013; Doherty, Lemieux, and Canally, 2014; Bakolis et al., 2018; Seresinhe et al., 2019; Kondo et al., 2020), air pollution (MacKerron and Mourato, 2009; Krekel and MacKerron, 2020), airport noise (Fujiwara, Lawton, and MacKerron, 2017), transport mode (Glasgow et al., 2019), sequence of physical activity (Chinapaw et al., 2019), characteristics of the urban environment (Birenboim, 2018), and adolescent physical activity (Bejarano et al., 2019).

In the context of EMA research, the weight attached to integration with GIS data or the use of accelerometery will depend on the study focus. Where these factors are central to the research design, an app-based approach dominates an SMS-based design simply because the latter cannot easily collect the required information. Similarly, techniques such as geofencing – where the respondent's location is used to trigger a prompt – are not possible through an SMS-based approach. For studies where this is not the case, the limitations of an SMS-based design with respect to data integration are less important. One key caveat here is that a number of prominent EMA studies (e.g. Killingsworth and Gilbert, 2010; MacKerron and Mourato, 2013) were – at least partially – speculative in design (i.e. designed with a wide range of applications in mind). In these circumstances there is real value in the richer dataset made available through a dedicated app.

4.2 Conclusion

Information on the pros and cons of different research methods on data collection is of fundamental importance in good research design. Nonetheless, much knowledge about these issues is tacit or unpublished and passed on between researchers as part of the "craft" of research. The comparison presented in this paper contributes to the accessibility of information on the design of EMA studies building on the relatively limited existing literature (e.g. Csikszentmihalyi and Larson, 2014; Liao et al., 2016; Jones et al., 2019).

Although the two studies discussed here are not perfect natural experiments for identifying methodological effects, they are robust enough to draw some clear conclusions. First, the SMS approach to EMA works and collects valid data. The PHUR study considered here is more complex and wide ranging in its scope than existing SMS-based EMA studies such as Tyler and Olson (2018) or Kitsaras et al. (2020). In particular, rather than focus on a single event (such as bedtime in the case of Kitsaras et al.), the PHUR study replicates the scope of the THH EMA with a week-long duration and multiple responses per day at (to the respondent) random times.

When used in this context the SMS approach used in the PHUR study replicates the key patterns of wellbeing over the day, across different social contexts and – less clearly – between activity types that we would expect to see from the wider literature providing good evidence of construct validity for the resulting data. In fact, in terms of data quality, the only strong difference between the THH and the PHUR study is the difference in rates of protocol compliance, which favoured the PHUR SMS study.

A second important point is the impact of aspects of study design on compliance with the EMA protocol. While demographic differences explain some part of the difference in compliance rates between the two studies, most of the difference is attributable to differences in aspects of study design: either mode or respondent burden. To the degree that app-based approaches to data collection have been in relatively widespread use for the past decade and that high compliance rates have been achieved (De Vries, Baselmens, and Bartels, 2021) it is reasonable to infer that the lower levels of protocol compliance in THH are not primarily driven by mode. This, then, points strongly towards the impact of respondent burden on EMA compliance rates.

While the lower compliance rates for the first wave of the THH EMA study discussed here do not undermine the validity of the data (which still performs well in terms of construct validity), the effective sample size of the dataset is smaller than would otherwise be the case, reducing statistical power. A clear lesson for the design of EMA studies (also identified in De Vries, Baselmens, and Bartels, 2021) is the importance of minimising avoidable respondent burden. The number and length of prompts play a particularly important role here.

Finally, from comparing the two EMA studies it is possible to draw some conclusions about the relative strengths of each approach and where each of the two methodologies are most likely to be useful. The SMS/text-based approach can be characterised as low fixed-cost, higher cost per respondent, and low respondent burden with a relatively high degree of customisability. In contrast, app-based approaches may have a low or high fixed cost (depending on whether the research project has access to an open source of off the shelf EMA app), lower cost per respondent and higher respondent burden at set-up combined with a higher cost involved in making changes to the EMA instrument. However, an app-based approach can collect a much wider range of data – including GPS coordinates and accelerometery – that cannot be picked up so easily with a simpler SMS instrument.

The differing features of SMS-based as opposed to app-based approaches to EMA suggest that the two methods should be considered complementary, with the choice of method guided by research objectives. Where a research project is relatively narrowly focused, requires only a small sample, and where ease of customisability is important an SMS based approach has clear advantages. This might be the case in the context of a doctoral thesis where resources for large scale data collection are limited or for studies focusing on small, hard to reach populations where even relatively minor impositions – such as downloading an app – may have a relatively large impact on respondent compliance. In contrast, when larger samples are needed or research goals make it important to link experienced wellbeing measures with place, an app-based approach is to be preferred.

Authors

Conal Smith
Kōtātā Insight Limited, Wellington, New Zealand
https://orcid.org/0000-0002-1253-9922
conal.smith@kotatainsight.co.nz

Matt Cross
Sports Performance Research Institute, Auckland University of Technology, New Zealand
https://orcid.org/0000-0003-1579-3720

Scott Duncan
Dept Physical Activity and Nutrition, Auckland University of Technology, New Zealand
https://orcid.org/0000-0002-8402-2930

Arthur Grimes
Motu Economic and Public Policy Research, and Victoria University of Wellington
https://orcid.org/0000-0001-9152-8051

Philippa Howden-Chapman
Dept Public Health, University of Otago, Wellington, New Zealand
https://orcid.org/0000-0002-1529-6735

Lydia Le Gros
New Zealand Centre for Sustainable Cities, University of Otago, Wellington, New Zealand

Amber Logan
New Zealand Centre for Sustainable Cities, University of Otago, Wellington, New Zealand

Lisa MacKay
Dept Physical Activity and Nutrition, Auckland University of Technology, New Zealand
https://orcid.org/0000-0002-7344-5794

Julia McPhee
Dept Physical Activity and Nutrition, Auckland University of Technology, New Zealand
https://orcid.org/0000-0001-5792-3349

Kate Murphy
New Zealand Centre for Sustainable Cities, University of Otago, Wellington, New Zealand

Anantha Narayanan
Dept Physical Activity and Nutrition, Auckland University of Technology, New Zealand
https://orcid.org/0000-0002-1577-803X

Guy Penny
EMPlan Services, Auckland, New Zealand
https://orcid.org/0000-0003-3031-9971

Nevil Pierse
New Zealand Centre for Sustainable Cities, University of Otago, Wellington, New Zealand
https://orcid.org/0000-0002-2726-8141

Tom Stewart
Dept Physical Activity and Nutrition, Auckland University of Technology, New Zealand
https://orcid.org/0000-0001-5915-3843

Author contribution statement

CS: Conceptualisation, Methodology, Instrument development, Data collection, Formal analysis, Investigation, Original Draft, Writing, Review & Editing; MC: Instrument development, Data collection, Review; SD: Instrument development, Data collection, Review; AG: Conceptualisation, Methodology, Instrument development, Review; PHC: Instrument development, Review; LLG: Instrument development, Data collection, Review; AL: Instrument development, Data collection, Review; LM: Writing, Instrument development, Data collection, Review; JMP: Instrument development, Data collection, Review; KM: Instrument development, Data collection, Review; AN: Writing, Instrument development, Data collection, Review; GP: Instrument development, Data collection, Review; NP: Instrument development, Data collection, Review; TS: Writing, Instrument development, Data collection, Review.

Funding

This work was supported by the MBIE Endeavour Programmes, Public housing and urban regeneration: maximising wellbeing (Grant ID: 20476 UOOX2003) and Enhancing the impact of major urban regeneration on community wellbeing (Grant ID: AUTX2002).

Conflict of interest statement

The authors have no competing interests to declare that are relevant to the content of this article.

Compliance with ethical standards

The research reported in this paper was reviewed by the AUT Ethics Committee (Te Hotonga Hapori) and the University of Otago Human Ethics Committee (Public Health and Urban Regeneration). All participants in both studies gave written informed consent for participation in the study.

AI statement

AI was not used in any capacity during the preparation of this article.

Data availability statement

The datasets generated and analysed during the current study are available from the corresponding author upon reasonable request.

Publishing Timeline

Received 5 January 2025
Revised version received 18 December 2025
Accepted 31 December 2025
Published 21 January 2026

References

Bakolis, I., Hammoud, R., Smythe, M., Gibbons, J., Davidson, N., Tognin, S., & Mechelli, A. (2018). Urban mind: Using smartphone technologies to investigate the impact of nature on mental well-being in real time. *Bioscience*, *68*(2), 134-145. https://doi.org/10.1093/biosci/bix149

Bejarano, C. M., Cushing, C. C., & Crick, C. J. (2019). Does context predict psychological states and activity? An ecological momentary assessment pilot study of adolescents. *Psychology of Sport and Exercise*, *41*, 146-152. https://doi.org/10.1016/j.psychsport.2018.05.008

Birenboim, A. (2018). The influence of urban environments on our subjective momentary experiences. *Environment and Planning B: Urban Analytics and City Science*, *45*(5), 915-932. https://doi.org/10.1177/239980831769014

Boarini, R., Comola, M., de Keulenaer, F., Manchin, R., & Smith, C. (2013). Can governments boost people's sense of well-being? The impact of selected labour market and health policies on life satisfaction. *Social indicators research*, *114*, 105-120. https://doi.org/10.1007/s11205-013-0386-8

Chinapaw, M. J., Wang, X., Andersen, L. B., & Altenburg, T. M. (2019). From Total Volume to Sequence Maps: Sophisticated Accelerometer Data Analysis. *Medicine and science in sports and exercise*, *51*(4), 814-820. https://doi.org/10.1249/mss.0000000000001849

Csikszentmihalyi, M., Larson, R. (2014). Validity and Reliability of the Experience-Sampling Method. In: *Flow and the Foundations of Positive Psychology*. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9088-8_3

De Vries, L. P., Baselmans, B. M., & Bartels, M. (2021). Smartphone-based ecological momentary assessment of well-being: A systematic review and recommendations for future studies. *Journal of Happiness Studies*, *22*(5), 2361-2408. https://doi.org/10.1007/s10902-020-00324-7

Doherty, S. T., Lemieux, C. J., & Canally, C. (2014). Tracking human activity and well-being in natural environments using wearable sensors and experience sampling. *Social science & medicine*, *106*, 83-92. https://doi.org/10.1016/j.socscimed.2014.01.048

Fujiwara, D., Lawton, R. N., & MacKerron, G. (2017). Experience sampling in and around airports. Momentary subjective wellbeing, airports, and aviation noise in England. *Transportation Research Part D: Transport and Environment*, 56, 43–54. https://doi.org/10.1016/j.trd.2017.07.015

Glasgow, T. E., Le, H. T., Geller, E. S., Fan, Y., & Hankey, S. (2019). How transport modes, the built and natural environments, and activities influence mood: A GPS smartphone app study. *Journal of Environmental Psychology*, *66*, 101345. https://doi.org/10.1016/j.jenvp.2019.101345

Hughes, T. (2021). *Towards a Living Standards Framework for all Aotearoa: Culture, children and wellbeing (No. 21/01).* New Zealand Treasury Discussion Paper.

Jones, A., Remmerswaal, D., Verveer, I., Robinson, E., Franken, I. H., Wen, C. K. F., & Field, M. (2019). Compliance with ecological momentary assessment protocols in substance users: A meta-analysis. *Addiction*, *114*(4), 609-619. https://doi.org/10.1111/add.14503

Kahneman, D. (1999). Objective happiness. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 3-25). Russell Sage Foundation.

Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *Journal of Economic perspectives*, *20*(1), 3-24. https://doi.org/10.1257/089533006776526030

Kahneman, D., & Sugden, R. (2005). Experienced utility as a standard of policy evaluation. *Environmental and Resource Economics*, *32*, 161-181. https://doi.org/10.1007/s10640-005-6032-4

Kahneman, D., Diener, E., & Schwarz, N. (Eds.). (1999). *Well-being: Foundations of hedonic psychology*. Russell Sage Foundation.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004a). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, *306*(5702), 1776-1780. https://doi.org/10.1126/science.1103572

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. (2004b). Toward national well-being accounts. *American Economic Review*, *94*(2), 429-434. https://doi/org/10.1257/0002828041301713

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics*, *112*(2), 375-406. https://doi.org/10.1162/003355397555235

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, *330*(6006), 932-932. https://doi.org/10.1126/science.1192439

Kitsaras, G., Goodwin, M., Allan, J., Kelly, M., & Pretty, I. (2020). An Interactive Text Message Survey as a Novel Assessment for Bedtime Routines in Public Health Research: Observational Study. *JMIR public health and surveillance, 6(4),* e15524. https://doi.org/10.2196/15524

Kondo, M., Triguero-Mas, M., Donaire-Gonzalez, D., Seto, E., Valentín, A., Hurst, G., Carrasco-Turigas, G., Masterson, D., Ambròs, A., Ellis, N., Swart, W., Davis, N., Maas, J., Jerrett, M., Gidlow,C., and Nieuwenhuijsen, M. (2020). Momentary mood response to natural outdoor environments in four European cities, *Environment International, 134*, 105237. https://doi.org/10.1016/j.envint.2019.105237

Krekel, C., & MacKerron, G. (2020). How environmental quality affects our happiness. *World happiness report*, 95-112. https://www.worldhappiness.report/ed/2020/how-environmental-quality-affects-our-happiness/

Krueger, A. B., Kahneman, D., Schkade, D., Schwarz, N., & Stone, A. A. (2009). National time accounting: The currency of life. In *Measuring the subjective well-being of nations: National accounts of time use and well-being* (pp. 9-86). University of Chicago Press.

Liao Y, Skelton K, Dunton G, Bruening M., (2016). A Systematic Review of Methods and Procedures Used in Ecological Momentary Assessments of Diet and Physical Activity Research in Youth: An Adapted STROBE Checklist for Reporting EMA Studies (CREMAS). *J Med Internet Res 18*(6): e151 https://doi.org/10.2196/jmir.4954

MacKerron, G., & Mourato, S. (2009). Life satisfaction and air quality in London. *Ecological Economics*, *68*(5), 1441-1453. https://doi.org/10.1016/j.ecolecon.2008.10.004

MacKerron, G., & Mourato, S. (2013). Happiness is greater in natural environments. *Global Environmental Change*, *23*(5), 992-1000. https://doi.org/10.1016/j.gloenvcha.2013.03.010

OECD. (2013). *OECD Guidelines on Measuring Subjective Well-being*. OECD Publishing, Paris.

Pudney, S. (2010). *An experimental analysis of the impact of survey design on measures and models of subjective wellbeing* (No. 2010-20). ISER Working Paper Series. https://hdl.handle.net/10419/65958

Seresinhe, C.I., Preis, T., MacKerron, G. and Moat, H. (2019). Happiness is Greater in More Scenic Locations. *Sci Rep 9*, 4498. https://doi.org/10.1038/s41598-019-40854-6

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, *4*, 1-32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Smith, Conal (2018). *Treasury Living Standards Dashboard: monitoring intergenerational wellbeing*. Kōtātā Insight. Wellington. https://hdl.handle.net/11159/2029

Stone, A. A., Schwartz, J. E., Schkade, D., Schwarz, N., Krueger, A., & Kahneman, D. (2006). A population approach to the study of emotion: diurnal rhythms of a working day examined with the Day Reconstruction Method. *Emotion*, *6*(1), 139. https://doi.org/10.1037/1528-3542.6.1.139

Tyler, K. A., & Olson, K. (2018). Examining the Feasibility of Ecological Momentary Assessment Using Short Message Service Surveying with Homeless Youth: Lessons Learned. *Field Methods, 30(2), 91–104.* https://doi.org/10.1177/1525822x18762111

Wang, S., Yang, C. H., Brown, D., Cheng, A., & Kwan, M. Y. W. (2025). Participant Compliance with Ecological Momentary Assessment in Movement Behavior Research Among Adolescents and Emerging Adults: Systematic Review. *JMIR mHealth and uHealth*, *13*, e52887. https://doi.org/10.2196/52887

Appendices

Appendix 1. THH EMA Instrument

Figure A1. Collecting experienced wellbeing data in Te Hotonga Hapori

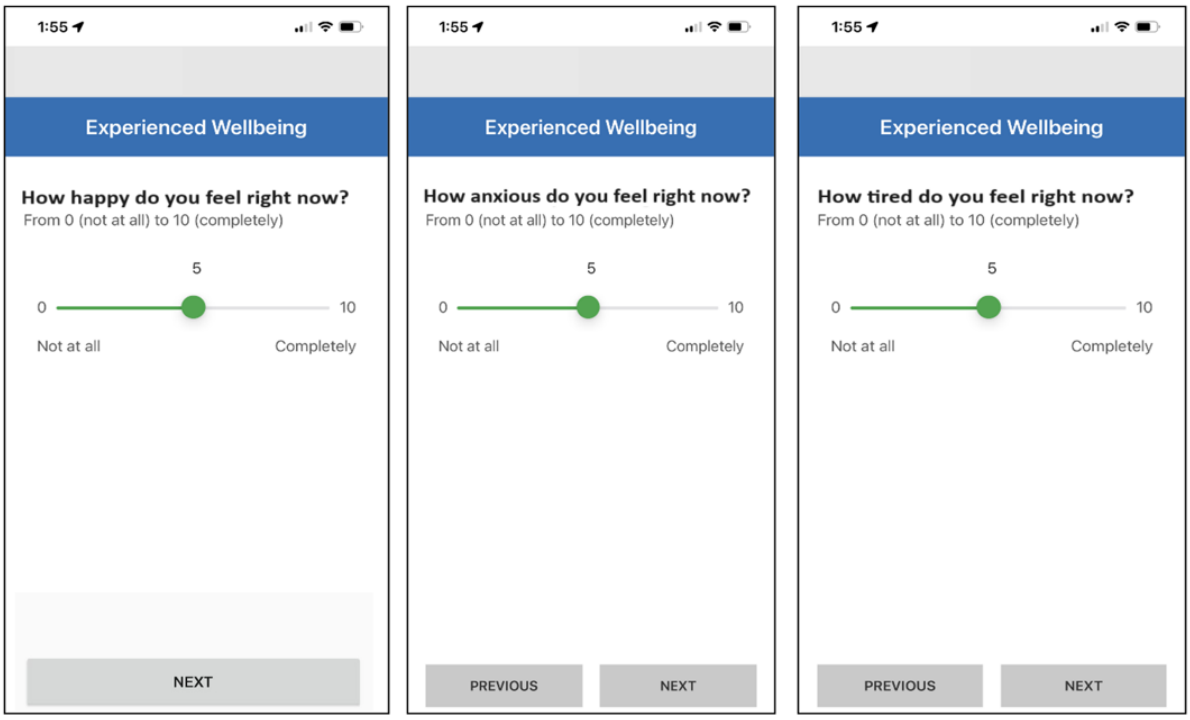
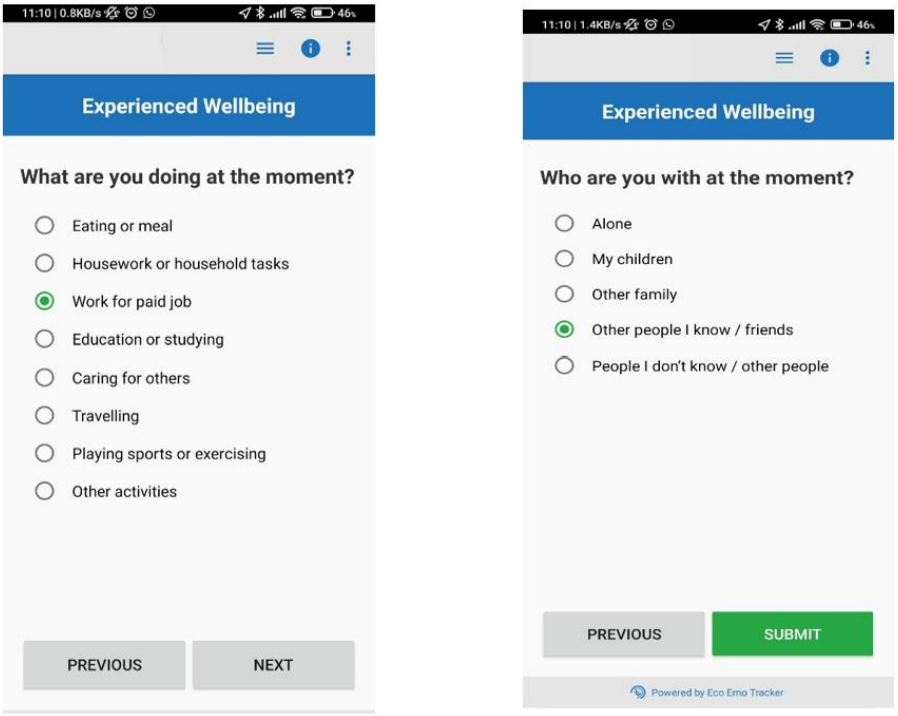


Figure A2. Collecting contextual information in THH



Appendix 2. PHUR EMA instrument

Kia ora.

This is the Measuring Experienced Wellbeing at Home Study.

On a scale from 0 to 10, where 0 is completely unhappy and 10 is completely happy, how happy do you feel right now?

Where are you right now?
1 Inside at home
2 Outside at home
3 In my neighbourhood
4 Away from home

What are you doing at the moment?
1 Paid work, housework, or household task
2 Travelling or commuting
3 Social activities
4 Other activities / something else

Who is with you at the moment?:
1 I am alone
2 I am with other people who live with me
3 I am with other people who do not live with me
4 I am with people who live with me and with other people

Right now, do you feel
1 Very cold
2 A little too cold
3 Just right
4 A little too warm
5 Very warm

Do you feel respected where you are right now?
1 Yes
2 No

Do you feel at ease where you are right now?
1 Yes
2 No

Thank you for your responses.