Radarweg 60
1043 NT Amsterdam
The Netherlands

www.tno.nl

T +31 88 866 50 10

TNO report

TNO 2018 P11413 | Final report

Validation of Quby 'Verspillingschecker' (Waste Checker)

Date	26 November 2018
Author(s)	M. van Lidth de Jeude, L. Dreijerink, C. Tigchelaar
Copy no	
No. of copies	
Number of pages	89 (incl. appendices)
Number of appendices	5
Sponsor	Quby
Project name	Validation of Quby 'Verspillingschecker' (Waste Checker) Quby
Project number	060.34019

Summary

The Verspillingschecker (Energy Waste Checker) is an application developed by Quby and launched at the Dutch Toon smart thermostat platform. The app generates automated individual insights about where energy is wasted in a user's home, and provides advice about how to reduce this. The app includes seven specific so-called use cases (namely showerhead, thermostat, refrigerator, washing machine, dishwasher, dryer, standby use) and 3 overall energy use cases (gas for heating home, gas for heating water and electricity use). Based on patterns in the energy data and additional calculations Quby provides diagnoses to the user: people get to see whether a specific usage is efficient or not and what a user could do about that (advice). For the seven use cases people receive either a green (efficient) or red (inefficient) message/flag.

The aim of this research was to look into the effects of the Verspillingschecker. We wanted to know if the automatically generated personalized advice can realize energy savings in households, and how effective this advice would be. Moreover Quby wanted to know the impact of the Verspillingschecker on a use-case specific level, as well as on indicators for future change.

The two main research questions were:

1. Can automatically generated personalized advice realize energy savings per household of 5-10%?
2. How effective is automatically generated personalized advice on energy efficiency measures in households?

Besides the hypotheses based on the research questions, four assumptions about how these effects are generated have been tested:

a. The amount of data collected by the Verspillingschecker application is adequate for detecting inefficient energy usage by the Verspillingschecker.
b. The generated personal advice offers home owners sufficient insight in their energy conservation potential.
c. The generated personal advice provides home owners insight in what energy conservation measures are the most interesting for them.
d. The generated personal advice helps home owners in taking investment decisions concerning energy conservation measures.

Method

We collected data on overall energy use (weekly gas and electricity) of the Verspillingschecker users and of a comparable control group. In addition we gathered data using a questionnaire in both the Verspillingschecker and control group. Participants of this questionnaire were asked to give their permission to link the two datasets. Finally we also used data on the seven specific use cases for the Verspillingschecker users.

In total we received energy and use case data of 53.931 Verspillingschecker users and energy data of 38.868 Eneco clients without Toon and the Verspillingschecker app. Of these two groups respectively 2.534 and 3.234 people filled in the

questionnaire. Of the two groups approximately 2.000 participants gave permission to link their two datasets (see Table 1).

Table 1 Overview of people per group participating in the research

	Energy use (m^3, kWh) per week	Questionnaire	Questionnaire with permission to link data
Verspillingschecker group	53.931	2.534	1.985
Control group	38.868	3.234	1.873

Results

Impact on overall energy use
After four weeks we found a small effect on electricity use among Verspillingschecker users: the decline of the electricity use was 0.01% stronger than in the control group. This was a statistically significant effect, although small. It means that in both groups the electricity use decreased, but in the Verspillingschecker group this decrease was a little steeper. For gas use we did not find any difference.

We also looked at the longer term effects. After 14 weeks we found a significant effect on electricity use: The decline of the electricity use was 0.03% stronger among the users of the Verspillingschecker than the control group. For gas use we found no significant differences over this time period.

Use cases
When focusing on the separate use cases, we see that providing feedback slightly increases the energy use of the refrigerator, washing machine, dishwasher and stand-by usage. Regarding the thermostat use case, the related energy use at night decreased slightly on average. The showerhead showed no significant change over time. The dryer was not analysed further, since every user received a red diagnosis.

When we take a closer look at the type of feedback given, it stands out that for the refrigerator, washing machine, dishwasher, thermostat and shower head a red diagnosis (inefficient) leads to a small decrease in the energy use of the appliance, while a green diagnosis (efficient) leads to a small increase in the energy use. Thus, the small positive effect of showing people a diagnosis that their usage of these appliances is inefficient, seems to be diminished by the increase of energy use of the people who receive a message that their usage of these appliances is efficient. This was not the case for stand-by usage. For stand-by usage both a red and green diagnosis lead to a small increase of related energy use.

Impact on attention, attitudes towards actions, and actions taken
In the questionnaire we asked people about their attention, interest, desire and actions regarding energy waste and savings. These factors are part of the so-called AIDA model that is often used in marketing. Moreover we asked people about their attitudes towards saving energy and attitudes towards the specific actions that are advised by the Verspillingschecker. Our results did not confirm the separate AIDA factors exist. In fact, attention, interest, desire were perceived as one factor that we labelled ‘attention to prevent energy waste at home’.

We found that the Verspillingschecker has an impact on overall attention for energy waste in home. The overall attention increase was stronger for Verspillingschecker users than for the control group. Moreover we found that attitudes about replacing a still working appliance; using appliances in Eco mode, lowering the thermostat when leaving home, and lowering the thermostat at night were more positive for users of the Verspillingschecker than for the control group respondents. Also Verspillingschecker users report to have undertaken more actions in 2018 to prevent energy waste in home than the control group: they describe to have done somewhat more small actions in their home to prevent energy waste. Moreover they indicate having purchased more new energy efficient appliances than the control group. Although this hardly results in a lower energy usage, this could clarify the little difference (-0,03%) of the electricity use in the Verspillingschecker group compared to the control group.

Adequacy
All tests that were done by Quby for the launched 7 use cases confirmed the reliability of the measurements used by Quby to detect inefficiency. Via the questionnaire we came to know that most users (56%) had indeed the idea advices were credible. They (somewhat) agreed with the statement that the recommendations they received were credible.

Impact on insights and experienced help
Next to attention, attitudes and actions we also asked participants of the questionnaire to answer a number of questions about gained insights related to energy efficiency and the extent to which they experience help from Eneco on that topic. Among the Verspillingschecker users we also asked about their satisfaction with the app and whether they had any feedback.

Results showed that in the Verspillingschecker group, users more strongly believe that they increased their knowledge about how much energy they can save, than in the control group. They also indicate to know better what they can do to save energy than control group. Home owners know this better than tenants. Moreover Verspillingscheck users perceived to have had more help with preventing energy waste at home than the control group.

On average Verspillingschecker user were satisfied with Toon (score 7,8) and a bit less satisfied with the app (score 6,4). All results are summarized in Table 2 below.

Table 2 Summary of the main findings per hypothesis and assumption

Hypotheses and assumptions	Findings
H1. The Verspillingschecker group will on average show a 5-10% stronger change in energy usage than the control group.	No, over 14 weeks the decline of the electricity use was 0,03% stronger in the Verspillingschecker than in the control group
H2. Relative energy use will change with 5-10% over time after subjects in the test group have used the Verspillingschecker application.	No, over 14 weeks the Verspillingschecker group saves 0,09% on electricity and 0,07% on gas

H3. A use case specific advice (either marked with a red or green 'flag') has a significant impact on the related use case specific behaviour.	Yes, for the refrigerator, washing machine, dishwasher stand-by usage, thermostat. There was no effect regarding the shower head.
H4. The type of use case specific advice (Red or Green) has an impact on the behaviour addressed.	Yes, for the refrigerator, washing machine, dishwasher stand-by usage, thermostat, and shower head.
H5. Attention, interest and desire to take measures, attitude towards measures and the amount of measures taken changes significantly after subjects in the test group used the Verspillingschecker app.	Yes, the overall attention for energy waste in home increased among Verspillingschecker users.
H6. Verspillingschecker users will on average show a stronger change in attention, interest and desire to take measures, attitude towards measures and the amount of measures taken, than the control group.	Yes, Attitudes about replacing a still working appliance; using appliances in Eco mode, lowering the thermostat when leaving home, and lowering the thermostat at night were stronger for Verspillingschecker users than for the control group. They also report to have undertaken more actions in 2018 to prevent energy waste in home than the control group.
A.a. The amount of data collected by the Verspillingschecker application is adequate for detecting inefficient energy usage	Yes, all tests that were done by Quby for the launched 7 use cases confirmed the reliability of the measurements. See Annex E.
A.b. The generated personal advice offers home owners sufficient insight in their energy conservation potential.	Yes, the Verspillingschecker users more strongly believe that they increased their knowledge about how much energy they can save, than in the control group.
A.c. The generated personal advice provides home owners insight in what energy conservation measures are the most interesting for them.	Yes, Verspillingschecker users know better what they can do to save energy than control group
A.d. The generated personal advice helps home owners in taking investment decisions concerning energy conservation measures.	Yes, Verspillingscheck users perceived to have had more help with preventing energy waste at home than the control group.

Conclusions

1. Can automatically generated personalized advice realize energy savings per household of 5-10%?

Overall energy use
The effects of the Verspillingschecker that we found in this study were smaller than expected. Our results show that the Verspillingschecker does not have the

hypothesized impact in the order of magnitude of 5-10% on the overall electricity and gas use over time. In fact the impact was under 1%.

Use cases
An interesting pattern evolves for most of the use-cases: The small positive effect of showing people a red flag, seems to be diminished by the increase of energy use of the people who receive a green flag. It is not the case that without the negative effects of the green flags on energy saving percentages, the Verspillingschecker effect would be in range of 5-10% savings. However, it makes one think about the effects of different types of feedback and unexpected side-effects.

2. How effective is automatically generated personalized advice on energy efficiency measures in households?

Attention for energy waste at home, attitudes towards specific actions, and amount of actions taken
When looking at the results of the questionnaire, more indicators for a positive effect were found: The Verspillingschecker had a positive impact on 'attention to prevent energy waste at home.' At the action-level, Verspillingschecker users reported to have done more actions to prevent energy waste at home than control group participants (in 2018: just after the introduction of the Verspillingschecker). It is not possible to claim the higher number of actions in the test group would not have exist without the Verspillingschecker, but as it corresponds with a stronger attention for energy waste among these users, it creates a strong case that the Verspillingschecker made an impact at this level.

Assumption a: The amount of data collected by the Verspillingschecker application is adequate for detecting inefficient energy usage by the Verspillingschecker.
This is true according to testing done by Quby. Most but not all users of the Verspillingschecker agreed with this statement. Further investigation of the comments made by the users via the questionnaire could lead to further improvements and/or ways to convince users of the credibility of the diagnoses.

Assumptions b,c,d: Increased insights and experienced help
Other results show that in the Verspillingschecker group, users gained more insights in how much energy they can save, know better what they can do to save energy, and perceive more help from Eneco with preventing energy waste at home than in the control group. It is not clear to what extend this is due to the Verspillingschecker versus due to the Toon itself.

In short, the Verspillingschecker app did increase attention for energy waste at home, and most likely also changed behaviour, but not systematically, and with little effect on energy use.

Discussion and recommendations

Studies show that percental savings due to feedback devices are often overestimated. A percental saving in the range of 5% could be realistic, provided that the application and data visualization connect to the practical preference and

interests of the consumer (RVO, 2014). It is possible that this was not the case (yet) in the first version of the Verspillingschecker.

Another difficulty is that energy saving in households can come with unexpected side effects, which makes finding absolute effects difficult. For instance, when buying a new energy efficient refrigerator, the old one sometimes ends up in the basement as a second fridge. Insulation of one's home can lead to using and heating more rooms and thereby reducing the expected impact; a so-called rebound effect. And a field experiment by Tiefenbeck, Staake, Roth and Sachs (2013) aimed at saving water, showed that people in the experimental group indeed decreased their water use, but at the same time their electricity use increased . This is an example of a negative spillover effect that could possibly be prevented by presenting net total effects and help users focus on that.

Another point of discussion is related to the negative impact of the 'positive' feedback to efficient users. This is an effect more often found and can be tackled: A well-known study by Schultz, Nolan and Cialdini (2007) for a American utility called OPower showed already that telling low-consuming households they had performed better than average, increases energy consumption of those households. This potential destructive effect of providing positive feedback about what one is doing compared to similar others, was eliminated with the addition of an 'injunctive' message. When a smiley face (☺) was added as a social approval of the performance, low-consuming households did not start to increase their energy use. By 'simply' adding a happy emoticon to the households that did better than average, and a frown (☹) to households that did worse than average, all households started to decrease their energy usage. This is something that can be incorporated in the next version of the Verspillingschecker as well.

At last discussed the net effect of Toon. As differences in energy saving percentages between the groups were almost zero, one could ask whether there is an effect of Toon at all. We argue this cannot be known from this experiment as we do not know energy consumption data from the period before the test groups had installed a Toon. It could be that the impact of Toon had already taken place before this experiment. In a follow-up experiment we would recommend to include a Toon-only group experiment and to include energy data from before the Toon was installed to further investigate the impact of Toon itself as well.

Contents

1 Introduction

1.1 What is the Verspillingschecker?

The Verspillingschecker (Energy Waste Checker) is an application developed by Quby and launched at the Dutch Toon smart thermostat platform. The app generates automated individual insights about where energy is wasted in a user's home, and advice about how to reduce this.

The app includes seven specific so-called use cases (showerhead, thermostat, refrigerator, washing machine, dishwasher, dryer, standby use) and 3 overall energy use cases (gas for heating home; gas for heating water and electricity use). Based on patterns in the energy data and additional calculations Quby provides diagnoses to the user: people get to see whether a specific usage is efficient or not and what a user could do about that (advice). For the seven use cases people receive either a green (efficient) or red (inefficient) message. In this report we call this message a flag or a diagnosis. In most cases, the app asks use-case specific questions to a user before giving the advice. For instance, a question about the energy label of an appliance that has been detected as inefficient.

The app is visible at the thermostat display of Toon itself and visible online via the Toon application that users can download at their devices of choice (smartphone, lap/desktop, tablet).

Figure 1 The Verspillingschecker is one of the services Quby develops to fulfil the goal of Eneco to act as a service provider for Eneco clients. In the figure the three domains are depicted: heating, cooling and electricity.

1.2 Aim and research questions

The main aim of this research was to answer two following research questions (as proposed in the request for a DEI subsidy):

1. Can automatically generated personalized advice realize energy savings per household of 5-10%?
2. How effective is automatically generated personalized advice on energy efficiency measures in households?

While the first question is very specific and leaves minor room for discussion, the second research question is less specific: it is unspecified what is meant with ‘effective’ and with ‘measures’. Therefore we used the definitions below.

Effective: We take a broader range of effects into account in line with the interest of Quby in the impact of the Verspillingschecker on actions and indicators for future actions: We proposed to take the AIDA (attention, interest, desire, action) spectrum into account (a well-known model amongst marketeers and as such also known by Quby employees; see figure 2), as well as attitudes towards saving energy and attitudes towards the specific actions that are advised by the Verspillingschecker. See data collection method in section 2.4.2 for a description of how we measured this.

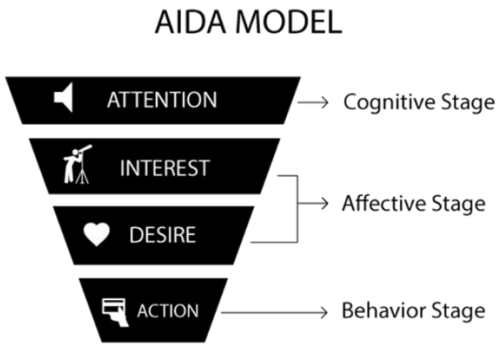

Figure 2 AIDA Model

Measures: Regarding measures, most often this refers to energy efficient investments; one-time actions that have a lasting effect. We proposed to add small investments (such as buying an efficient shower head), setting-changes (such as using the ecomode of appliances or changing the night temperature to a standard of 15 degrees Celsius, for example) and habitual behaviour changes (such as washing at low temperatures, lowering the thermostat when leaving the home) to this definition as well, as these changes do not have an investment barrier, and are targeted by the Verspillingschecker as well.

In addition Quby wanted to know the impact of the Verspillingschecker on a use-case specific level, as well as on indicators for future change. The use-case specific level gives additional insights on how to further improve the impact of the application. The indicators of future change are of interest as not all advices will most likely be translated into actions that can be measured before the end of the research period possible under DEI subsidy (granted to develop and test the Verspillingschecker app). We therefore included the impact of Verspillingschecker on use-case specific indicators as well as indicators for future action, being attention, interest, desire and attitudes towards saving energy and the specific actions advised by Verspillingschecker.

With this information the overall aim is to further improve automatically generated personalized advice which can contribute to energy savings in households.

2 Methodology

2.1 Basic design

The first research question (Can automatically generated personalized advice realize energy savings per household of 5-10%?) could only be answered reliably by measuring real energy usage of Verspillingschecker users over a longer period of time (**experimental group**) and compare this with a **control group**. In addition, *as well as* to answer the second question, and to gain support for the underlying assumptions of any effects found, both Verspillingschecker users and control group member filled out a **questionnaire** about themselves and the topics of this study.

Group comparisons were performed to isolate the change (before and after Verspillingschecker advices have been seen by a user) from any possible 'natural' change such as weather effects, the impact of events, national or international energy saving campaigns, etc.

To rule out the net impact of Toon itself, the control group ideally is a group that has a Toon without the Verspillingschecker. Unfortunately this group was not available as the Verspillingschecker was made available to all Eneco clients with a Toon. After exploration with Quby, the next best available control group option was a smart meter users group without a Toon. We thus not measure the pure effect of the Verspillingschecker app, but measure the effect of the Verspillingschecker app *plus Toon*. See the discussion for implications of this situation.

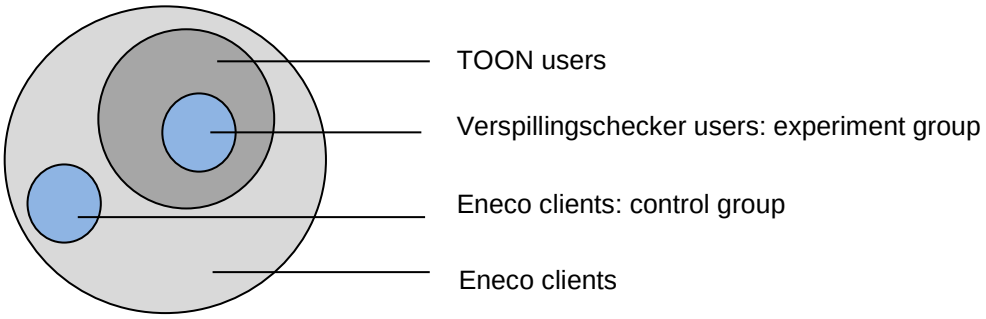

Figure 3 Choice of experiment and control group among Eneco clients

2.2 Hypotheses

For the two research questions we generated a hypothesis about the impact of the Verspillingschecker on a user, as well a hypothesis about whether this difference is stronger than whatever happens in a control group.

Hypothesis: Usage of the Verspillingschecker app changes the amount of overall energy use with 5-10%.

1. The group that used the Verspillingschecker application (test group) will on average show a 5-10% stronger change in energy usage than the control group (that has no application and no Toon).

2. Relative energy use will change with 5-10% over time after subjects in the test group have used the Verspillingschecker application.

Hypothesis: A use case specific advice (these are energy-efficiency related diagnoses in the Verspillingschecker app, marked with a red or green 'flag') has a significant impact on the related use case specific behaviour.

3. An advice has an impact on the behaviour addressed.
4. The type of advice (Red or Green) has an impact on the behaviour addressed.

Hypothesis: The Verspillingschecker has an impact on attention, interest and desire to take measures (see our definition of 'measures' written above), the attitude towards measures and the amount of measures taken.

5. Attention, interest and desire to take measures, attitude towards measures and the amount of measures taken changes significantly after subjects in the test group used the Verspillingschecker application.
6. The group that used the Verspillingschecker application (test group) will on average show a stronger change in attention, interest and desire to take measures, attitude towards measures and the amount of measures taken, than the control group,that has no application and no Toon.

Besides the hypotheses based on the research questions, four assumptions about how these effects are generated were included in the subsidy request granted under DEI. These four assumptions, which are also hypotheses, have been tested too:

e. The amount of data collected by the Verspillingschecker application is adequate for detecting inefficient energy usage by the Verspillingschecker.
f. The generated personal advice offers home owners sufficient insight in their energy conservation potential.
g. The generated personal advice provides home owners insight in what energy conservation measures are the most interesting for them.
h. The generated personal advice helps home owners in taking investment decisions concerning energy conservation measures.

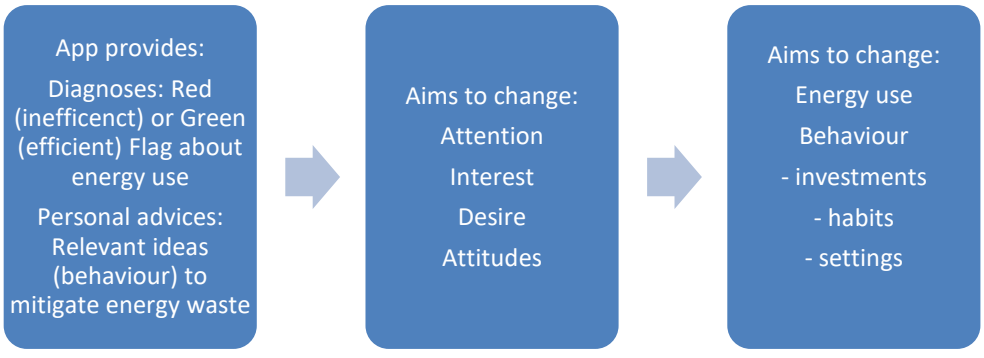

Figure 4 Schematic presentation of the behavioural effect of the Verspillingschecker

2.3 Participant characteristics

All Verspillingschecker users with a smart meter and without PV were selected. The app was only available in the Netherlands. We selected smart meter users only to be able to exclude the impact of generated electricity (PV users).

To minimize any further possible alternative explanation for differences between these groups, we selected the control group based on the same mix of profile characteristics as the Verspillingschecker group, being smart meter clients with the same mix of house type (corner house, flat, etc.) and household size (number of people in household). These latter two profile characteristics have the strongest impact on gas versus electricity use respectively.

Because house type (corner house, flat, etc.) and household size are such important impact factors, and data reliability about these factors was insecure, we also asked participants to give us these characteristics via the questionnaire used in this study. Because of the differences between what was reported by them and what was registered by Quby and Eneco, we decided to select the users that filled out the questionnaire and gave us permission to link their questionnaire data to their energy usage data to answer the first research question (1. Can automatically generated personalized advice realize energy savings per household of 5-10%?). We excluded some users that did not answer a single question in the app. The group that was left answered at least one question in the Verspillingschecker app and had filled out the questionnaire. We think these users can prove best what the impact is of the Verspillingschecker, because we know they have actively interacted with the app and we can control for household size, house type, home ownership and/or any other characteristic that appears to have an impact.

Every of the 1985 Verspillingschecker users, were randomly matched with a user in the control group based on household size and house type. As a result, pairs were created resulting in two groups similar in terms of household size and house type, see the Table below.

For research question 1, after this matching process, removing missing matches (n=196) and outlier exclusion, we included in total 2,680 participants: In the Verspillingschecker group 1,242 users with a smart meter participated. In the Control group 1.438 households with a smart meter were included. The so-called fuzzy process made it possible that a Verspillingschecker user was matched with In the table below you see how we came down to the number of people per group that participated. Table 3 describes the main characteristics. More details can be found in Annex A.

Table 3 Overview of people per group participating in the research

	Energy use (m^3, kWh) per week, 21 weeks (Nov-April)	Questionnaire (demographics + attention, attitude, behaviours), (May)	Questionnaire (demographics + attention, attitude, behaviours) with permission to link data	Outlier exclusion + Matching users (on household size and house type)
Verspillingschecker group	53.931	2.534	1.985	1.242
Control group*	38.868	3.234	1.873	1.438

* Eneco clients without Toon and Verspillingschecker app

Table 4 Main characteristics of the two groups for research question 1.

	Verspillingschecker group	Control group
Household size	Average 2,6 persons per household (44% 2-person households)	Average 2,6 persons per household (46% 2-person households)
House type	Mostly town houses (68%)	Mostly town houses (67%)
Home ownership	Mostly home-owners (73%)	Mostly home-owners (70%)
Surface area size	Mostly with a surface area of 100-125 m^2 (39%)	Mostly with a surface area of *unknown*
PV users	No PV (100%)	Mostly No PV (78%)*[1]
Building year	Building year 1975-1987 (27%)	Building year *unknown*
Gender	Men (75%)	Men (67%)
Age	On average 53 years old	On average 60 years old
Net Income level	On average €3,500-4,000 per month (35% did not want to indicate their income level)	On average €3,500-4,000 per month (35% did not want to indicate their income level)
Baseline electricity use	72 kWh per week	62 kWh per week
Baseline gas use	40 m^3 per week	40 m^3 per week
Environmental identity*[2]	Average of 4,9 on a 7 point scale	Average of 5,0 on a 7 point scale

*[1] For our analysis on electricity use we filtered these respondents out: from their actual use data we could not find out how much of their generated energy was used directly.

*[2] Based on three statements (reliability of the three statements is high): Saving energy is an important part of who I am (Energie besparen is een belangrijk deel van wie ik ben); I am the type of person that saves energy (Ik ben het type persoon dat energie bespaart); I see myself as an energy saving person" (Ik zie mijzelf als een energiebesparend person).

Regardless this matching process, the groups differ on some characteristics. For example, we found that the Verspillingschecker and the Control group differed systematically on their baseline electricity use; on average around 10 kWh per week. We checked various explanations but the difference was not caused by household size, house type or people's environmental identity. Eneco then

confirmed that Toon users on average use more energy. In the analyses we correct for these and other characteristics of significant impact, by inserting them as a so-called 'covariate' to correct for these systematic differences.

To answer research question 2, we included all respondents that filled out the questionnaire, because we did not have to link their questionnaire data to their energy usage for this question. In total we included 5,771 respondents: 2,534 in the Verspillingschecker group and 3,237 in the Control group. Table 5 describes the main characteristics. More details can be found in Annex A.

Table 5 Main characteristics of the two groups for research question 2.

	Verspillingschecker group	Control group
Household size	Average 2,5 persons per household (48% 2-person households)	Average 2,3 persons per household (53% 2-person households)
House type	Mostly town houses (64%)	Mostly town houses (58%)
Home ownership	Mostly home-owners (74%)	Mostly home-owners (70%)
PV users	Mostly no PV (81%)	Mostly No PV (77%)
Gender	Men (72%)	Men (63%)
Age	On average 54 years old	On average 61 years old
Net Income level per household	On average €3.500-4.000 per month (35% did not want to indicate their income level)	On average €3.500-4.000 per month (37% did not want to indicate their income level)
Environmental identity	Average of 4,9 on a 7 point scale	Average of 4,9 on a 7 point scale

2.4 Operationalisation of factors

2.4.1 Can automatically generated personalized advice realize energy savings per household of 5-10%?

Automatically generated personalized advice

In the Verspillingschecker group we included users that did see *at least one* diagnosis about their usage. We did know if and when this was the case, by their first time of use' registration. This is the date that is registered when a user answers a question that had been asked in the Verspillingschecker app after seeing a use case specific diagnosis (for example, a question following the diagnosis: Your fridge is inefficient/efficient). This way we secured at least some interaction with the application did happen.

N.B. Click data registrations were not available for the current research (frequency of use, seconds spend at the application) so we could not measure any further intensity of the interaction with the app.

Overall impact: energy savings (gas and electricity)

To calculate overall energy savings we determined energy consumptions for each household. We analysed consumptions on a week level -in line with the update frequency of Verspillingschecker data to Verspillingschecker users.

In the Verspillingschecker group, users started using the Verspillingschecker app in their own time; in different weeks. Therefore, for every user the first week of interaction with the Verspillingschecker was determined based on the first moment a user answered a question in the Verspillingschecker app. From that week on, the app could have an impact on their energy consumption. The weeks *after* this week we therefore labelled 'impact weeks' (defined as week I, II, III, IV, etc.). The average gas and electricity usage of the four weeks before the first week of interaction was labelled someone's 'baseline consumption'. The percental difference between the baseline consumption and the impact weeks has been calculated to determine overall energy savings (gas and electricity).

In the control group the same calendar weeks as the baseline consumption and impact weeks of the users in the Verspillingschecker group were taken in analyses to make sure time-effects were ruled out.

Use case specific behaviour change

To test the third and fourth hypothesis we looked at the data of the different use cases in the app. For each of 7 use cases (showerhead, thermostat, refrigerator, washing machine, dishwasher, dryer, standby use) and 3 overall energy usages (gas for heating home; gas for heating water and electricity use), use case specific calculations lead to diagnoses from Quby for the user. It is made known whether a specific usage is efficient or not and what a user could do about that (advice). For the 7 use cases people receive either a green (efficient) or red (inefficient) message (see example below, Figure 5).

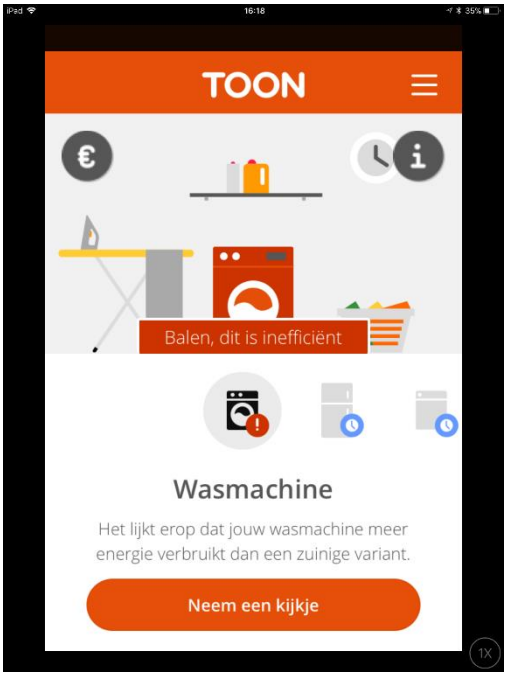

Figure 5 Example of a use case specific inefficient message (in Dutch)

Annex B contains the operationalization of impact per use case which we insert in our analyses at a use case level.

2.4.2 How effective is automatically generated personalized advice on energy efficiency measures in households?

The variables of interest were mainly operationalized by statements that users are asked to agree or disagree with on a 7 point scale (from totally disagree to totally agree). All variables were measured using two or more statements (see Annex C for the operationalization of all factors, and Annex D for the full questionnaires).

As a first step, a factor analyses and reliability analysis determined which statements *in fact* measured the same underlying factor AND together had an acceptable reliability (α >.70). The following factors appeared to exist, measured by the statements given underneath:

Attention for energy waste at home (Reliability of this scale is high; α = .875.)
This appeared to be a combination of items originally included to measure attention, interest, desire to act, and an item that was supposed to measure Assumption 2 ('insight in saving potential').

1. I now am more interested in the possibilities to prevent energy waste in my home than a year ago. (Ik heb nu meer interesse in mogelijkheden om energieverspilling in huis tegen te gaan dan een jaar geleden.)
2. I now have a stronger desire to prevent energy waste in my home than a year ago. (Ik heb nu een sterkere wil dan een jaar geleden om energieverspilling in mijn huis tegen te gaan.)
3. I pay more attention to energy waste in my home than a year ago. (Ik heb nu meer aandacht voor verspilling van energie in huis dan een jaar geleden.)
4. I now know better how much energy I can save in my home than one year ago (Ik weet nu beter hoeveel energie ik in mijn huis kan besparen dan een jaar geleden.)
5. An energy efficient home is higher on my wishlist than a year ago.(Een energiezuinig huis staat hoger op mijn 'verlanglijst' dan een jaar geleden)

Attitude towards specific measures/actions
We asked respondents to what extent they (dis)agreed to statements about specific measures/actions advised by the Verspillingschecker. The statements addressed the extent to which the action 'can help to prevent energy waste at home', 'is worth the effort', 'is a hassle', 'gives a good feeling', is seen as 'normal behaviour'. Together the scores represent the attitude about that specific measure/action. After a reliability check of the answers to the statements, some statements were removed to get the acceptable reliabilities presented below.

The measures/actions for which attitudes have been measured by the questionnaire, are:

- Replacing a still functioning fridge, dryer, washing machine or dishwasher for a more efficient one (Reliability of this scale is; α = .714.)
- Using the Eco mode (or energy-efficient program) of appliances (Reliability of this scale is; α = .739.)
- Washing at low temperatures (Reliability of this scale is; α = .741.)

- Switching appliances off instead of using stand-by mode (Reliability of this scale is; α = .821.)
- Efficient hot water use in the shower (Reliability of this scale is; α = .770.)
- Lowering thermostat when leaving home (Reliability of this scale is; α = .723.)
- Lowering thermostat at night (Reliability of this scale is; α = .772.)

Experienced help from Eneco (Reliability of this scale is high; α = 875.)
This is a scale with a combination of items originally included to measure 'interest in actions' versus 'insight in most relevant measures to take' (assumption 4). After factor analysis, we labeled this scale 'experienced help from Eneco' as the word Eneco appears to be leading. This can be any kind of experienced help from Eneco about energy conservation/preventing energy waste at home.

1. Eneco helps me to limit wasting energy.
2. Eneco provides me insight in the benefits of energy saving activities or measures.
3. Eneco has triggered my curiosity regarding measures to prevent waste in my home.

Realized actions to decrease energy waste
For both groups we summed the reported number of actions taken (daily (habitual) behaviours, setting changes, small investments) and the reported number of appliances bought (investment behaviour) in 2018, to measure the amount of actions taken per household.

2.5 Procedure

2.5.1 Verspillingschecker gradual roll out

The Verspillingschecker was launched in December 2017. Users could download a new version of their Toon application wherein the Verspillingschecker was made available.

Users did not receive all advices at once. These were released gradually over a few months. The Verspillingschecker needs time to read and assess the data before it provides a specific advice. The first use cases were released in December 2017. The last use case was released in February 2018. In total 7 appliance-specific and 3 overall use cases were available during the measurement period of this research.

- First time of interaction (week of 04-12-2017): dryer, washing machine, fridge, dishwasher, stand-by usage, gas for hot water (overall), electricity (overall), gas for heating home (overall)
- First time of interaction (22-12-2017): thermostat
- First time of interaction (23-02-2018): showerhead.

During all times a user could see which cases were already available to him/her and which cases would be available to him/her in the future.

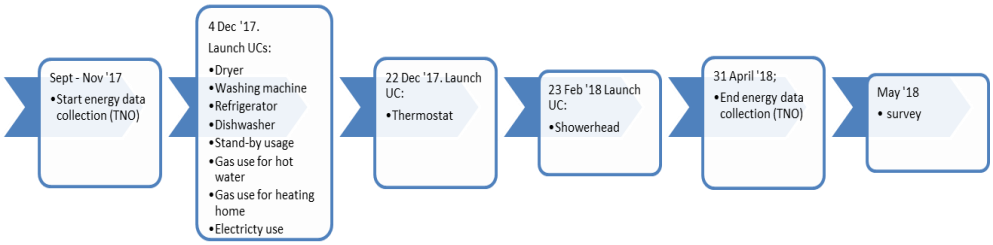

Figure 6 Timeline research project

2.5.2 Questionnaire

After we finished collecting energy consumption data Quby and Eneco facilitated to send out a questionnaire to both Verspillingschecker users and control group members about themselves and the topics of this study. The questionnaire was sent via email to the two groups (20.000 in each group) at the beginning of May 2018. We did not provide an incentive (like money or a present) in return for their participation. After one week a reminder was sent out. We aimed at 2.000 respondents for both the Verspillingschecker group and the control group.

After two weeks 2.534 Verspillingschecker users (13% response) had filled in the questionnaire and within this group 1.985 (78%) gave permission to link their data to their energy usage data (see also Table 3). The control group was somewhat more difficult: initially 20.000 respondents were emailed with the request to fill in the questionnaire. 2.247 (11% response) participated and 1.077 (48%) consented to link their data. We then decided to approach another 10.000 of the 38.868 people in the control group to increase the number. They again were sent a reminder after one week. This led to the total control group of 3.234 participants. Of this total group 1.873 (58%) gave permission to link their data.

The questionnaires were distributed with help of research agency Totta. Totta works together with Eneco regularly and was used to their type of communication and with Eneco's procedures for sharing privacy sensitive information.

3 Results

In paragraph 3.1 the results of tests done to answer research question 1 (Can automatically generated personalized advice realize energy savings per household of 5-10%?) are shown: overall impact of the Verspillingschecker on energy usage (3.1). We provide a more detailed view on the impact of the diagnoses per use-case on the related specific behaviour in paragraph 3.2.

In paragraph 3.3 the test results for answers to research question 2 (How effective is automatically generated personalized advice on energy efficiency measures in households?) and the related assumptions are shown.

3.1 Overall impact

For research question 1 (Can automatically generated personalized advice realize energy savings per household of 5-10%?), we start with the presentation of the weekly electricity and gas usages per group to give an idea of how these usage patterns evolved over time.

1) Weekly usages:
Below we present the overall electricity and overall gas consumption graphs of week -IV to -I (baseline consumption weeks) versus week I – XIV (impact weeks). After 14 impact weeks the number of users that could be included in analyses dramatically declined. Note that all weeks can be in a different period of the year for each user, since users did not start using the application at the same time.

a) Overall electricity:

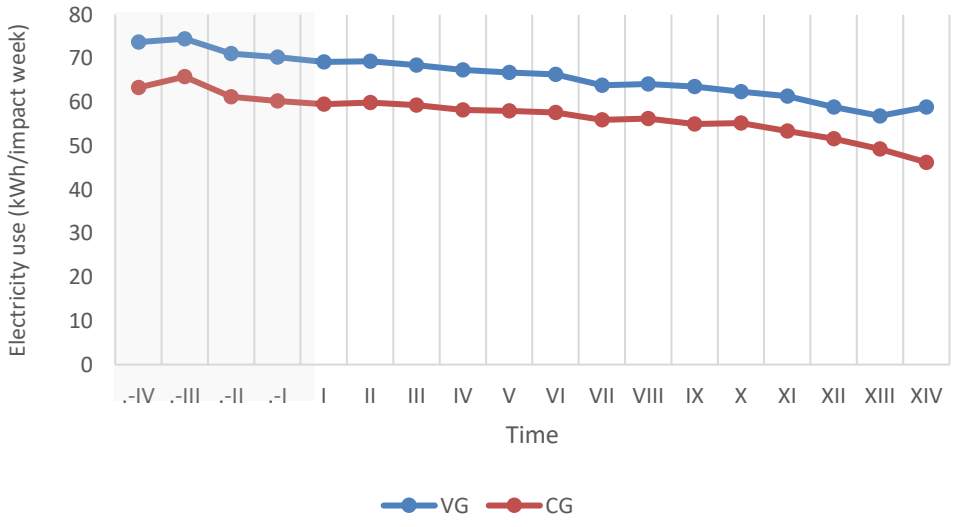

Figure 7 Electricity use per impact week (in kWh). VG = Verspillingschecker Group, CG = Control Group (no Toon).

Figure 7 shows the amount of energy used (Y-axis) per impact week (X-axis). Dashed (grey) weeks represent the baseline consumption weeks, before the Verspillingschecker had been used.

b) Overall gas:

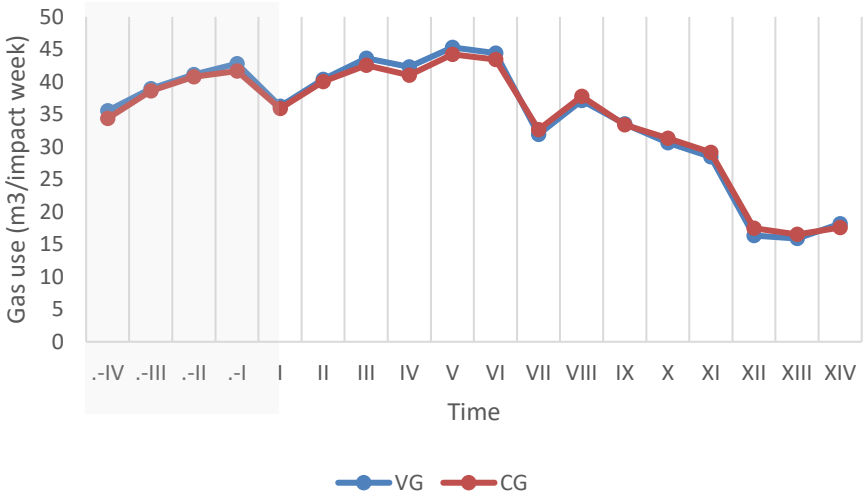

Figure 8 Gas use per impact week (in m^3). VG = Verspillingschecker Group, CG = Control Group (no Toon).

Figure 8 shows the amount of energy used (Y-axis) per impact week (X-axis). Dashed (grey) weeks represent the baseline consumption weeks; before the Verspillingschecker had been used.

2) Saving percentages:

As a next step we created saving percentages (comparison with baseline usage), see paragraph 2a and 2b for weekly saving percentages and differences in patterns between groups. We summarize and test total percental impact (short term and longer term saving percentages) in 2c.

1. Short term (first impact week versus first 4 impact weeks. The number of 4 weeks has been chosen based on the amount of weeks that are included in baseline consumption being 4 as well)
2. Longer term (first 14 impact weeks. After 14 impact weeks the number of users that could be included in analyses dramatically declined because of dropout.

a) Overall Electricity:

The analyses of differences between groups for electricity saving percentages showed that groups did not differ from each other. There was no difference between the impact of the Toon + Verspillingschecker on overall weekly electricity saving percentages and the impact of no Toon in the control group. This is shown in Figure 9. See Table 6 for means and standard deviations for all weeks.

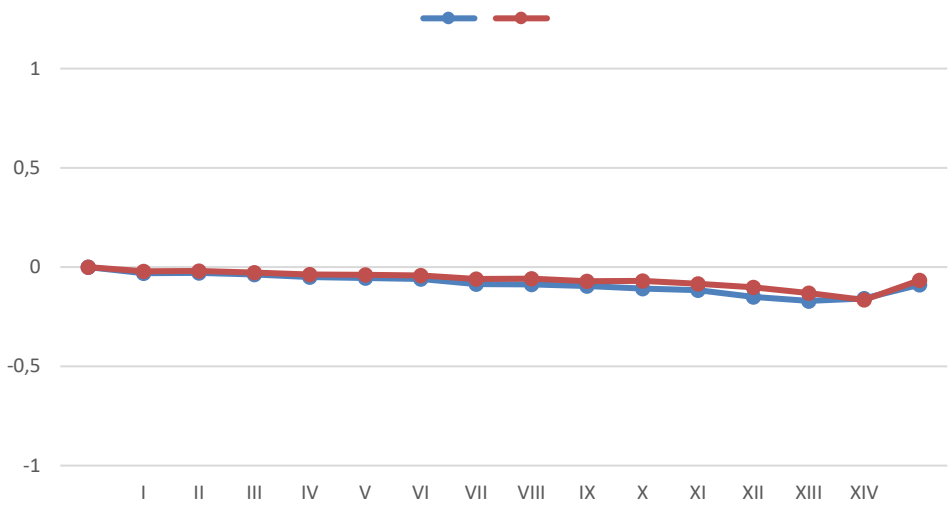

Figure 9 Average effect on electricity use (kWh/week) for Verspillingschecker (VG) and Control group (CG). (Percental difference with the baseline).

How did we test this?
A 1×2 repeated measures ANOVA with Group (Toon + Verspillingschecker vs. No Toon) as between-subject factor, weekly savings as within subjects factor and percentage energy saved as dependent variable revealed no effect of Group over time, F(1,1442)=1.804, p=.179. Baseline Electricity consumption and Household size were controlled for as these were inserted as covariates (independent factors).

Table 6 Means and standard deviations for percentages electricity conservation relative to baseline per week and overall over 14 weeks

		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	MEAN
VG	***M***	**-0,03**	**-0,03**	**-0,04**	**-0,05**	**-0,06**	**-0,06**	**-0,09**	**-0,09**	**-0,10**	**-0,11**	**-0,12**	**-0,15**	**-0,17**	**-0,16**	**-0,09**
	SD	0,16	0,16	0,18	0,19	0,20	0,28	0,23	0,20	0,18	0,20	0,21	0,22	0,24	0,23	0,21
CG	***M***	**-0,02**	**-0,02**	**-0,03**	**-0,04**	**-0,04**	**-0,04**	**-0,06**	**-0,06**	**-0,07**	**-0,07**	**-0,08**	**-0,10**	**-0,13**	**-0,17**	**-0,07**
	SD	0,13	0,17	0,17	0,16	0,19	0,19	0,17	0,18	0,18	0,15	0,15	0,16	0,21	0,23	0.17

b) Overall Gas:

The analyses of differences between groups showed that groups did not differ from each other over time for gas conservation. Toon + Verspillingschecker did not lead to more weekly gas conservation than no Toon. This is shown in Figure 10. See Table 7 for means and standard deviations for all weeks.

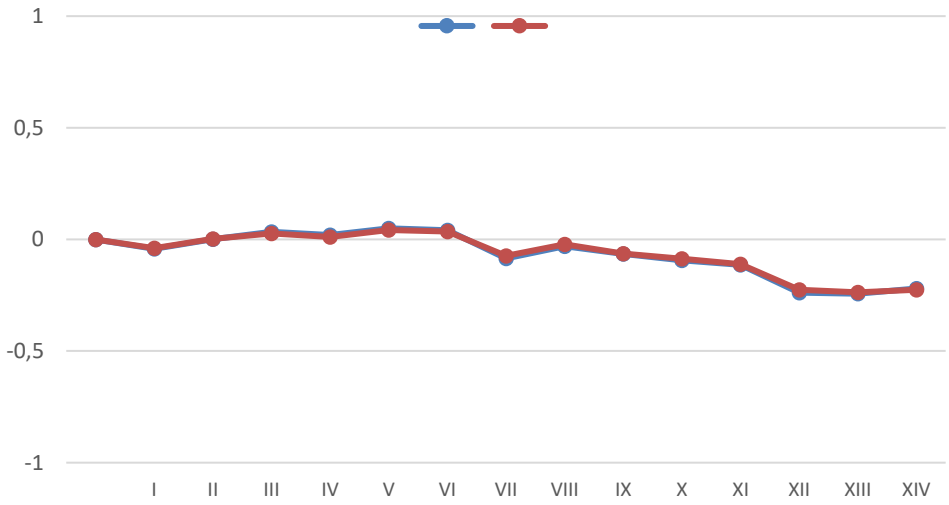

Figure 10 Average effect on gas use (m^3/week) in Verspillingschecker (VG) and Control group (CG). (Percental difference with the baseline).

How did we test this?
A 1×2 repeated measures ANOVA with Group (Toon + Verspillingschecker vs. No Toon) as between-subject factor, weekly gas savings as within subjects factor and percentage gas saved as dependent variable revealed no effect of Group: Toon + Verspillingschecker did not lead to more gas conservation than no Toon, F(1,1605)=0.426, p=.514.

Baseline Gas consumption, Household size and House Type were controlled for as these were inserted as covariates (independent factors).

Table 7 Means and standard deviations for percentages gas conservation relative to baseline per week and overall over 14 weeks

		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	MEAN
VG	***M***	**-0,04**	**0,00**	**0,04**	**0,02**	**0,05**	**0,04**	**-0,08**	**-0,03**	**-0,06**	**-0,09**	**-0,11**	**-0,24**	**-0,24**	**-0,22**	**-0,07**
	SD	0,11	0,11	0,14	0,14	0,19	0,25	0,15	0,17	0,15	0,18	0,26	0,17	0,19	0,17	0,17
CG	***M***	**-0,04**	**0,00**	**0,03**	**0,01**	**0,04**	**0,04**	**-0,07**	**-0,02**	**-0,06**	**-0,09**	**-0,11**	**-0,23**	**-0,24**	**-0,22**	**-0,07**
	SD	0,10	0,10	0,14	0,13	0,18	0,25	0,14	0,17	0,14	0,17	0,25	0,18	0,21	0,19	0,17

c) Short term impact

The impact of the Verspillingschecker (VC) in the **first week** is *nonsignificant* for electricity and for gas usage:

- In the VC group a **0,01%** stronger decline than in the Control group can be seen for electricity use.

One-tailed T-tests comparing the means (the mean saving percentage of week 1) of the Verspillingschecker group (M=-0.03, SD=0.16) and the control group (M=-0.02, SD=0.13) showed an insignificant difference, t(2135)=1.47, p=.141.

- In the VC group a **0,00%** stronger decline than in the Control group can be seen for gas use.

One-tailed T-tests comparing the means (the mean saving percentage of week 1) of the Verspillingschecker group (M=-0.04, SD=0.11) and the control group (M=-0.04, SD=0.10) showed an insignificant difference, t(2380)=0.403, p=.687.

The impact of the Verspillingschecker (VC) in the **first 4 weeks** is *significant for electricity* and *nonsignificant for gas usage*:

- In the VC group a **0,01%** stronger decline than in the Control group can be seen for electricity use.

One-tailed T-tests comparing the means (the mean saving percentage of week I-IV) of the Verspillingschecker group (M=-0.04, SD=0.15) and the control group (M=-0.02, SD=0.11) showed a significant difference, t(2131)=2.18, p=.030. The effect should be labelled as a 'small' effect (Cohen's *d*= 0.15).

- In the VC group a **0,00 %** stronger decline than in the Control group can be seen for gas use.

One-tailed T-tests comparing the means (the mean saving percentage of week I-IV) of the Verspillingschecker group (M=-0.00, SD=0.10) and the control group (M=-0.00, SD=0.09) showed an insignificant difference, t(2375)=-0.771, p=.441.

d) Longer term impact

The impact of the Verspillingschecker in the first 14 weeks is *significant for electricity* versus *nonsignificant for gas usage*:

- In the VC group a **0,03%** stronger decline than in the Control group can be seen for electricity use.

One-tailed T-tests comparing the means (the mean saving percentage of week I - XIV) of the Verspillingschecker group (M=-0.08, SD=0.16) and the control group (M=-0.06, SD=0.12) showed a significant difference, t(1447)=3.65, p=.000. The effect should be labelled as a 'small' effect (Cohen's *d*= 0.14).

- in the VC group a **0,00%** stronger decline than in the Control group can be seen for gas use.

One-tailed T-tests comparing the means (the mean saving percentage of week I - XIV) of the Verspillingschecker group (M=-0.04, SD=0.08) and the control group (M=-0.04, SD=0.07) showed an insignificant difference, t(1611)=0.320, p=.749.

With the results we tested the hypothesis: Usage of the Verspillingschecker app changes the amount of overall energy use with 5-10%.

1. The group that used the Verspillingschecker application (test group) will on average show a 5-10% stronger change in energy usage than the control group (that has no application and no Toon).
2. Relative energy use will change with 5-10% over time after subjects in the test group have used the Verspillingschecker application.

Both hypotheses were not confirmed.

Usage of the Verspillingschecker app changes the amount of overall energy use with less than 5-10%.

3.2 Use case-specific impact

Per use case, the only diagnosis (red=inefficient and green=efficient) that we are sure of people saw, was at their first time of interaction with that use case. For the other weeks that followed, we cannot be sure whether people saw the diagnoses or not. In addition, for each respondent we defined the first, second, third etc weeks of use of the app (namely, week I, II, III, etc.). Moreover, we defined their baseline consumption as the value of the week before the start week.

In our analysis we wanted to take into account as many weeks as possible. We calculated, given the number of weeks, number of groups, effect size and desired power, how many respondents (who also filled in the survey) we needed to include, in order to provide reliable results. We compared this to our data and the decided how many weeks we could include in our analysis. This therefore differs between the use cases; from five weeks for the shower head to 15 weeks for the refrigerator and washing machine use cases.

For all seven use cases we tested if any advice in the app has a significant impact on the related use case specific behaviour (Hypothesis 3) and whether the type of advice (Red of Green) has an impact on user behaviour (Hypothesis 4).

The effect of any advice and of specific advice (red or green) on use case behaviour was tested over time in a within-subjects comparison (T0, T1, T2, T3, etc), using GLM repeated measures analysis. In this analysis, the different time measurements represent the different levels of the factor time. In this analysis we include a between subjects factor (colour of the advice) to see whether differences over time differ for users that received a green or a red advice.

3.2.1 Use case 11: washing machine

The washing machine use case was introduced at the beginning of December 2017 (week 16). During our measurement period in total 38.592 participants started using this use case, with peaks in December ’17 (week 18) and January ’18 (week 23); see Figure 11.

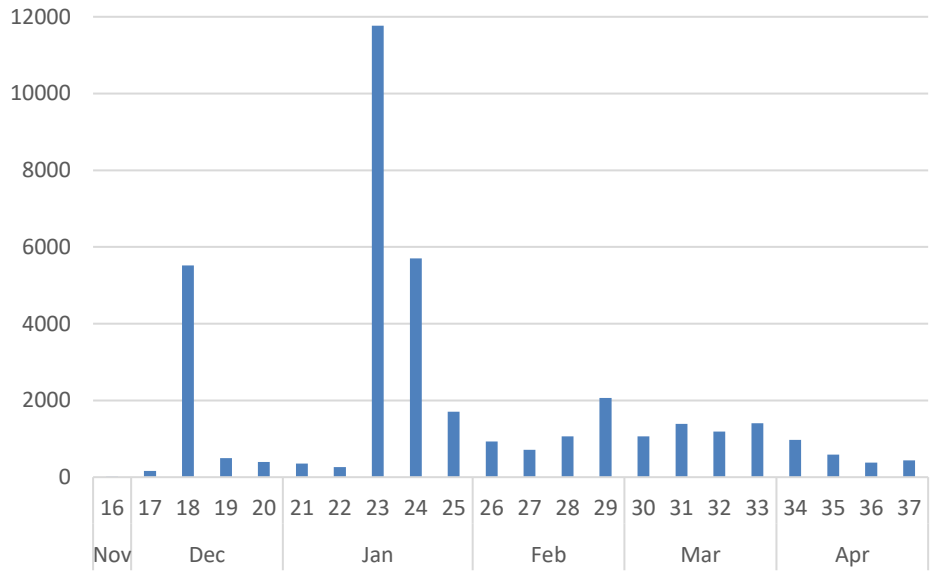

Figure 11 Number of respondents per start week

The red or green advice was based on the duration of the heating element in the washing cycle (in minutes), for a 3 month period. We did not look further at the exact threshold at which people receive either a red or green flag.

When looking at the group of respondents that had access to this use case for a minimum of fifteen weeks (n=18.199), at the start week 48% of them received a red flag and 52% received a green flag (see Figure 12). In week XV, 52% received a red flag and 48% a green one. So over time the share of red flags increased somewhat.

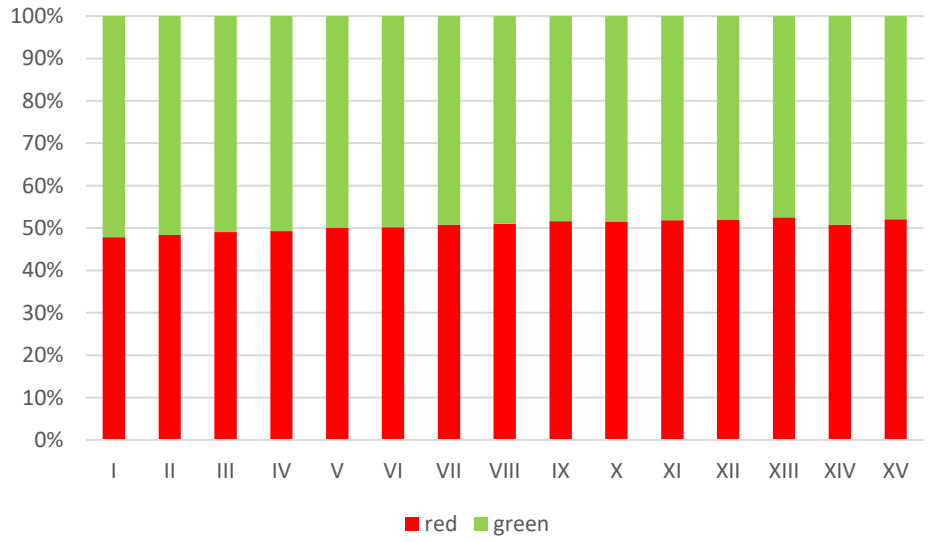

Figure 12 Share of red/green flags per week (in %).

When zooming in at changes per respondent we find a comparable result (see Table 8). The majority of the people with a red flag at week I, also receives a red flag at week XV: 37% of the total switches. This also holds for the green flags: 37% of all respondents receive a green flag at both times. In addition we see that more people (16%) made a switch from green to red, than from red to green (11%).

Table 8 Number of switchers from Red to Green and vice versa from week I until week XV

		Week XV		
		Red	**Green**	**Total**
Week I	**Red**	6.633 37%	2.039 11%	8.672
	Green	2.812 16%	6.652 37%	9.464
	Total	9.445	8.691	18.136

Finally, we statistically tested the effect of the flag people received and saw at week I. First, we wanted to test the effect of any advice (being it Red or Green) on the 'duration of the heating element' (H3). We therefore performed a 1x1 repeated measures ANOVA with 'time (I, II, III, …, XV)' as a within subjects factor and 'duration of the heating element' as the dependent variable. We found that time indeed affected the duration of the heating element. F(14, 94388)= 54,953,p=.000. Over time the duration of the heating element increased somewhat (see Figure 13), but the size of this effect is rather small (η^2=.008).

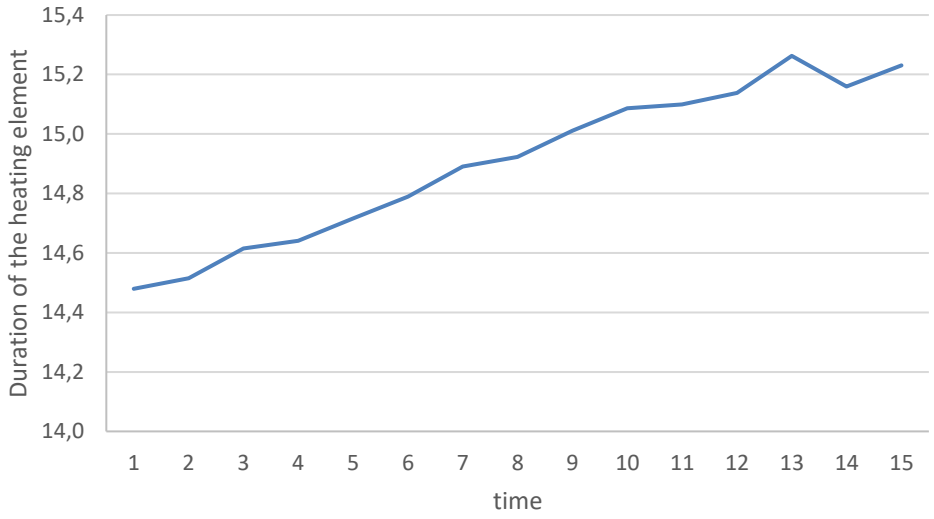

Figure 13 Development of the duration of the heating element (in minutes) over 15 weeks.

To check whether the colour of the flag had an impact on the duration of the heating element over time (Hypothesis 4), we split the users into two groups: users that started with a red flag versus green flag. A 1×2 repeated measures ANOVA with 'Flag colour at start week' (Red or Green) as between-subject factor, 'time (I, II, III, …, XV)' as within subjects factor and 'duration of the heating element' as dependent

variable revealed an effect of Flag colour, F(1, 6741)= 5180,232, p=.000. Users that received a red flag showed a decrease in the duration of heating element over time (see Figure 14). Users that received a green flag showed an increase in their duration of the heating element. The effect size is quite large (partial η^2=.435).

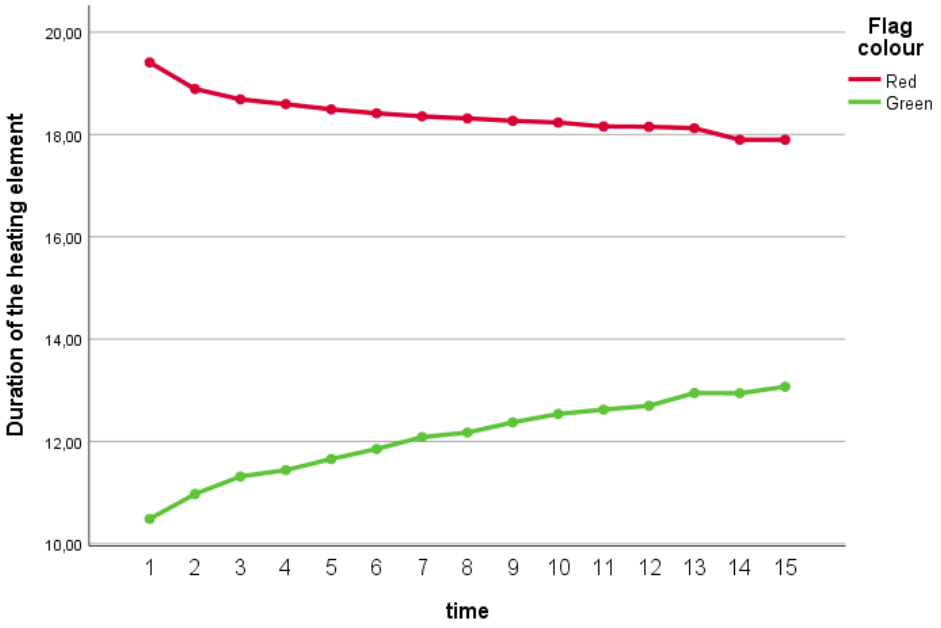

Figure 14 Development of duration of the heating element (in minutes) over 15 weeks, for red and green flag groups.

Use case specific behaviours from the survey

In the survey we asked people about specific behaviours advised by the app and if people performed these in 2018. We expected that more people with a red flag at the first week would start performing these behaviours in 2018 (i.e. in the period after they installed the app). It showed that indeed more people with a red flag stated to have bought a new washing machine (13%) than in the group who received a green flag (10%).

Table 9 Number of people buying a new washing machine in 2018 per flag group

	No	Yes	Total
Red	697	103	800
	40%	6%	46%
	87%	13%	100%
Green	839	94	933
	48%	5%	54%
	90%	10%	100%
Total	1.536	197	1.733
	89%	11%	100%

It also showed that more people who received a green flag washed at 30 degrees (see Table 10). This difference could however also be due to older habits: people who have the habit of washing at 30 degrees would do this in 2018, but would also have done this in 2017 thus receiving a green flag.

Table 10 Number of people washing at 30 degrees per flag group

	No	Yes	Total
Red	437	363	800
	25%	21%	46%
	55%	45%	100%
Green	471	462	933
	27%	27%	54%
	51%	50%	100%
Total	908	825	1.733
	52%	48%	100%

3.2.2 Use case 10: refrigerator

The refrigerator use case was introduced at the beginning of December 2017 (week 16). During our measurement period in total 32.757 participants started using this use case, with peaks in December ’17 and January ’18 (see Figure 15).

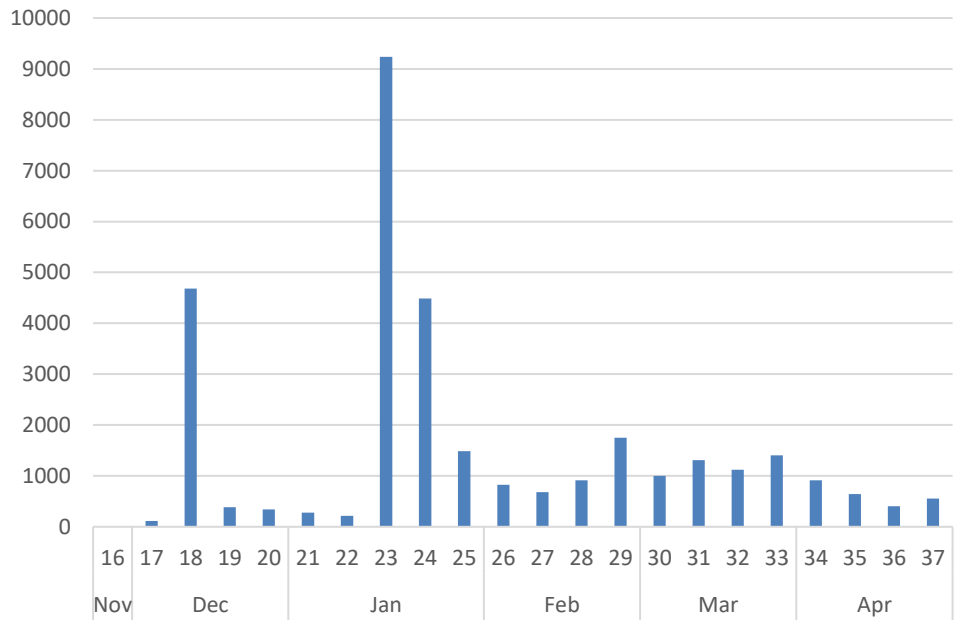

Figure 15 Number of respondents per start week

The red or green advice was based on the indicator load duration: the time that the compressor is running (in minutes), over a 3 month period.

When looking at the group of respondents that had access to this use case for a minimum of fifteen weeks (n=15.250), at the start week 27% received a red flag and

73% received a green flag (see Figure 16). In week XV, 26% received a red flag and 74% a green one. So over time the shares of red and green flags remained almost alike.

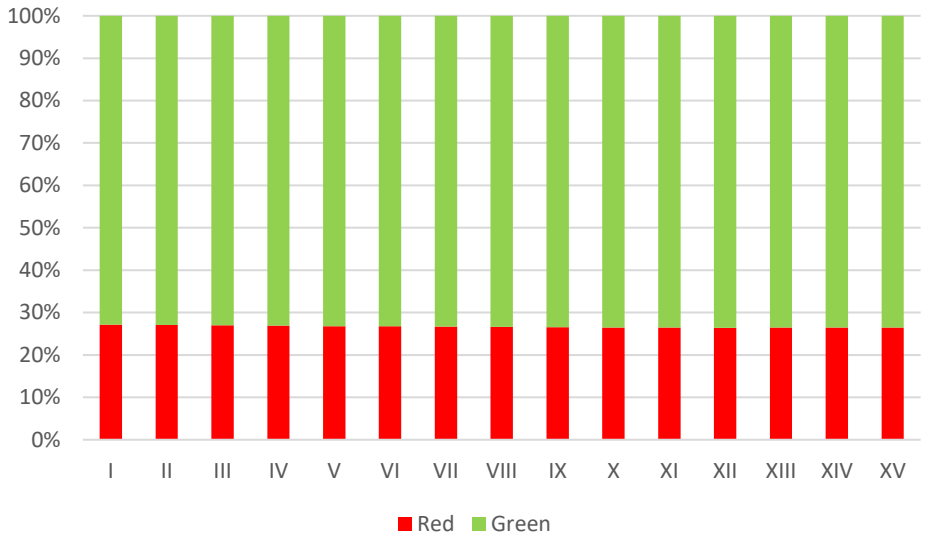

Figure 16 Share of red/green flags per week (in %).

When zooming in at changes per respondent we find a comparable result (see Table 11). The majority of the people with a red flag at week I, also receives a red flag at week XV: 24% of the total switches. This also holds for the green flags: 70% of all respondents receive a green flag at both times. In addition we see that a small group switches from red to green (3%), and a small group changes from green to red (2%).

Table 11 Number of switchers from Red to Green and vice versa from week I until week XV

		Week XV		
		Red	**Green**	**Total**
Week I	**Red**	3.538 24%	467 3%	4.005
	Green	360 2%	10.369 70%	10,729
	Total	3.898	10.836	14.734

Finally, we statistically tested the effect of the first flag people received and saw. First, we wanted to test the effect of any advice (being it Red or Green) on the 'load duration' (H3). We therefore performed a 1x1 repeated measures ANOVA with 'time (I, II, III, …, XV)' as a within subjects factor and 'load duration' as the dependent variable. Load duration did significantly change over time (see Figure 17), $F(14, 201264)= 17.041, p=.000$. The effect size is however small (partial $\eta^2=.001$).

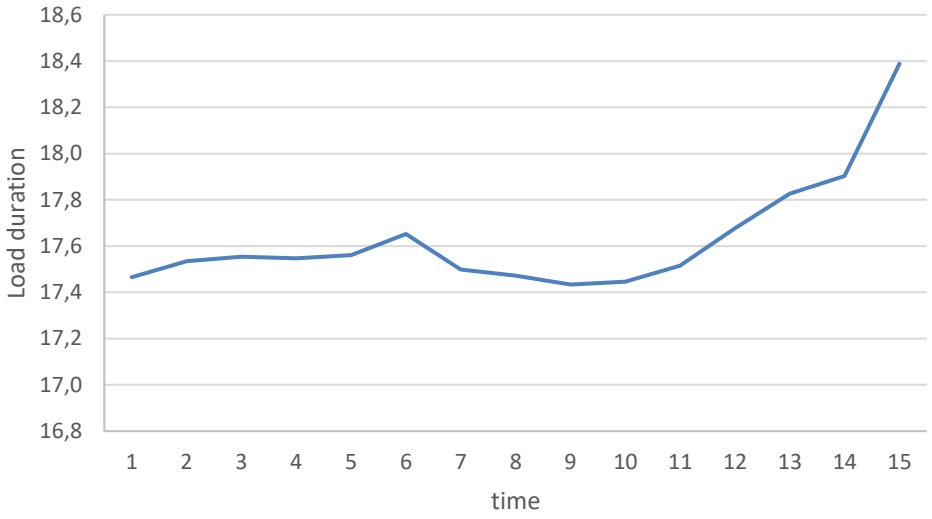

Figure 17 Development of the load duration (in minutes) over fifteen weeks.

To check whether the colour of the flag had an impact on the load duration over time (H4), we split the users into two groups: users that started with a red flag versus green flag. A 1×2 repeated measures ANOVA with 'Flag colour at start week' (Red or Green) as between-subject factor, 'time (I, II, III, …, XV)' as within subjects factor and 'load duration' as dependent variable revealed an effect of Flag colour, $F(1, 13891)= 945.513$, $p=.000$. Users that received a red flag showed a small decrease in duration of heating element over time (see Figure 18). Users that received a green flag showed an increase in duration of heating element. The effect size is rather small ($\eta^2=.064$).

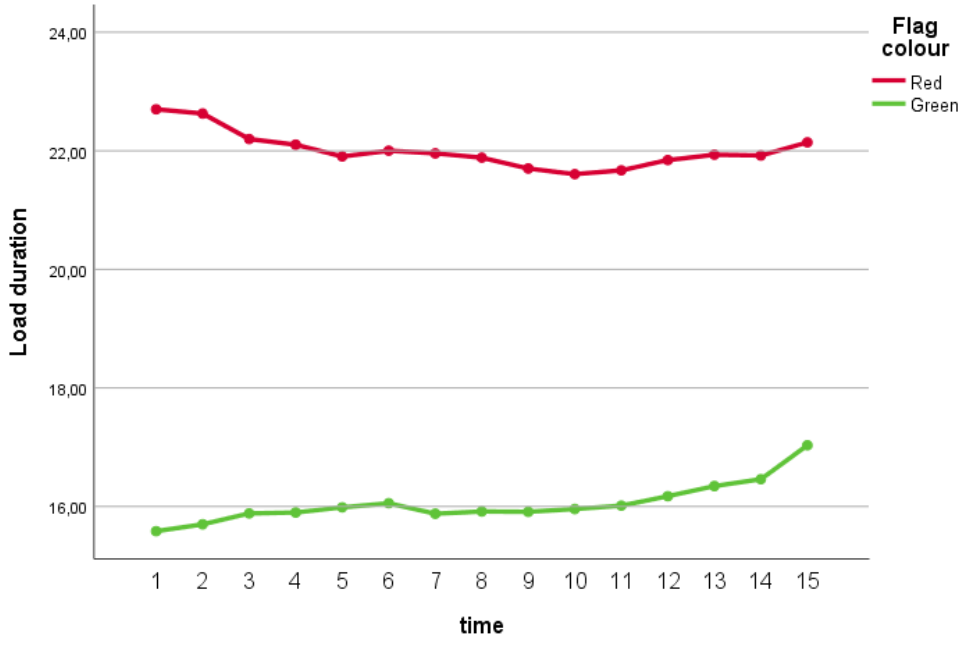

Figure 18 Development of the load duration (in minutes) over 15 weeks, for red and green flag groups.

Use case specific behaviours from the survey

In the survey we asked people about specific behaviours advised by the app and if people performed these in 2018. We expected that more people with a red flag at the first week would start performing these behaviours in 2018. Contrary to what was expected it showed that more people with a green flag stated to have bought a new refrigerator (11%) than in the group who received a red flag (9%); see Table 12.

Table 12 Buying a new refrigerator in 2018 per flag group.

	No	Yes	Total
Red	330 24% 91%	34 3% 9%	364 27% 100%
Green	881 65% 89%	108 8% 11%	989 73% 100%
Total	1.211 90%	142 11%	1.353 100%

On the other hand, as expected, it showed that more people who received a red flag (36%) set their temperatures between 3 and 4 degrees than people who received a green flag (33%); see Table 13.

Table 13 Set temperature of the refrigerator between 3 and 4 degrees C., per flag group

	No	Yes	Total
Red	230 17% 63%	134 10% 37%	364 27% 100%
Green	659 49% 66%	330 24% 33%	989 73% 100%
Total	889 66%	464 34%	1.353 100%

3.2.3 Use case 8: thermostat

The thermostat use case was introduced halfway December 2017 (week 19). During our measurement period in total 29.432 participants started using this use case, with a high peak at the end of January/ beginning of February ’18 (see Figure 19).

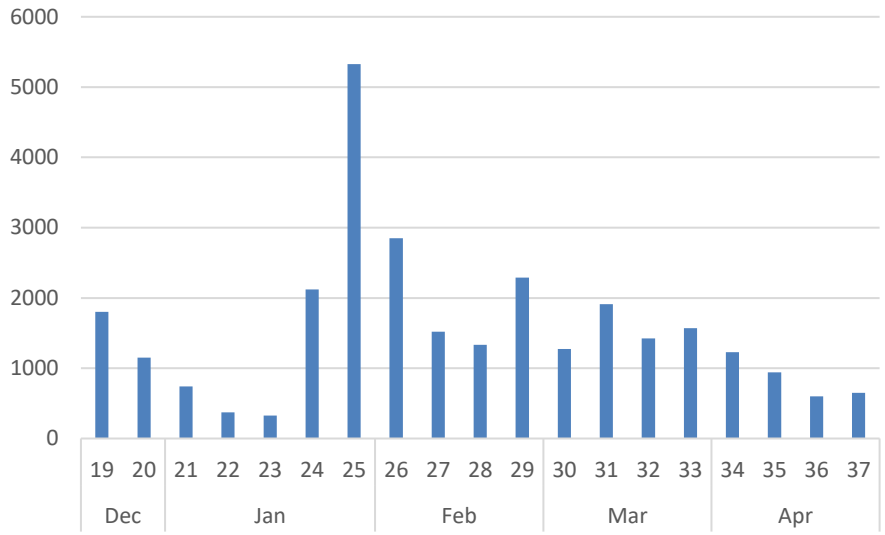

Figure 19 Number of respondents per start week

The red or green advice was based on the indicator 'lowest night setpoint'. This is the most common thermostat setpoint in the last 14 days during the night (in Degrees Celsius).

When looking at the group of respondents that had access to this use case for a minimum of thirteen weeks (n=11.841), at the start week 91% received a red flag and 9% received a green flag (see Figure 20). In week XIII, 74% received a red flag and 26% a green one. So over time the shares of red and green flags shifts to more green flags.

Figure 20 Share of red/green flags per week (in %).

When zooming in at changes per respondent we find a comparable result (see Table 14). The majority of the people with a red flag at week I, also receives a red flag at week XIII: 73% of the total switches. In addition we see that 18% switches from red to green.

Table 14 Number of switchers from Red to Green and vice versa from week I until week XIII

		Week XIII		
		Red	**Green**	**Total**
Week I	**Red**	7.644 73%	1.903 18%	9.547
	Green	148 1%	824 8%	972
	Total	7.792	2.727	10.519

Finally, we statistically tested the effect of the first flag people received and saw. First, we wanted to test the effect of any advice (being it Red or Green) on the 'lowest night setpoint' (H3). We therefore performed a 1x1 repeated measures ANOVA with 'time (I, II, III, …, XIII)' as a within subjects factor and 'lowest night setpoint' as the dependent variable. Lowest night setpoint did significantly change over time (see Figure 21), $F(12, 108300)= 38{,}909, p=.000$. The effect size is rather small (partial $\eta^2=.004$).

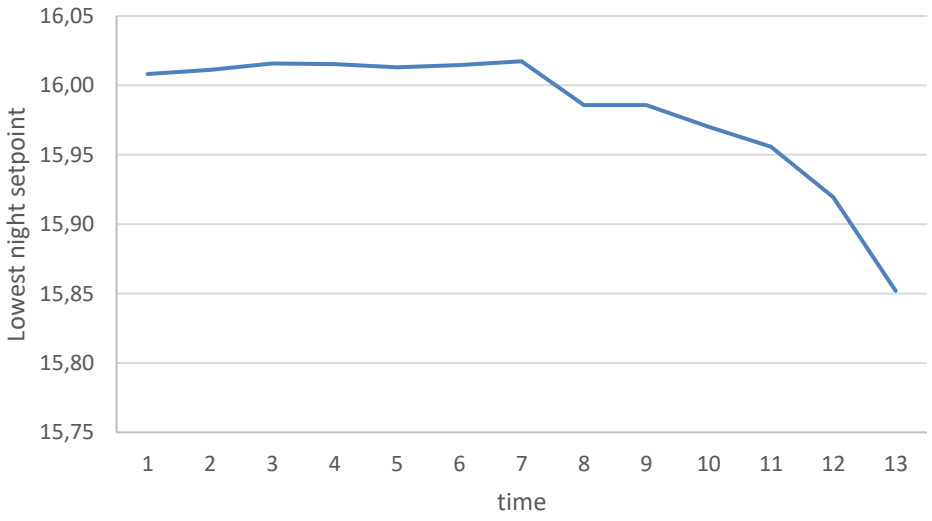

Figure 21 Development of lowest night setpoint over 13 weeks.

To check whether the colour of the flag had an impact on the lowest night setpoint' over time (H4), we split the users into two groups: users that started with a red flag versus green flag. A 1×2 repeated measures ANOVA with 'Flag colour at start week' (Red or Green) as between-subject factor, 'time (I, II, III, …, XIII)' as within subjects factor and 'lowest night setpoint' as dependent variable revealed an effect of Flag colour, $F(1, 7998)= 637{,}040$, $p=.000$. Users that received a red flag showed a decrease in 'lowest night setpoint' over time (see Figure 22). Users that received a green flag showed a small increase in 'lowest night setpoint'. The effect size is however small (partial $\eta^2=.074$).

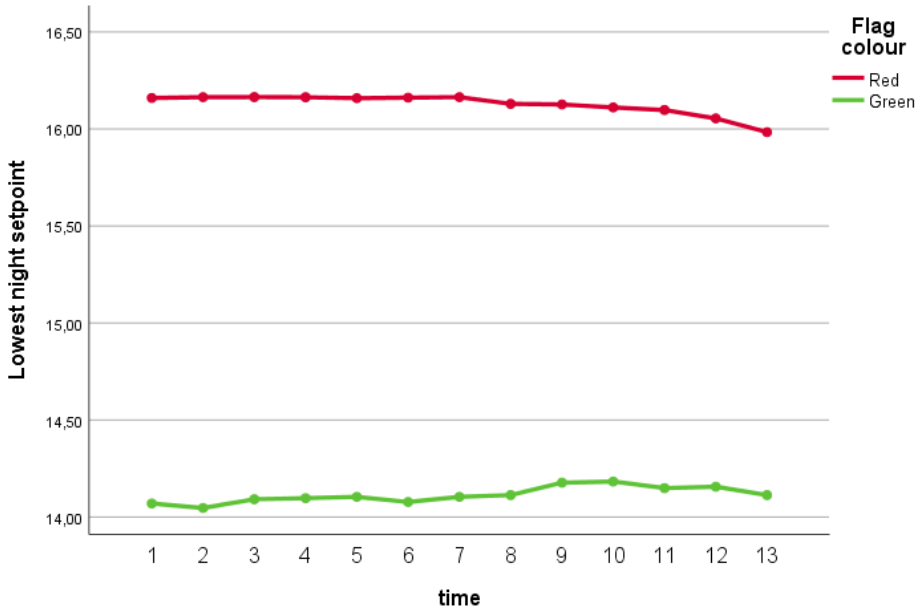

Figure 22 Development of lowest night setpoint over thirteen weeks for red and green flag groups.

Use case specific behaviours from the survey

In the survey we asked people about specific behaviours advised by the app and if people performed these in 2018. We expected that more people with a red flag in the first week would start performing these behaviours in 2018. Results showed that indeed more people with a red flag stated to have set a weekly program on their thermostat (see Table 15). In addition more people with a green flag indicated they turned off radiators in specific rooms (see Table 16). About the same number of people indicated to have put radiator foil between the wall and the radiator (see Table 17).

Table 15 Set weekly program for thermostat, per flag group

	No	Yes	Total
Red	443	720	1.163
	34%	55%	88%
	38%	62%	100%
Green	86	73	159
	7%	6%	12%
	54%	46%	100%
Total	529	793	1.322
	40%	60%	100%

Table 16 Turn off radiators in specific rooms, per flag group

	No	Yes	Total
Red	540	623	1.163
	41%	48%	88%
	46%	54%	100%
Green	63	96	159
	5%	7%	12%
	40%	60%	100%
Total	603	719	1.322
	46%	54%	100%

Table 17 Put radiator foil between the wall and radiator, per flag group

	No	Yes	Total
Red	1.038	125	1.163
	79%	10%	88,0%
	89%	11%	100%
Green	141	18	159
	11%	1%	12%
	89%	11%	100%
Total	1.179	143	1.322
	89%	11%	100%

3.2.4 Use case 12: dishwasher

The dishwasher use case was introduced at the beginning of December 2017 (week 16). During our measurement period in total 35.290 participants started using this use case, with a peak in January ’18 (see Figure 23).

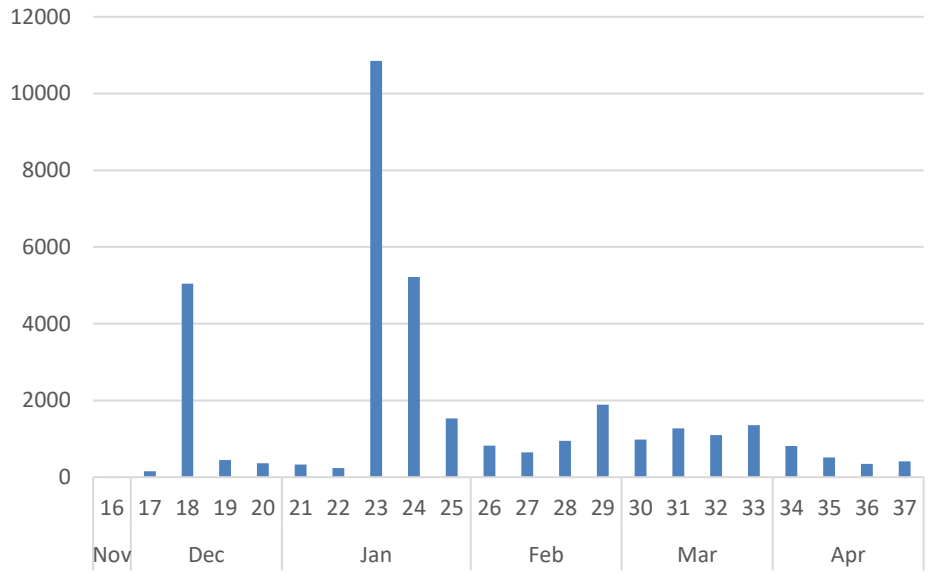

Figure 23 Number of respondents per start week

The red or green advice was based on the 'sum of energy use in kWh, over 3 month period'.

When looking at the group of respondents that had access to this use case for a minimum of twelve weeks (n=21.893), at the start week 42% received a red flag and 58% received a green flag (see Figure 24). In week XII, 47% received a red flag and 54% a green one. So over time the shares of red and green flags changed somewhat.

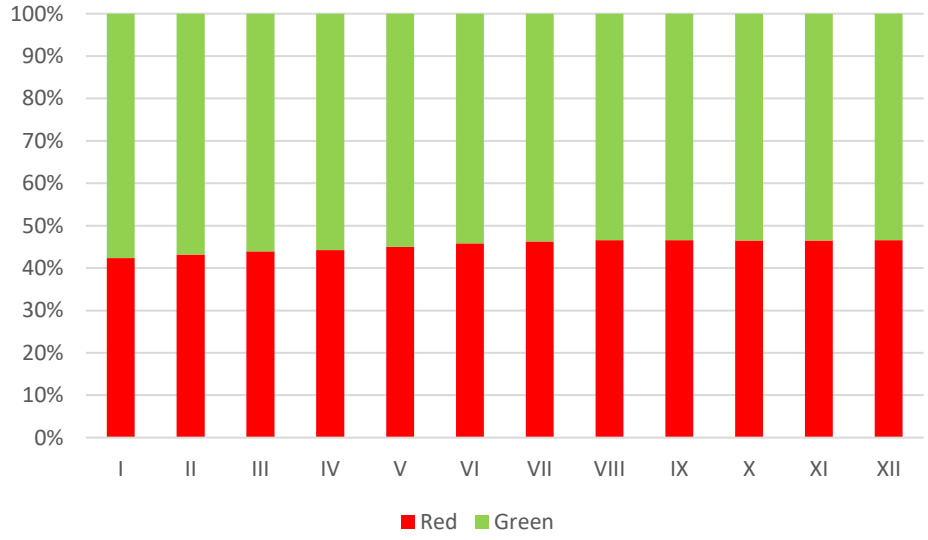

Figure 24 Share of red/green flags per week (in %).

When zooming in at changes per respondent we find a comparable result (see Table 18). The majority of the people with a red flag at week I, also receives a red flag at week XII: 35% of the total switches. This also holds for the green flags: 46%

of all respondents receive a green flag at both times. In addition we see that 7% group switches from red to green, and 12% switches from green to red.

Table 18 Number of switchers from Red to Green and vice versa from week I until week XII

		Week XII		
		Red	**Green**	**Total**
Week I	**Red**	7.602 35%	1.605 7%	9.207
	Green	2.553 12%	9.994 46%	12.547
	Total	10.155	11.599	21.754

Finally, we statistically tested the effect of the first flag people received and saw. First, we wanted to test the effect of any advice (being it Red or Green) on the 'Sum of energy use in kWh, over 3 month period' (H3). We therefore performed a 1x1 repeated measures ANOVA with 'time (I, II, III, …, XII)' as a within subjects factor and 'Sum of energy use in kWh, over 3 month period' as the dependent variable. 'Sum of energy use in kWh, over 3 month period' increased significantly over time (see Figure 25), $F(11, 237149)= 233{,}324, p=.000$. The effect size is however rather small (partial $\eta^2=.011$).

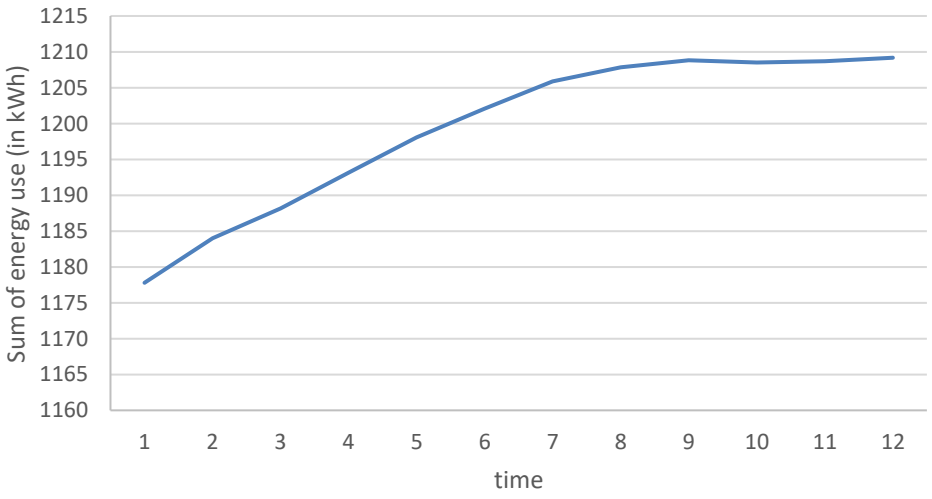

Figure 25 Development of the sum of energy use in kWh, over 12 weeks.

To check whether the colour of the flag had an impact on the 'sum of energy use in kWh, over 3 month period' (H4), we split the users into two groups: users that started with a red flag versus green flag. A 1×2 repeated measures ANOVA with 'Flag colour at start week' (Red or Green) as between-subject factor, 'time (I, II, III, …, XII)' as within subjects factor and 'Sum of energy use in kWh, over 3 month period' as dependent variable revealed an effect of Flag colour, $F(1, 21558)=25369{,}731$, $p=.000$. Users that received a red flag showed a small decrease in their 'sum of energy use in kWh, over 3 month period' over time (see Figure 26). Users that received a green flag showed a small increase in their 'sum of energy use in kWh, over 3 month period'. The size of this is quite large (partial $\eta^2=.54$).

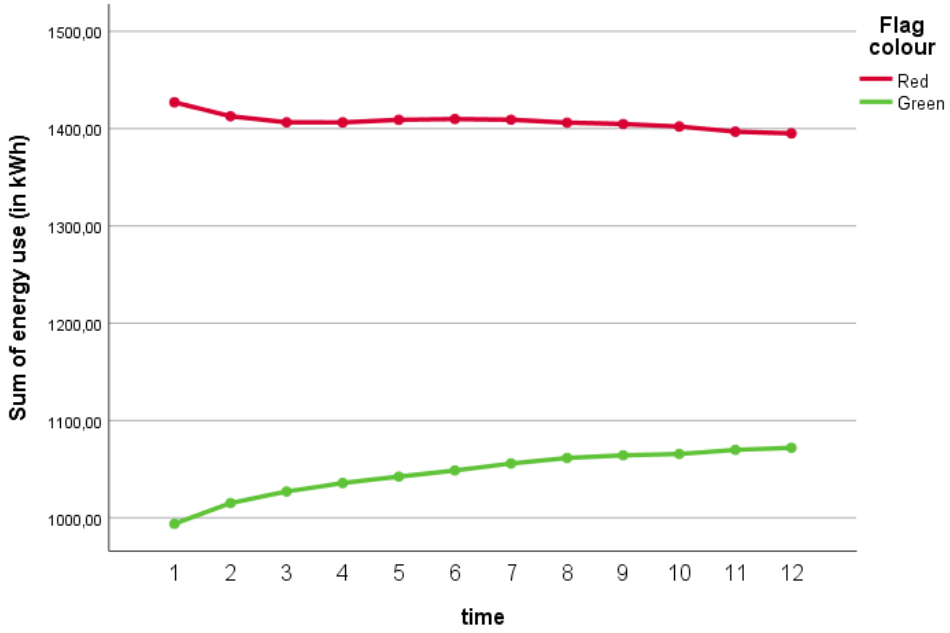

Figure 26 Development of sum of energy use in kWh, over 12 weeks, for red and green flag groups.

Use case specific behaviours from the survey

In the survey we asked people about specific behaviours advised by the app and if people performed these in 2018. We expected that more people with a red flag at the first week would start performing these behaviours in 2018 (i.e. after they downloaded the app). It showed that more people with a green flag stated to have bought a new refrigerator (12%) than in the group who received a red flag (10%).

Table 19 Buying a new dishwasher in 2018, per flag group

	No	Yes	Total
Red	548	63	611
	35%	4%	39%
	90%	10%	100%
Green	834	109	943
	54%	7%	61%
	88%	12%	100%
Total	1.382	172	1.554
	89%	11%	100%

In addition results showed that more people who received a green flag stated to have used the eco mode (52%) than people who received a red flag (44%), see Table 20. On the other hand more people who received a red flag indicated to only run the dishwasher when its full (75%) compared to the people who received a green flag (71%), see Table 21.

Table 20 Use dishwasher Eco mode (50 degrees), per flag group

	No	Yes	Total
Red	343	268	611
	22%	17%	39%
	56%	44%	100%
Green	445	498	943
	29%	32%	61%
	47%	53%	100%
Total	788	766	1.554
	51%	49%	100%

Table 21 Run the dishwasher only when its full

	No	Yes	Total
Red	153	458	611
	10%	30%	39%
	25%	75%	100%
Green	273	670	943
	18%	43%	61%
	29%	71%	100%
Total	426	1.128	1.554
	27%	73%	100%

3.2.5 Use case 14: standby use

The standby use case was introduced at the beginning of December 2017 (week 17). During our measurement period in total 25.855 participants started using this use case, with a peak in January ’18 (see Figure 27).

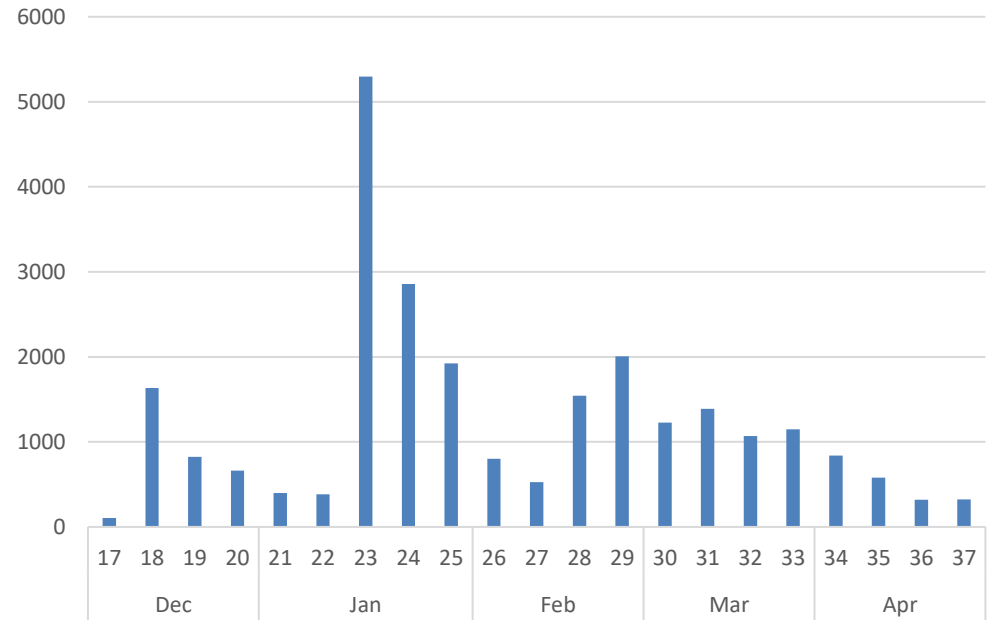

Figure 27 Number of respondents per start week

The red or green advice was based on the 'lowest electricity consumption in 5 minutes of the last day [in W]'. Inefficiency is determined by calculating an inefficient standby threshold on a per group basis. Groups are formed from households with similar characteristics such as house type, living area and number of residents. The threshold is set such that the bottom 20% of households within each group are considered efficient (green) and all others are inefficient (red). Over time the shares of red and green therefore remain 20/80.

Respondents can however switch. When zooming in at changes per respondent we find that the majority of the people with a red flag at week I, also receives a red flag at week X: 77% of the total switches (see Table 22). This also holds for the green flags: 16% of all respondents receive a green flag at both times. In addition we see that only a small group switches from red to green (4%). And a small group changes from green to red (3%).

Table 22 Number of switchers from Red to Green and vice versa from week I until week X

		Week X		
		Red	**Green**	**Total**
Week I	**Red**	11.996 77%	612 4%	12.608
	Green	505 3%	2.577 16%	3.082
	Total	12.501	3.189	15.690

Finally, we statistically tested the effect of the first flag people received and saw. First, we wanted to test the effect of any advice (being it Red or Green) on the 'lowest electricity consumption in 5 minutes of the last day' (H3). We therefore performed a 1x1 repeated measures ANOVA with 'time (I, II, III, …, X)' as a within subjects factor and 'lowest electricity consumption in 5 minutes of the last day' as the dependent variable. 'Lowest electricity consumption in 5 minutes of the last day'

did change significantly over time (see Figure 28), F(9, 131535)=6,587,p=.000. The effect size is however very small (partial η^2=.000).

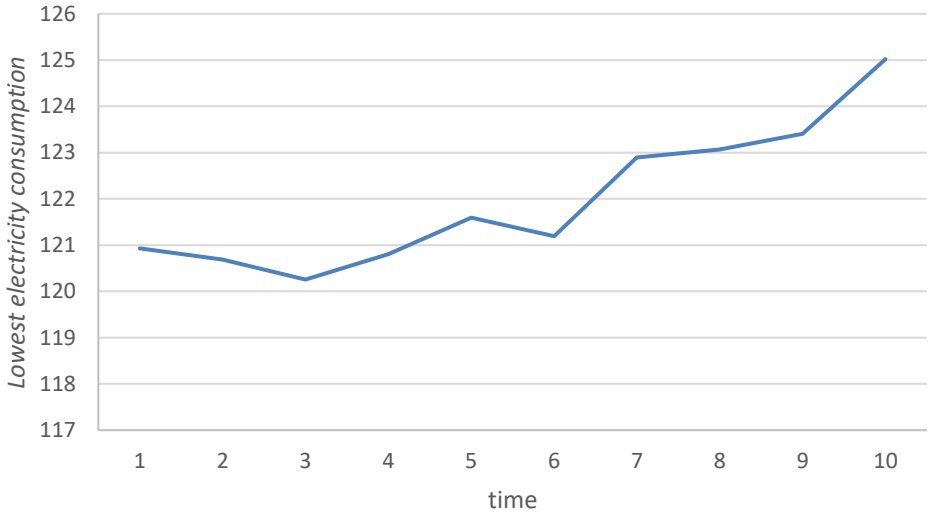

Figure 28 Development of the lowest electricity consumption in 5 minutes of the last day, over ten weeks.

To check whether the colour of the flag had an impact on the lowest electricity consumption in 5 minutes of the last day (H4), we split the users into two groups: users that started with a red flag versus green flag. A 1×2 repeated measures ANOVA with 'Flag colour at start week' (Red or Green) as between-subject factor, 'time (I, II, III, …, X)' as within subjects factor and 'lowest electricity consumption in 5 minutes of the last day' as dependent variable revealed an effect of Flag colour, F(1,13548)=1693.671, p=.000. Users that received a red flag showed a small increase in lowest electricity consumption in 5 minutes of the last day over time (see Figure 30). Users that received a green flag also showed a small increase in lowest electricity consumption in 5 minutes of the last day. The effect size can be called small to medium (partial η^2=.11).

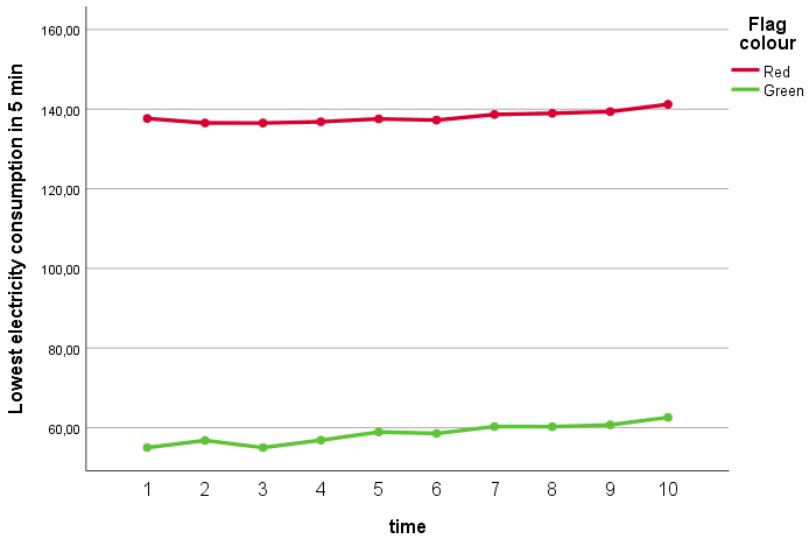

Figure 29 Development of lowest electricity consumption in 5 minutes of the last day over ten weeks, for red and green flag groups.

Use case specific behaviours from the survey

In the survey we asked people about specific behaviours advised by the app and if people performed these in 2018. We expected that more people with a red flag at the first week would start performing these behaviours in 2018. It showed that somewhat more people with a red flag (36%) used a switch timer compared to people who received a green flag (32%), see Table 23.

Table 23 Use a switch timer to turn off appliances

	No	Yes	Total
Red	643	369	1.012
	51%	29%	81%
	64%	36%	100%
Green	167	77	244
	13%	6%	19%
	68%	32%	100%
Total	810	446	1.256
	65%	36%	100%

In contrast to our expectations more people with a green flag stated to have actively sought for stand-by use (60% versus 51%) and to have used plug blocks (27% versus 24%), see Table 24 and Table 25.

Table 24 Actively look for stand-by use

	No	Yes	Total
Red	499	513	1.012
	40%	41%	81%
	49%	51%	100%
Green	97	147	244
	8%	12%	19%
	40%	60%	100%
Total	596	660	1.256
	48%	53%	100%

Table 25 Put plugs in one block and turn them off all at once

	No	Yes	Total
Red	765	247	1.012
	61%	20%	81%
	76%	24%	100%
Green	179	65	244
	14%	5%	19%
	73%	27%	100%
Total	944	312	1.256
	75%	25%	100%

3.2.6 Use case 3: showerhead

The showerhead use case was introduced at the end of February 2018 (week 28). During our measurement period in total 5.138 participants started using this use case, with a peak at the beginning (see Figure 30).

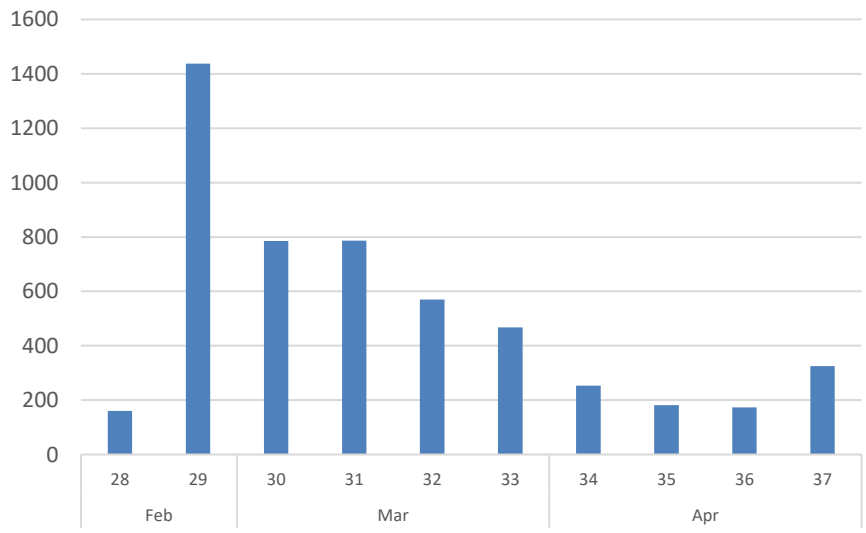

Figure 30 Number of respondents per start week

The red or green advice is based on two indicators: 'hot water gas rate' which is the gas rate while using hot water in [l/minute]) and 'gas hot water' which is the gas aggregate for hot water (based on 30 most recent non-heating days) in [m^3].

Since this use case started at the end of February there was only a small group that participated for 10 weeks (n=156). Given our choice of statistical method, number of groups and effect sizes, we calculated that at week V we had a large enough sample to do our statistical calculations. When looking at the group of respondents that had access to this use case for a minimum of five weeks (n=4.174), at the start week 49% received a red flag and 51% received a green flag (see Figure 31). In week V, 49% received a red flag and 51% a green one. These shares did not seem to differ much over time.

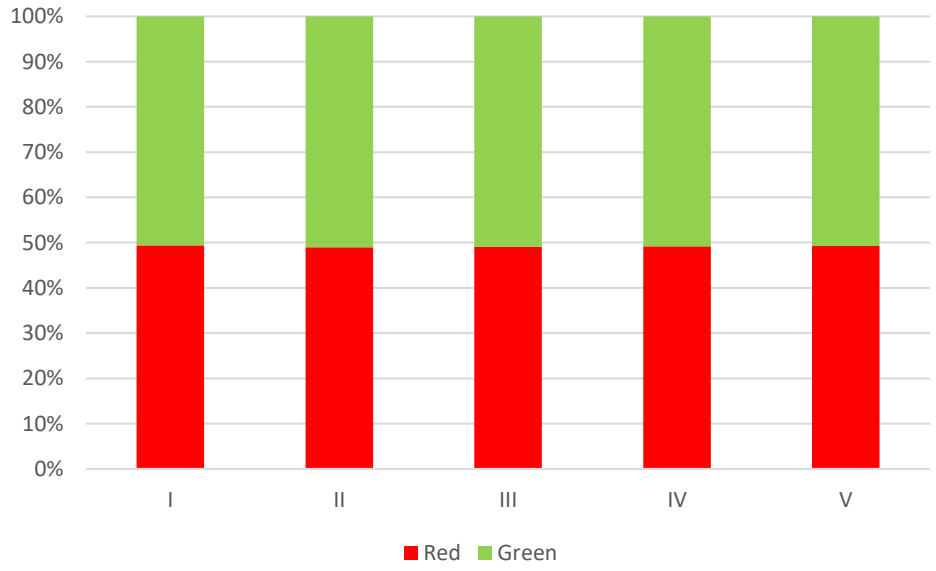

Figure 31 Share of red/green flags per week (in %) for five weeks

When zooming in at changes per respondent we find a comparable result (see Table 26). The majority of the people with a red flag at week I, also receives a red flag at week V: this 48% of the total switches. This also holds for the green flags: 49% of all respondents receive a green flag at both times. In addition we see that a small group changes from red to green (2%) and from green to red (2%).

Table 26 Number of switchers from Red to Green and vice versa from week I until week V

		Week V		
		Red	**Green**	**Total**
Week I	**Red**	1.729 48%	68 2%	1.797
	Green	55 2%)	1.786 49%	1.841
	Total	1.784	1.854	3.638

Finally, we statistically tested the effect of the first flag people received and saw. First, we wanted to test the effect of any advice (being it Red or Green) on the 'hot water gas rate' (H3). We therefore performed a 1x1 repeated measures ANOVA with 'time (I, II, …, V)' as a within subjects factor and 'hot water gas rate' as the dependent variable. Hot water gas rate did not significantly change over time although it seems to rise in the figure due to the intervals on the y-axis, this change is not significant (see Figure 32), F(4, 492)= 1,569,p=.181.

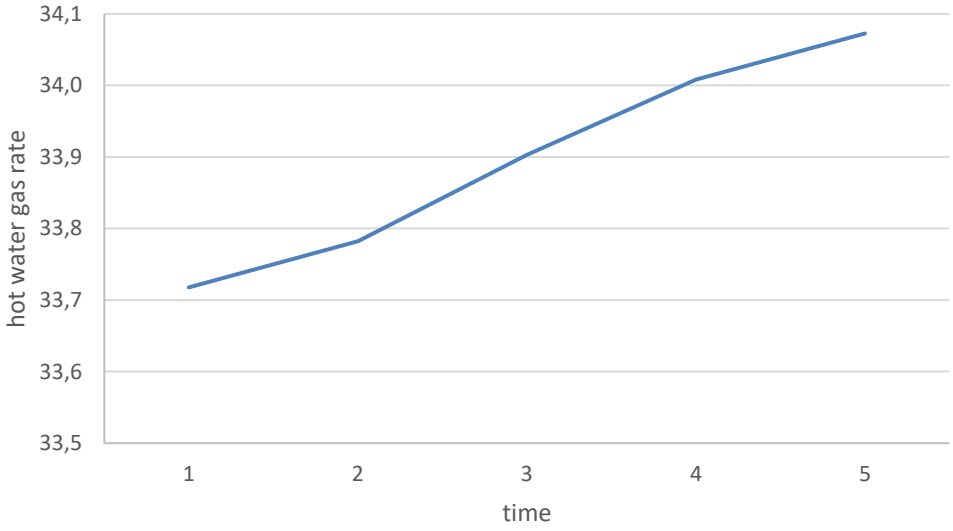

Figure 32 Development of hot water gas rate over five weeks.

To check whether the colour of the flag had an impact on the hot water gas rate (H4), we split the users into two groups: users that started with a red flag versus green flag. A 1×2 repeated measures ANOVA with 'Flag colour at start week' (Red or Green) as between-subject factor, 'time (I, II, …, V)' as within subjects factor and 'hot water gas rate' as dependent variable revealed an effect of Flag colour, F(1,76)= 62,894, p=.000. Users that received a red flag showed a small decrease in 'hot water gas rate' over time (see

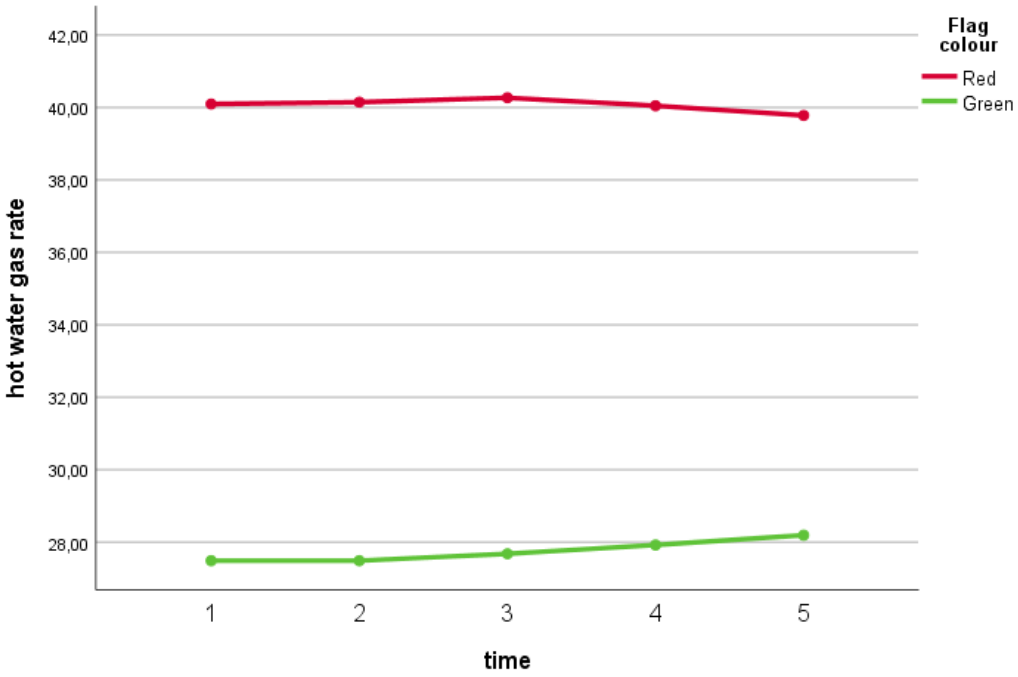

Figure 33). Users that received a green flag showed a small increase in 'hot water gas rate'. The size of this effect is quite large (η^2=.453).

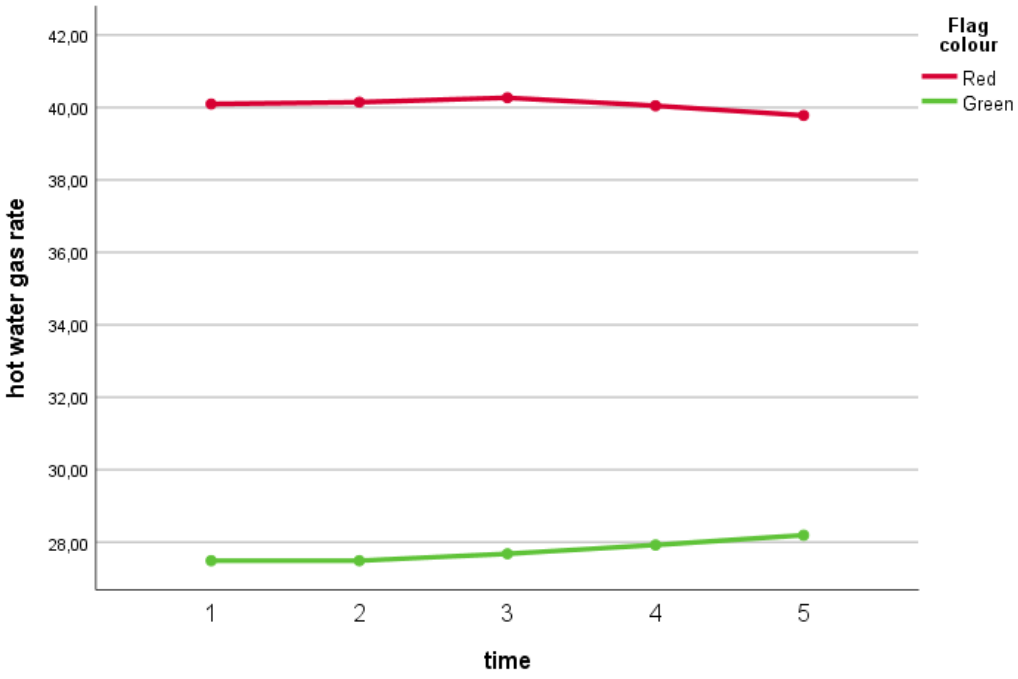

Figure 33 *Development of hot water gas rate over five weeks, for red and green flag groups.*

Use case specific behaviours from the survey

In the survey we asked people about specific behaviours advised by the app and if people performed these in 2018. We expected that more people with a red flag at the first week would start performing these behaviours in 2018. Contrary to our

expectations, it showed that more people with a green flag stated to have bought a new shower head (30%) than in the group who received a red flag (20%), see Table 27.

Table 27 Buying a water saving shower head, per flag group

	No	Yes	Total
Red	93	23	116
	36%	9%	45,1%
	80%	20%	100%
Green	99	42	141
	39%	16%	55%
	70%	30%	100%
Total	192	65	257
	75%	25%	100%

3.2.7 Use case 13: dryer

The dryer use case was introduced at the beginning of December 2017 (week 16). During our measurement period in total 24.521 participants started using this use case, with a peak in January '18 (see Figure 34).

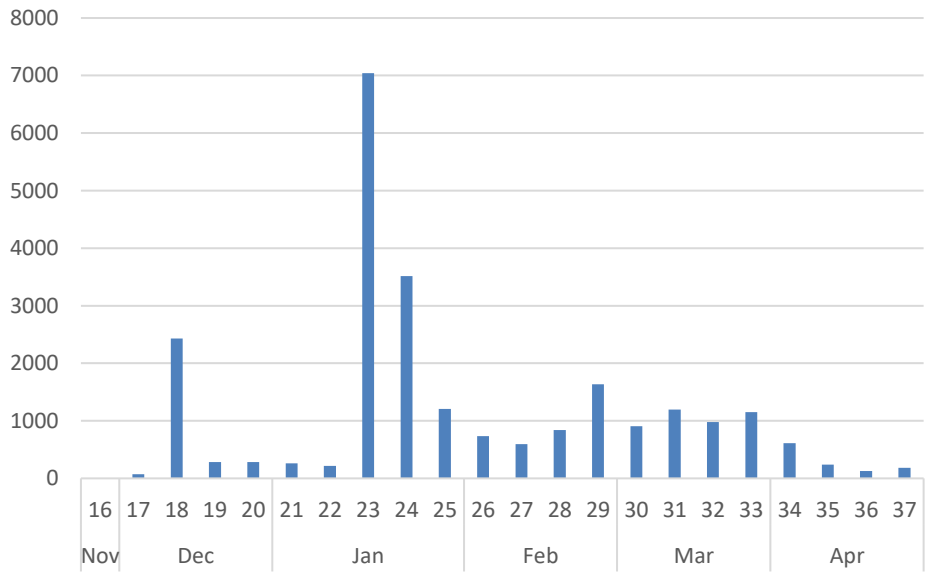

Figure 34 Number of respondents per start week

This use case differs from the other use cases, since all people with a dryer receive a red flag at week I and this remains a red flag over time. This red advice was based on the detection of any heating block for the user.

Use case specific behaviours from the survey

In the survey we asked people about specific behaviours advised by the app and if people performed these in 2018. Since all Verspillingschecker users received red coloured feedback if a dryer heating block was detected, there is no distinction to be made between a red and green flag group. It showed that 7% stated to have bought a heat pump dryer in 2018 (see Table 28). Moreover 60% indicated to have dried their clothes on a rack or line (see Table 29).

Table 28 Buying a heat pump dryer in 2018

	No	Yes	Total
Red	680	51	731
	93%	7%	100%

Table 29 Drying clothes on a rack or line

	No	Yes	Total
Red	292	439	731
	40%	60%	100%

3.3 How effective is automatically given personalized advice on energy efficiency measures in households?

For this research question we tested the impact of the Verspillingschecker on

1. attention for energy waste at home,
2. attitudes towards the specific actions addressed,
3. realized actions (the amount of measures taken: daily habitual behaviours/setting changes/small investments versus larger investment behaviours).

For hypothesis 5

We tested whether 'attention for energy waste at home' changed significantly in the Verspillingschecker group, as was hypothesized. We could do so because we measured overall attention for energy waste at home *compared to 'a year ago'*. We did use 'a year ago' because we thought this was the easiest comparison with a relatively recent period in their life without the Verspillingschecker.

We did not incorporate that 'a year ago'-reference in the statements about the specific actions addressed to prevent energy waste at home, as we think it is not realistic to assume (and it is annoying as well to ask) respondents know how they now think differently about a specific act, or how much more often they now carry out specific acts, compared to a year ago. Strictly put, we therefore tested hypotheses 5 for 'attention for energy waste at home' only.

How did we test this?
A one sample T-test was conducted to check whether the mean scores of the scales measuring attention were significantly different from 'neutral' (4 on a 1 to 7 point scale). In this case users of the Verspillingschecker (dis)agreed to statements about an increased attention for energy waste at home compared to last year, which can be taken as an indicator for the impact of the Verspillingschecker app.

For hypothesis 6
We tested whether the *changed* attention for energy waste at home was *stronger* in the Verspillingschecker group than in the Control group, as was hypothesized.
We tested hypothesis 6 differently for attitudes towards the specific actions addressed, as we did not measure change on specific action level: We tested whether the means of the attitudes in the Verspillingschecker group were different than the means of the attitudes in the Verspillingschecker group. For the amount of realized actions in 2018 we did the same: We tested whether the means of the groups on these factors were different from each other.

How did we test this?
Independent sample T-tests were conducted on each behaviour to see whether the test group had significantly different scores than the control group.

1) Attention for energy waste at home

The Verspillingschecker users somewhat agreed with statements about a stronger focus on the topic than 'last year'. Their average scores on the overall attention scale (M=4,75, SD=1,27) were significantly above neutral (4 is representing 'neutral' on a 7 point scale, ranging from 1 = totalle disagree to 7 = totally agree), t (2532)=29,62, p=.000
In addition their scores on attention for energy waste at home werehigher than in the Control group (M=4,08, SD=1,43), t (5767)=18,37, p=.000. This should be seen as a medium-size effect (Cohen's *d* =0,50).

Hypotheses 5 and 6 were confirmed for the impact factor attention for energy waste at home:

The Verspillingschecker had an impact on overall **attention for energy waste in home**. The attention growth for this topic was stronger for users of the Verspillingschecker than for the control group respondents.

2) Attitudes towards the specific behaviours addressed

Small effects were found: The Verspillingschecker group had a significantly somewhat stronger positive attitude than the control group about four of the seven behaviours addressed in the Verspillingschecker, being (in the order of most postive towards least positive attitude):
- Attitude towards lowering the termostat at night,
- Attitude towards lowering the thermostat when leaving home,
- Attitude towards using appliances in Ecomode,
- Attitude towards replacing a still working appliance.

The Control group had a somewhat stronger positive attitude than the Verspillingschecker group about two of the seven behaviours addressed in the Verspillingschecker, being:

- Attitude towards efficient hot water use in the shower
- Attitude towards switching off appliances instead of keeping them in stand-by mode.

There was no difference between the two groups for:

- Attitude towards washing at low temperatures.

The means, standard deviations and statistics for the attitude scales mentioned above can be found in the table below.

Table 30 Differences between groups for attitudes towards Verspillingschecker related behaviours

	Group	*N*	Mean (1 to 7 point scale)	Std. Deviation	p	Effect size (Cohen's *d*).
Attitude towards replacing a still working appliance	VC	2.532	**4,19**	1,24	,000	0,13 (small effect)
	Control	3.237	4,03	1,29		
Attitude towards using appliances in Ecomode	VC	2.532	**5,29**	1,06	,048	0,05 (small effect)
	Control	3.237	5,23	1,14		
Attitude towards washing at Low temperatures	VC	2.532	5,18	1,09	,478 (non significant)	na
	Control	3.237	5,16	1,16		
Attitude towards switching off appliances instead of keeping them in stand-by mode	VC	2.532	5,04	1,32	,000	0,14 (small effect)
	Control	3.237	**5,23**	1,37		
Attitude towards efficient hot water use in the shower	VC	2.532	5,19	1,16	,003	0,08 (small effect
	Control	3.237	**5,28**	1,20		
Attitude towards lowering the thermostat when leaving home	VC	2.532	**5,93**	1,03	,000	0,22 (small effect)
	Control	3.237	5,69	1,19		
Attitude towards lowering the thermostat at night	VC	2.532	**6,20**	0,93	,000	0,18 (small effect
	Control	3.237	6,01	1,17		

All average scores on the attention scales were significantly above neutral (4 is representing 'neutral' on a 7 point scale, ranging from 1 = totally disagree to 7 = totally agree). The Verspillingschecker user have positive attitudes about the specific behaviours addressed, although not all of them are as positive as the other (see the mean scores). In both groups respondents have the most positive attitude about lowering the thermostat at night, and the least positive attitude about replacing a still working appliance.

Hypotheses 6 can be confirmed for attitudes towards 4 out of 7 actions:

> The **attitudes about replacing a still working appliance; using appliances in Ecomode, lowering the thermostat when leaving home, and lowering the thermostat at night** were stronger positive for users of the Verspillingschecker than for the control group respondents. The differences between the groups should be interpreted as small differences.

3) Realized actions (the amount of measures taken due to Verspillingschecker)

We compared the number of actions reported in the Verspillingschecker group with the number of actions reported in the Control group. Respondents had to indicate which actions they performed in 2018 (this is after the Verspillingschecker was launched). The difference in total actions that were reported in 2018, is an indicator of the impact of the Verspillingschecker on saving behaviour/preventing energy waste at home.

a) daily (habitual) behaviours, setting changes, small investments

The average amount of in home small actions to prevent energy waste in 2018, mentioned by Verspillingschecker users was 5,33 (SD = 2,80) (N=2,532) compared to 4,61 (SD= 2,97) (N=3,237) by the control group. The VC users reported more actions than the control group respondents, $t(5767)=9,29$, p=.000.
This should be seen as a small effect (Cohen's d =0,25).

The two differences that stand out most are: looking for standby usage and setting a weekly thermostat program.

- "Looking for Standby power" (Sluipverbruik) gets more positive reactions in the Verspillingschecker group (45% yes) than in Control group (21%).
- "Set weekly program for thermostat" gets more positive reaction among VG (58% yes) than in CG (25%).

b) Larger investment behaviour

In the control group fewer new appliances were bought in 2018 (see Table 31) than in the Verspillingschecker group. The average amount of appliances bought in 2018 reported by Verspillingschecker users was 0,38 (SD = 0,76) (N=2.532) compared to 0,27 (SD= 0,63) (N=3.237) by the control group. The VC users reported more actions than the control group respondents, $t(5767)=5,84$, p=.000.
This should be seen as a small effect (Cohen's d =0,16).

Table 31 Newly purchased appliances per group

	Verspillingschecker group (N=2.532)	Control group (N=3.237)
Dishwasher	9,6%	5,8%
Heatpump dryer	5,9%	3,7%
Fridge	10,5%	8,3%
Washing machine	11,7%	9,2%

Hypotheses 6 can be confirmed for actions to prevent energy waste at home:

Verspillingscheck users report more **actions** to prevent energy waste in home than control group participants.

3.4 Assumption 1: The amount of data collected by the Verspillingschecker application is adequate for detecting inefficient energy usage.

To support this assumption Quby provided Annex E.

Perceived credibility of advices In addition, we present how strong users of the Verspillingschecker believed the credibility of the diagnoses of the Verspillingschecker.

There were 1.841 users that gave their response to the statement: 'The advices that I received from The Verspillingschecker were credible.' (De adviezen die ik van de Verspillingschecker kreeg zijn geloofwaardig). 55,6% Did (somewhat) agree with that statement. 19,1% Did (somewhat) disagree, and 25,4% did not disagree or agree. See the Table 32 below for more frequencies and percentages.

Table 32 Frequencies and percentages on the statement of credibility

"The recommendations I received from The Verspillingschecker were credible"					
		Frequency	Percent	Valid Percent	Cumulative Percent
Value	1 (totally disagree)	86	3,4	4,7	4,7
	2	123	4,9	6,7	11,4
	3	143	5,6	7,8	19,1
	4	467	18,4	25,4	44,5
	5	487	19,2	26,5	70,9
	6	425	16,8	23,1	94
	7 (totally agree)	110	4,3	6,0	100,0
	Total	1.841	72,7	100,0	
Missing	System	693	27,3		
Total		2.534	100,0		

The amount of data collected by the Verspillingschecker application is adequate for detecting inefficient energy usage. Most users (56%) (somewhat) agreed **with the statement that the recommendations were credible.**

3.5 Assumption 2: The generated personal home owners advice offers sufficient insight in their energy conservation potential.

To test this assumption we asked home owners versus tenants to assess whether they disagreed or agreed with the following statement:
"**I now know better now how much energy I can save in my home than a year ago**." ('Ik weet nu beter hoeveel energie ik in mijn huis kan besparen dan een jaar geleden')

An Independent-Samples T Test has been conducted to check whether the scores given by the test group differ significantly from the control group.

The **home owners** that used the Verspillingschecker scored higher than the control group average score (**M=4,72**, SD=1,55) was higher than in the Control group (**M=3,90**, SD=1,68), t (4132)=16,16, p=.000

The **tenants** that used the Verspillingschecker more strongly agreed (**M=4,96**, SD=1,51) with the statement than tenants in the Control group (**M=4,31**, SD=1,75), t (1633)=7,74, p=.000.

Support for assumption 2 was found:

In the Verspillingschecker group, users more strongly believe that they **increased their knowledge about how much energy they can save**, than in the control group without a Toon and Verspillingschecker.

3.6 Assumption 3: The generated personal advice provides home owners insight in what energy conservation measures are the most interesting for them.

To test this assumption we asked home owners versus tenants to assess whether they disagreed or agreed with the following statement:
I do not know what I can do in my home to save energy ('Ik weet niet wat ik nog meer in mijn huis kan doen om energie te besparen').

An Independent-Samples T Test has been conducted to check whether the scores given by the test group differ significantly from the control group.

On average **home owners** somewhat disagreed with statement (meaning they agree that they know what they can do to save energy), but there was a difference between the two groups: Home owners that used the Verspillingschecker, more strongly disagreed than control group respondents,meaning Verspillingschecker users more strongly believe that they know what to do.. The Verspillingschecker

group had a lower average score (**M=3,60**, SD=1,66) than the Control group (**M=3,76**, SD=1,77), t (4132)=-2,93, p=.003

The **tenants** had means at the other end of the scale: above 4 meaning they agree that they NOT know what they can do to save energy. The 669 tenants that used the Verspillingschecker (less strongly) agreed with that (**M=4,29**, SD=1,71) than 966 tenants in the Control group (**M=4,57**, SD=1,76), t (1633)=-3,18, p=.002. This should be seen as a small effect (Cohen's *d* =0,16).

Assumption 3 can be accepted:

Verspillingschecker users know better what they can do to save energy than control group respondents without a Toon and Verspillingschecker app. Home owners know this better than tenants. Tenants more strongly disagree than home owners that they know what to do.

3.7 Assumption 4: The generated personal advice helps home owners in taking investment decisions concerning energy conservation measures.

To test this assumption we asked respondents (including home owners) the following two questions:

- Eneco helps me to tackle energy waste (‘Eneco helpt mij om energieverspilling tegen te gaan.’)
- Eneco provides me insight in the yield of energy saving actions or measures (‘Eneco geeft mij inzicht in wat energiebesparende acties of maatregelen mij opleveren’).

Factor analyses showed a higher reliability with the addition of the third statement:

- Recently, Eneco made me curious about what I can do to prevent energy waste at home (‘Eneco heeft mij de afgelopen tijd nieuwsgierig gemaakt naar wat ik kan doen om verspilling in mijn huis tegen te gaan.’)

The three statements were therefore taken together as one factor measuring: **Experienced help from Eneco.**

An Independent-Samples T Test was conducted to check whether the scores given by the test group differ significantly from the control group.

The Verspillingschecker had an impact on Experienced help from Eneco.
Their average scores (M=5,00, SD=1,29) were higher than in the Control group (M=4,09, SD=1,41), t (5767)=25,17, p=.000.

Verspillingscheck users **perceive more help with preventing energy waste** at home than control group participants.

It is doubtful how much this confirms the assumption 4. Strictly put, 'Help from Eneco with preventing energy waste at home' is not the same as 'Help with taking investment decisions'.

3.8 Satisfaction

Satisfaction Verspillingschecker
In addition, we present results about how users of the Verspillingschecker perceive the app.

From the 2.532 respondents in the Verspillingschecker group that filled out the questionnaire, 83% did download and keep the Verspillingschecker app.

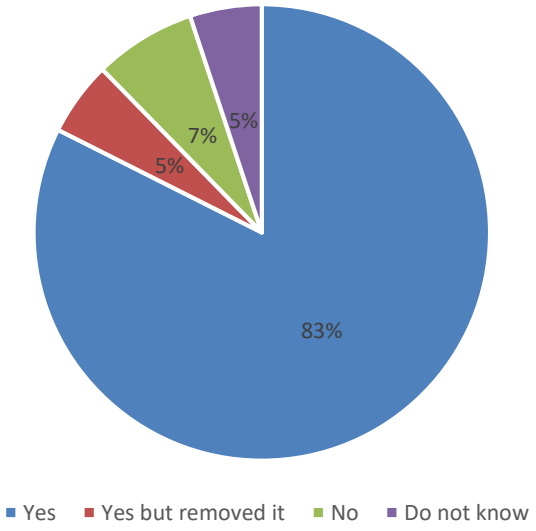

Figure 35 Number of people who downloaded the app, in % (N=2.532)

These 83% were given the question: **How satisfied are you with the Verspillingschecker?** (In Dutch: Hoe tevreden bent u over de Verspillingschecker?)
The 1694 users that filled out this question show the app is valued with a mean score of **6,4** (SD=1,6) (on a 10 point scale ranging from 1= very unsatisfied to 10=very satisfied).

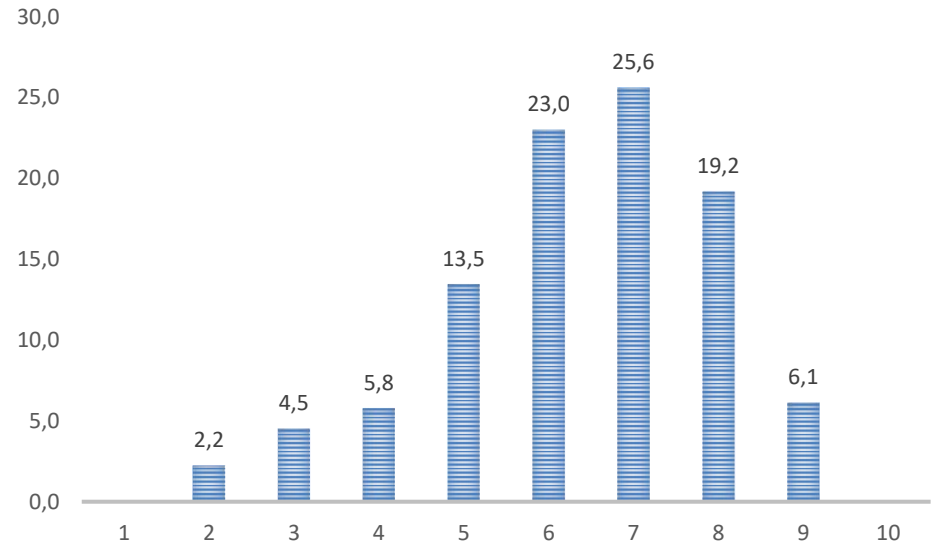

Figure 36 Satisfaction with the Verspillingschecker, in % (N=1.694, Mean=6,4)

Respondents did have room to comment on their score. These comments were provided separately to Quby for further analyses. A few examples of comments are:

- 'Results are unclear'
- 'The last time I looked there was only one appliance'
- '… I doubt if it is right'
- 'The app is very clear and easy to use'
- 'It provides a good overview'
- 'It gives practical advice'

Satisfaction Toon

All respondents in the Verspillingschecker group (N=2.532) were asked: How satisfied are you with Toon? (Hoe tevreden bent u over Toon®?). Users gave a 7,8 on average. See Figure 37 below.

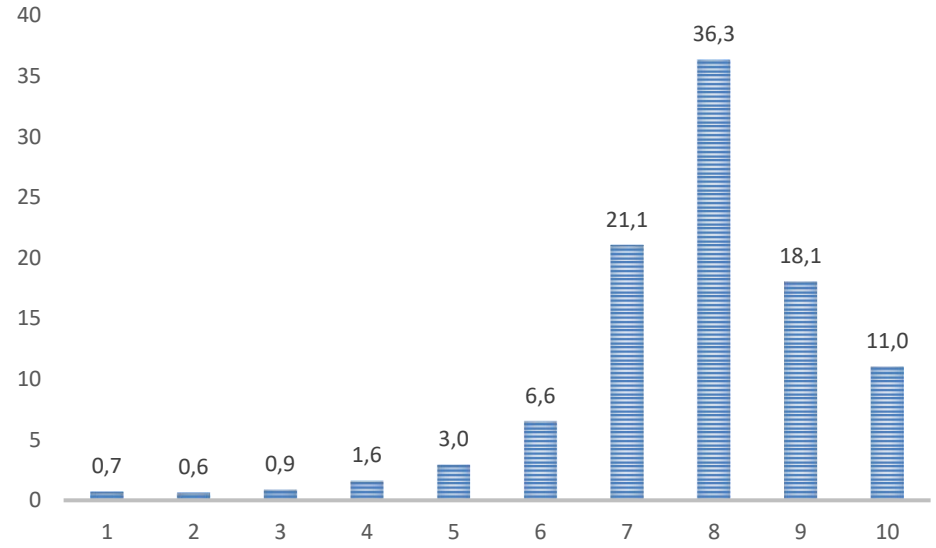

Figure 37 Satisfaction with Toon, in % (N=2.532, Mean=7,8)

Perceived feasibility of advices

In figure 39 can be seen that the respondents in the Verspillingschecker group had different opinions about how feasible the advices were to them. Most of them (29,3%) did neither disagree or agree with the statement: 'The advices that I received from The Verspillingschecker are feasible.' (De adviezen die ik van de Verspillingschecker kreeg zijn voor mij uitvoerbaar.)

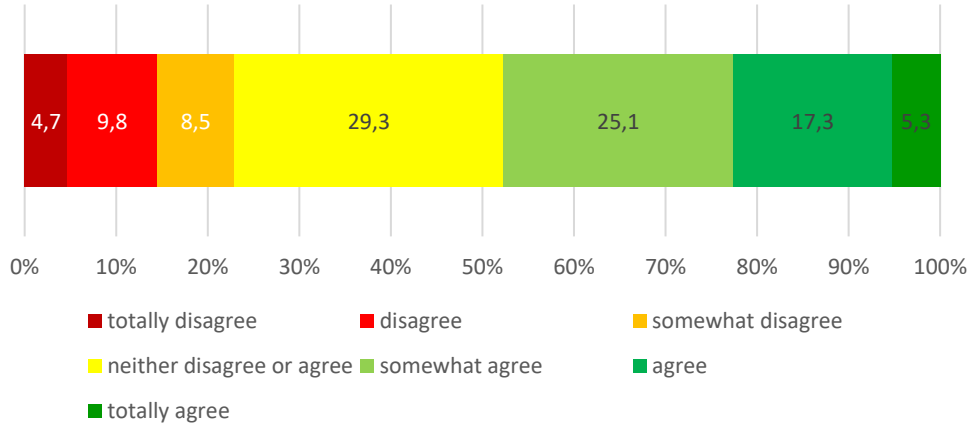

Figure 38 The advice from the app is feasible, in % (n=1.841)

Perceived interest/attraction of advices

Figure 39 shows that the respondents in the Verspillingschecker group had different opinions about how interesting the advices were for them. A higher percentage of the users agreed (47,8%), some disagreed (6,6%) or were indifferent (15,3%) to the statement: 'The advices that I received from The Verspillingschecker are of interest to me.' (De adviezen die ik van de Verspillingschecker kreeg vind ik interessant.)

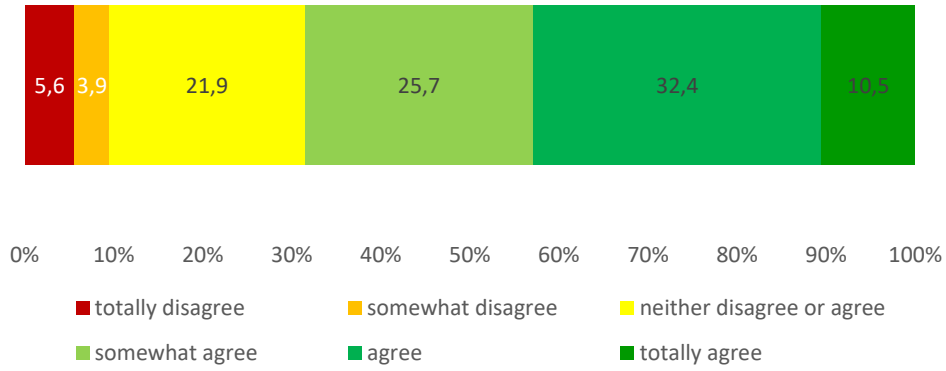

Figure 39 The advices in the app are of interest to me, in % (n=1.768)

Match with own assumption about preventing energy waste in home

Users indicated to what extent the Verspillingschecker did match with their own ideas/assumptions of how to tackle energy waste at home. (In Dutch: In hoeverre sluit de Verspillingschecker aan bij uw eerdere vermoedens/eigen ideeën over hoe u energieverspilling in huis tegen kunt gaan?). For most users the advices matched

with what they had already thought of (47,2%). Almost the same share of users did not have assumptions about how to tackle this (41,9%); see Figure 40.

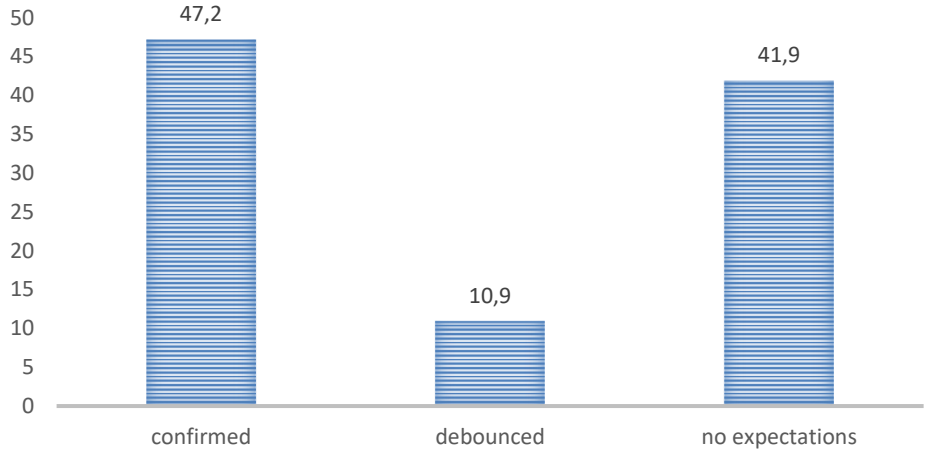

Figure 40 Match with own ideas and assumptions, in % (n=1.841)

Ease of use

Most respondents (69%) agreed with the statement: 'The Verspillingschecker is easy to use.' (De Verspillingschecker is gemakkelijk in het gebruik.) See Figure 41.

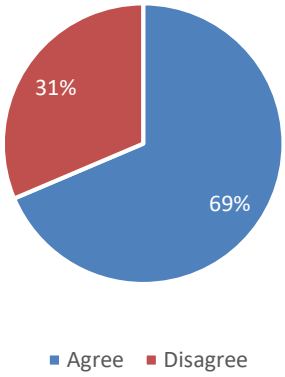

Figure 41 The Verspillingschecker is easy to use, in % (n=1.841)

Suggestions for improvement

Suggestions for improvement were of all kinds. A few examples are given below:

- 'Send push messages with advice'
- 'Combine saving advice with offers from third parties (Mediamarkt, etc)'
- 'Add lighting. I would have liked to see the change when I replaced my lightbulbs with LED.'

4 Conclusions, discussion and recommendations

4.1 Conclusions

In our research we focused on answering two research questions:

1. Can automatically generated personalized advice realize energy savings per household of 5-10%?
2. How effective is automatically generated personalized advice on energy efficiency measures (actions) in households?

Ad 1. Energy savings of 5-10%?
Our results show that the Verspillingschecker does not have the hypothesized impact in the order of magnitude of 5-10% on the overall electricity and gas use over time. When looking at average saving percentages, a very small impact on overall electricity usage can be seen: In the first four impact weeks a 0,01% stronger decline in electricity use was seen for the Verspillingschecker users compared to the Control group. When averaging the first fourteen impact weeks, a 0,03% stronger decline in electricity use was found for the Verspillingschecker users compared to the Control group.

Since these differences were quite small we went to look for explanations in the various use cases.

A closer look at the use-cases
After usage of the Verspillingschecker during the test period, at the use-case level, we see a slight increase in the energy use of the refrigerator, washing machine, dishwasher and stand-by consumption, versus a small decrease of the energy use related to lowering the thermostat at night.
An interesting pattern evolves for most of the use-cases: For the refrigerator, washing machine, dishwasher and thermostat settings at night, it holds that a red flag leads to a small decrease of the energy use -as expected-, but a green flag leads to a small increase of the energy use. Thus, the small positive effect of showing people a red flag, seems to be diminished by the increase of energy use of the people who receive a green flag.
For stand-by usage this was not the case: Here both a red and green flag led to a small increase of energy use.

Ad 2. Effect on attention for energy waste at home, attitudes towards proposed actions, and the amount of energy efficiency actions taken
We took the AIDA (attention, interest, desire, action) factors into account, as well as attitudes towards saving energy at home, and attitudes towards the specific actions that are advised by the Verspillingschecker. For the AIDA spectrum our factor analyses showed that respondents had perceived three out of the four AIDA factors (attention, interest, desire) as one single factor. We therefore took the items on these scales together into one factor that we labelled ‘attention for energy waste at home’.

In both the Verspillingschecker group and the control group we asked how much their attention for this topic had been changed since last year (the year before the Verspillingschecker had been introduced). Our results show that attention increase

for this topic was stronger in the Verspillingschecker group than in the control group. This leads to our conclusion that the Verspillingschecker had a positive impact on ‘attention to prevent energy waste at home.’

At the use-case specific level we found some small differences between groups that point in the same direction: The attitudes about replacing a still working appliance; using appliances in Eco mode, lowering the thermostat when leaving home, and lowering the thermostat at night were stronger positive for users of the Verspillingschecker than for the control group respondents. However, we also found two actions that point in the opposite direction: The attitude towards efficient hot water use in the shower, and towards switching off appliances instead of keeping them in stand-by mode, was stronger positive in the control group than in the Verspillingschecker group. For washing at low temperatures, no differences between groups existed.

We cannot claim the differences between groups are due to the Verspillingschecker, as we did not measure a *change* in these attitudes. It could have been that they already differed on these factors before the Verspillingschecker was introduced (see discussion).

At the action-level Verspillingschecker users report more actions to prevent energy waste in home than control group participants. Verspillingschecker users describe to do somewhat more small actions in their home to prevent energy waste. And they indicate having purchased more new energy efficient appliances than the control group.

Besides the two research questions, four assumptions were tested.

Assumption 1: Adequacy of the Verspillingschecker data
Up to 5 validation checks per use case were done to check whether the amount of data collected by the Verspillingschecker application is adequate for detecting inefficient energy usage.All tests for the launched 7 use cases confirmed that the reliability of the measurements used by Quby to detect inefficiency was sufficient. Most users (56%) (somewhat) agreed with the statement that the recommendations were credible**.** This supports the assumption that the amount of data collected by the Verspillingschecker application is adequate for detecting inefficient energy usage.

Assumption 2: Insight in saving potential
In the Verspillingschecker group, users more strongly believe that they increased their knowledge about how much energy they can save, than in the control group without a Toon and Verspillingschecker. Means are above neutral, pointing in the positive direction. This supports the assumption that the generated personal advice offers home owners ‘sufficient’ insight in their energy conservation potential.

Assumption 3: Insight in most interesting actions
The generated personal advice provides home owners insight in what energy conservation measures are the most interesting for them.
Verspillingschecker users know better what they can do to save energy than control group respondents without a Toon and Verspillingschecker app. Home owners

know this better than tenants. Tenants more strongly disagree than home owners that they know what to do.

Assumption 4: Help with taking investment decisions
The generated personal advice helps home owners in taking investment decisions concerning energy conservation measures.
Verspillingschecker users perceive more help with preventing energy waste at home than control group participants.

In short, the app does inform people, and change their behaviour, but not systematically, and with little effect on energy use.

4.2 Discussion and recommendations

In this research we analyzed the effect of the Verspillingschecker on energy use of its users compared to a control group. There are a number of points of discussion we would like to stress here.

Small effects
The effects of the Verspillingschecker that we found in this study were smaller than expected. There can be various explanations for that.

One of the explanations is that percentages are often overestimated.
In 2010, KEMA estimated that in the Netherlands the smart meter in combination with bimonthly feedback (by email or mail) could on average lead to a 3.2% decrease of household electricity use and 3.7% decrease of household gas use[1]. In addition, combining this feedback per mail with other methods (like apps or displays) could lead to savings of 6.4% on electricity and 5.1% on gas, they say. A 2014 RVO report stated that these estimated potentials seem relatively high in the Netherlands, among other things since the effect of the bimonthly feedback has not proven to be of this size. Moreover these estimations seem high compared to experiences with direct feedback in the United Kingdom and Ireland (average of 2% to 4% for electricity and 3% for gas)[2]. Various pilots in the Netherlands however indicated that these potential outcomes could be realistic, provided that the application and data visualization connect to the practical preference and interests of the consumer (RVO, 2014). It is possible that this was not the case (yet) in the first version of the Verspillingschecker.

Another explanation is that many of the use cases provided delayed feedback; feedback based on averages of a time period of for example three months or 30 days. Behaviour changes of its users (for example investing in a new refrigerator or using the eco program on the dishwasher) are therefore only visible in the app after a longer period of time. Except for the standby use case that is based on the previous day, the feedback is therefore indirect. Our research period of December 2017 up until April 2018 was rather short for some of the use cases; especially the ones that involved many participants in February 2018. This can mean that the

[1] KEMA (2010) Intelligente meters in Nederland. Een herziene financiële analyse en adviezen voor beleid.
[2] RVO (2014). Dutch Energy Savings Monitor for the Smart Meter. Downloaded from: http://www.smart-energy.com/wp-content/uploads/2014/06/Dutch-Smart-Meter-Energy-savings-Monitor-final-version.pdf

feedback provided later on, and its effect, were not taken into account in this study, thereby underestimating the effect of the application.

Control group
It was not possible to include a group of Toon users without a Verspillingschecker. Therefore the *net* or pure effect of Verspillingschecker could not be measured. This results in the fact that the effects found in this study cannot be isolated from the pure effect of the Toon itself. With this experimental group we therefore looked at the effect of the Verspillingschecker + Toon, and not solely at the effect of the Verspillingschecker app.

As the differences in energy saving percentages between the groups were almost zero, one could ask whether there is an effect of Toon at all. We argue this cannot be known from this experiment as we do not know energy consumption data from the period before the test groups had installed a Toon. It could be that the impact of Toon had already taken place before this experiment.

A Toon only group could also help to clarify the baseline difference in electricity consumption that existed: The Verspillingschecker and the Control group differed systematically on their electricity use; on average around 10 kWh per week. Eneco confirmed that on average Toon users use more energy. Possibly this is the reason these clients purchase a Toon. They possibly have more energy saving potential. Another explanation could be that Toon users, and especially the Verspillingschecker users, in general use more appliances or gadgets than other groups.

Recommendation We advise to include a Toon only group in the next experiment. This could isolate the effect of the (next version of) the Verspillingschecker.

Increase of energy consumption after receiving a 'green' (efficient) flag
It is not the case that without the negative effects of the green flags on energy saving percentages, the Verspillingschecker effect would be in range of 5-10% savings. However, it makes one think about the effects of feedback and unexpected side-effects. A well-known study by Schultz, Nolan and Cialdini (2007) for a American utility called OPower showed already that telling low-consuming households they had performed better than average, increases energy consumption of those households. This potential destructive effect of providing positive feedback about what one is doing compared to similar others, was eliminated with the addition of an 'injunctive' message. When a smiley face (☺) was added as a social approval of the performance, low-consuming households did not start to increase their energy use. By 'simply' adding a happy emoticon to the households that did better than average, and a frown (☹) to households that did worse than average, all households started to decrease their energy usage.

Recommendation
Although the effects we found are small, we would advise Quby to think about possible side-effects of positive feedback and how to improve this. It is not uncommon that positive reinforcements backfire and cause an opposite effect. Positive feedback could possibly be expanded by adding a so-called 'injunctive norm' to the story: these refer to perceptions of what is commonly approved or disapproved of within a culture. For example by adding a message (a compliment) indicating that the desired behaviour is approved.

Verspillingschecker app development
The app and a selection of its use cases was launched at the beginning of December 2017. Quby used a start-up period in order to develop a well programmed and well-tested app. However there were of course some start-up issues. For example, the detection of the home appliances in the energy patterns was still under development during our research period, and therefore users sometimes received feedback on an appliance that they did not own, or they owned an energy efficient heat pump dryer but the app indicated they would have an energy inefficient dryer. These kinds of issues might have become apparent by comments given in the questionnaire as we asked to comment on their level of satisfaction and provide improvement suggestions.

Recommendation
For further development of the Verspillingschecker rich information can be found via the open ended questions of the questionnaire filled out by hundreds of Verspillingschecker users.

For this study, Quby provided weekly data of the energy use for each use case, and the associated diagnosis (inefficient or efficient) that was shown to a user. Diagnoses could change over time due to a change in behaviour (for example a more efficient appliance has been installed or people showed new behaviour). Unfortunately it was not possible to check whether users came back to the app an saw the updated diagnoses, as click-data was available. We only knew the so-called 'first time of interaction' of a user with a use case. At this time a user had to answer an online question about the use case, before they got to see the use-case related advice. Because we did not know whether users did see feedback that was given after the first time of use, we could not measure the impact of a change in diagnosis.

Recommendation
In order to be more certain that people see the feedback that is offered, and to be able to learn whether a change in diagnosis has an impact on behaviour, we would advise Quby to collect click data next to the data on the first time of interaction with a use case. In addition we would recommend to also collect click data about interaction with overall energy use. This would improve a possible ongoing or recurring evaluation of the effect of the app.

Direct feedback: Shorten time periods of seeing effect by a user.
From the literature it is clear that direct feedback can be very effective in changing behaviour. For instance using a display in the shower that provides direct feedback on water and energy use, led to average savings between 19% and 21% of energy consumption in the shower (Staake, Tiefenbeck, Schöb and Kupfer, 2016)[3]. When feedback is delayed the immediate call to act disappears and there is less effect on actual behaviour. At the moment of this research, the Verspillingschecker had a delay in its feedback. For example, if users were to buy a new appliance the app would register and display this after a number weeks (depending on the measurement period as described in Annex B).

Recommendations
It would be more rewarding for users of the app if their good behaviour would be acknowledged sooner. The option to add a new appliance by the user itself could possibly improve this.

Since the literature shows that direct feedback is most effective in changing behaviours, Quby could consider providing more direct feedback, for instance in the form of push messages to activate the users more. Because push messages can be experienced as obtrusive, this could for example be combined with a specific campaigning period in which people commit to limit energy waste as much as possible.

Next to the timing of feedback, the type of feedback is of importance. We know Quby explored for example including financial feedback (saving on costs), environmental feedback (saving on CO_2 emissions) or social feedback (what similar others do). Now the app is running, it might be possible to do small experiments with different feedback approaches and combinations.

Link between questionnaire and energy data
We did not find a strong effect of the Verspillingschecker on the level of overall energy use. When looking at the use cases we saw that small positive actions of users with an inefficient, red diagnosis, were in some case diminished by small negative actions of users who received an efficient diagnosis. The questionnaire showed that people indeed indicated undertaking actions to reduce energy waste. For example buying a new energy efficient appliance or using an appliance in an energy efficient way. These actions were however undertaken by people regardless of the type of diagnosis they received (either red or green), and therefore they were hard to retrieve from the data on a use case level Some actions were possibly also too small to make much of a difference on the use case and on the overall levels. At the same time we should also acknowledge that there can be a difference between what people say they do and what they actually do. It is a common thing that people provide socially desirable answers.

[3] Staake, T., Tiefenbeck, V., Schöb, S., & Kupfer, A. (2016). Effects of Real-Time Feedback on Hot Water Use . Final report on theAmphiro-PWN-study. Downloaded via https://www.amphiro.com/wp-content/uploads/2016/05/Amphiro_PWN_FinalReport_MainPart_2016_04_28.pdf

Recommendation
We did not look at so-called mediation effects on the overall energy use level, since the differences within the Verspillingschecker group and between the Verspillingschecker and control group were so small. At this point we also did not look into possible mediation effects on the use case level, but this could be of interest. For example, to what extent is the effect that results from receiving a green flag on the washing machine influenced by age or gender, or attitudes.

Unexpected side effects
Energy saving in households can come with unexpected side effects, which makes finding absolute effects difficult. For instance, when buying a new energy efficient refrigerator, the old one sometimes ends up in the basement as a second fridge. Insulation of one's home can lead to using and heating more rooms and thereby reducing the expected impact; a so-called rebound effect. And a field experiment by Tiefenbeck, Staake, Roth and Sachs (2013) aimed at saving water, showed that people in the experimental group indeed decreased their water use, but at the same time their electricity use increased[4]. This is an example of a negative spillover effect. And finally, as shown in the OPower study, a message intended to provide positive feedback and behaviour can lead to an increase of energy use.

Recommendation
In the Verspillingschecker app users receive per specific use case a message of being efficient (green) or being inefficient (red). Additionally they get advice on how to improve their energy use related to this use case. In general people dislike receiving a negative assessment and tend to focus more on negative than on positive messages. A negative message can be a motivation to do better, but it can also trigger thoughts like 'this app doesn't work', or 'I am not convinced it shows real measurements'. It can lead to rejecting the app or to bury one's head in the sand. On the other hand a positive message can provide people with a license to behave less energy efficient. Although it was no part of this research, these negative spillovers can also exist between different behaviours, for instance between buying a new appliance and showering behaviour. Finding the right tone of voice and frame of each message is therefore a precise matter.

[4] Tiefenbeck, V., Staake, T., Roth, K., & Sachs, O. (2013). For better or for worse? Empirical evidence of moral licensing in a behavioural energy conservation campaign. *Energy Policy*, *57*, 160-171.

Household interaction

The results from the questionnaire showed that among the Verspillingschecker users mainly males from two person households participated. As household energy use is divided among different persons, actions can differ from person to person. Choosing which program to use on the dishwasher or washing machine can differ between the one using the app or that filled in the questionnaire versus the one who actually mostly does the washing. It is therefore possible that actions reported in the questionnaire do not depict how it really goes.

Recommendation

Saving energy is a household matter, but often not the most urgent one that is regularly discussed. The app could provide in conversation starters between the app users and their partners, children, or other house mates. For example, by making it easy to share the advice from the app, or by proposing topics to discuss, like 'Do you think all programs on your washing machine are equally energy efficient? What program do your use? Ask the other members in your household which program they use'.

Different groups or an effect of the Verspillingschecker?

For a number of attitudes towards specific actions, we found small differences per group. Verspillingschecker users were for instance somewhat more positive about replacing a still working appliance than the control group. As it was not possible to do a questionnaire before and after the Verspillingschecker was launched, that could have provided insight in actual change, we developed questionnaire statements that included a time period (like, 'In 2018 I bought…'). For the stamements on attitude on actions we decided not to make a comparison with 'a year ago', since we expect people can not a make a reliable estimation on changes in attitudes over a one year period. It is therefore difficult to make sure that the difference between groups exist due to the Verspillingschecker. The difference in attitude towards replacing a still working appliance could well be caused by other things (for example, by Toon itself); factors that were already of effect before the Verspillingschecker app started to be used.

We tried as much as possible to rule out these alternative explanations. For instance we measured 'environmental identity' of all respondents. This is a stable factor, as identity does not change easily over time. The Verspillingschecker and Control group were equal on their environmental identity, indicating that general attitudes towards environmental acts should be quite equal between groups. We thus think that indeed it is likely that the app did have an effect here, although we cannot claim this as a fact.

Recommendation

In follow-up studies it would be best to incorporate a pre-test questionnaire as well in both groups. This way a change in attitude (and any other behavioural impact) could be ascribed to the Verspillingschecker as one can control for differences measured by the pre-test questionnaire.

The AIDA model

We used the AIDA model as a bases of the questionnaire. This model is often used in marketing, but has not been tested scientifically on energy behaviour. Also on other behaviours there is not scientific research on the model. We therefore had to operationize the model and develop various questions ourselves. This can be seen as a first step towards a validated set of questions. However, unfortunately we did not find a difference between the factors attention, interest and desire. These were all perceived as one thing, we called ‘attention for energy waste at home’. One explanation could be that our questions were not right. Another explanation could be that the factors the AIDA model distinghuishes are no separate things.

Recommendation

We would advice other researchers working on the AIDA model, to learn from our operationalization and improve where possible.

A Characteristics research groups research question 1 on overall energy use

As described to answer Question 1 we performed our analyses on an active group of Verspillingschecker users and matched this group with the control group. In this Verspillingschecker group 75% of the participants was male (and 25% was female), while in the control group 67% was male (versus 33% female). The groups differed somewhat in age, with an average age of 53 years in the Verspillingschecker group and an average of 60 years in the Control group. Figure 42 shows that in the control group more people fall in the pensionable age categories of 66-80 and >80 years.

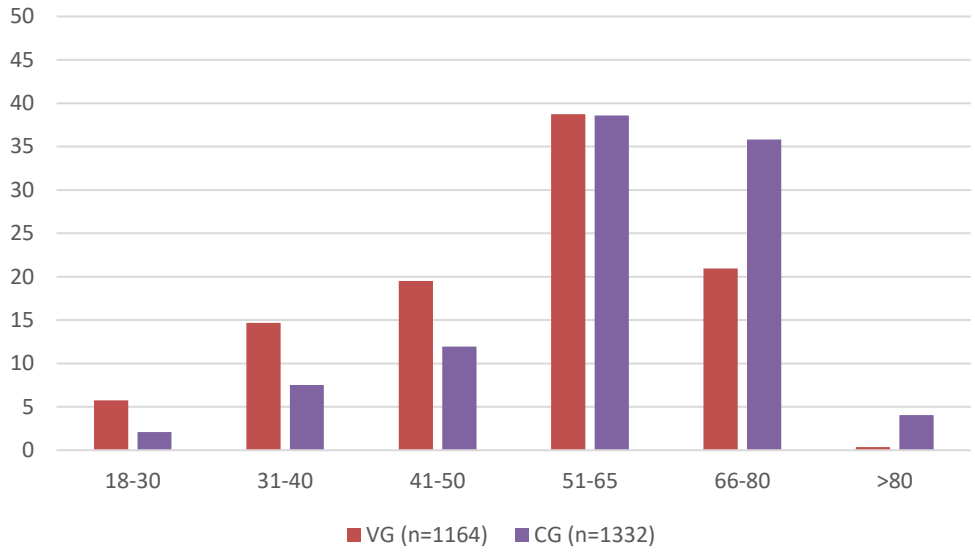

Figure 42 *Age categories (in %, per group)*

A majority of the participants was part of a two-person household (see Figure 43). The average household size for both the Verspillingschecker group and the Control group was 2,6. We saw some differences between the data on household size we received from Eneco/Quby and the figures respondents provided in the survey.

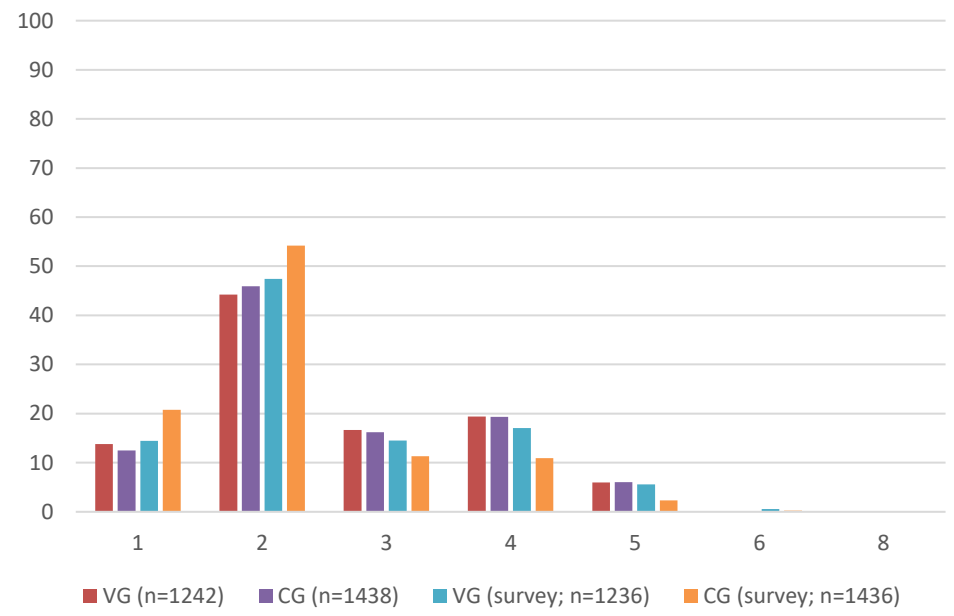

Figure 43 Household size (in % per group)

Most respondents (VC:68% and CG:67%) lived in a town house (see Figure 44). The figures we received from Eneco/Quby differed somewhat from the data respondents provided in the survey.

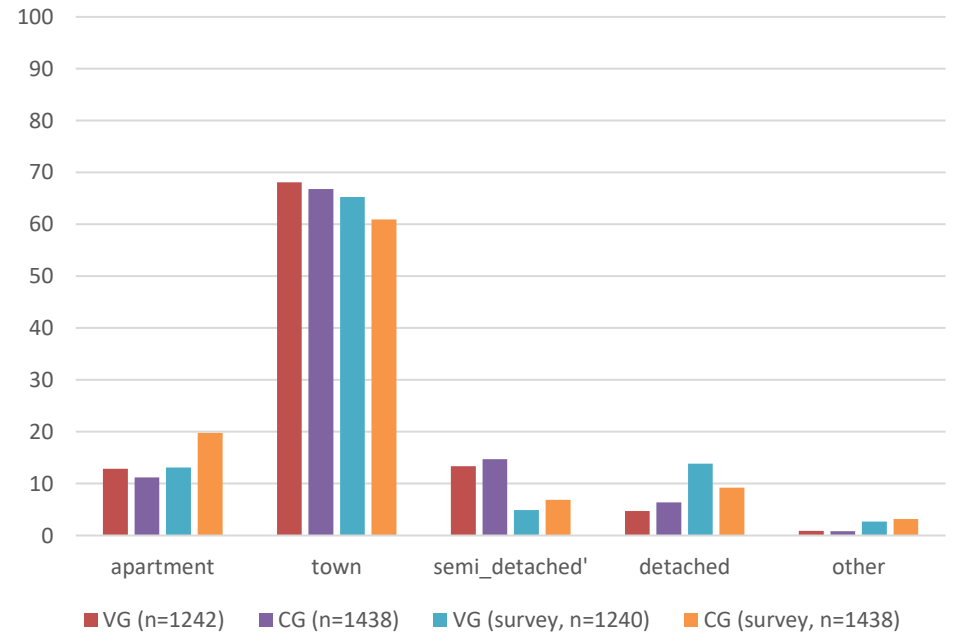

Figure 44 House type (in %)

In both groups around 70% of the respondents were home owners, while around 30% were tenants (see Figure 45).

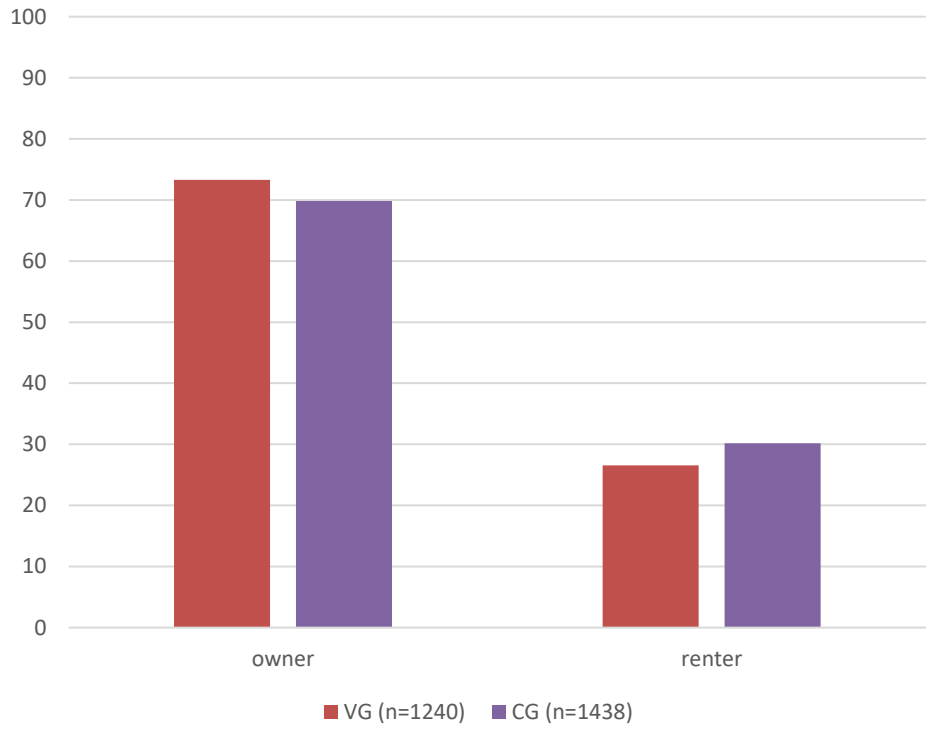

Figure 45 Home ownership (in %)

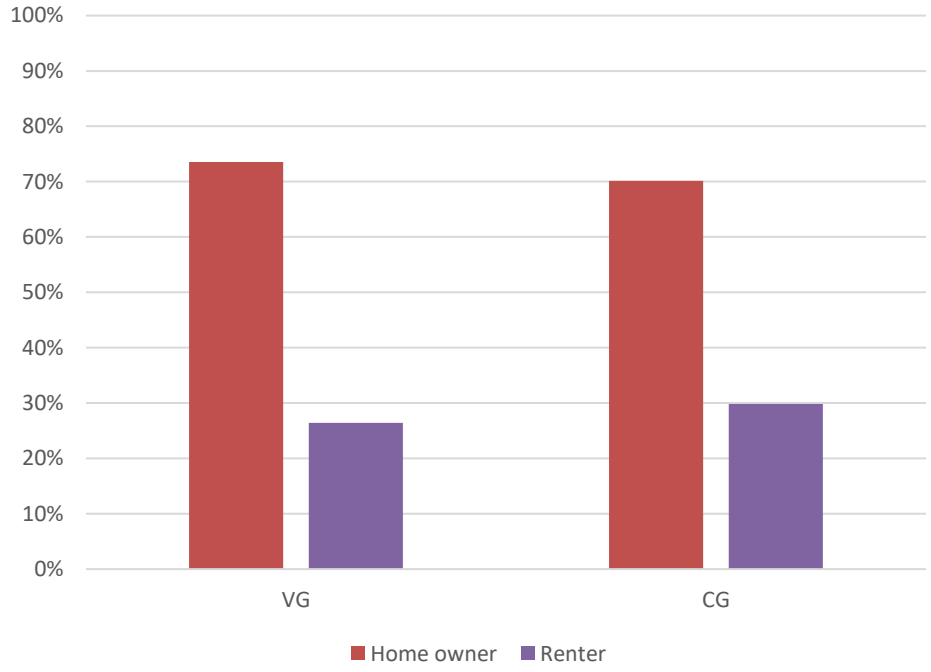

Figure 46 Home ownership (in %)

To answer Question 2 we included all participants that filled in the survey (N=5,796). Within the Verspillingschecker group 72% was male (n=2,532), while in the Control group 64% was male (n=3,237). The average age in the Verspillingschecker was 54 years while in the Control group the average age was 61 years. Figure 47 shows that in the control group more people fall in the pensionable age categories of 66-80 and >80 years.

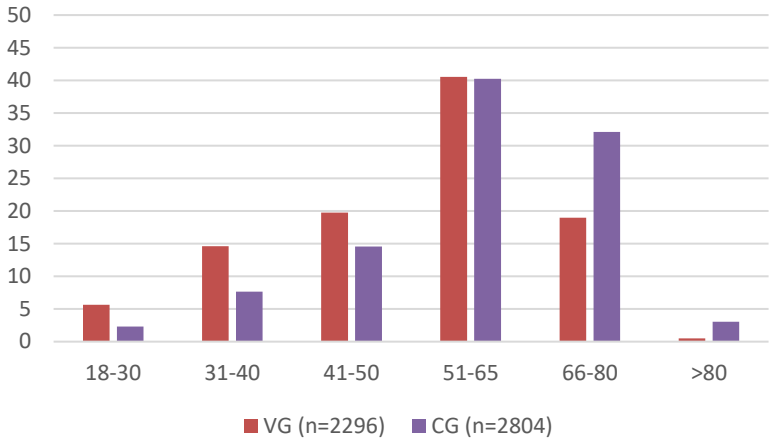

Figure 47 Age categories (in %, per group)

The average household size was 2.5 in the Verspillingschecker group and 2.3 in the control group. Most respondents lived in a 2-person household (see Figure 48).

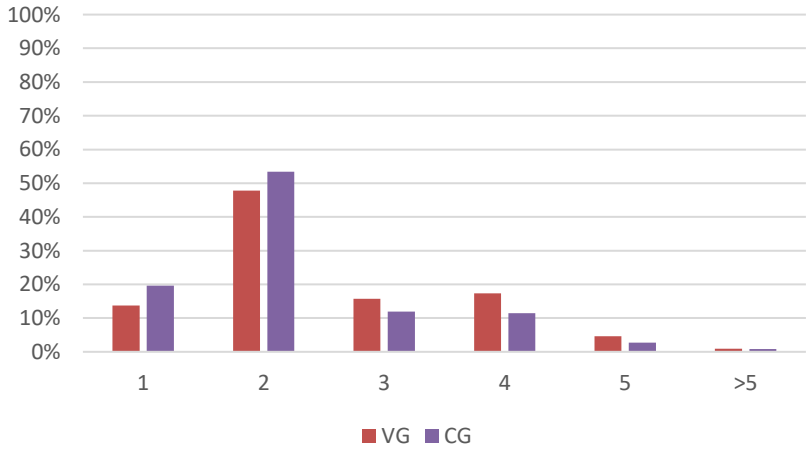

Figure 48 Household size (in % per group)

Most respondents (VC:64% and CG:58%) lived in a town house (see Figure 49).

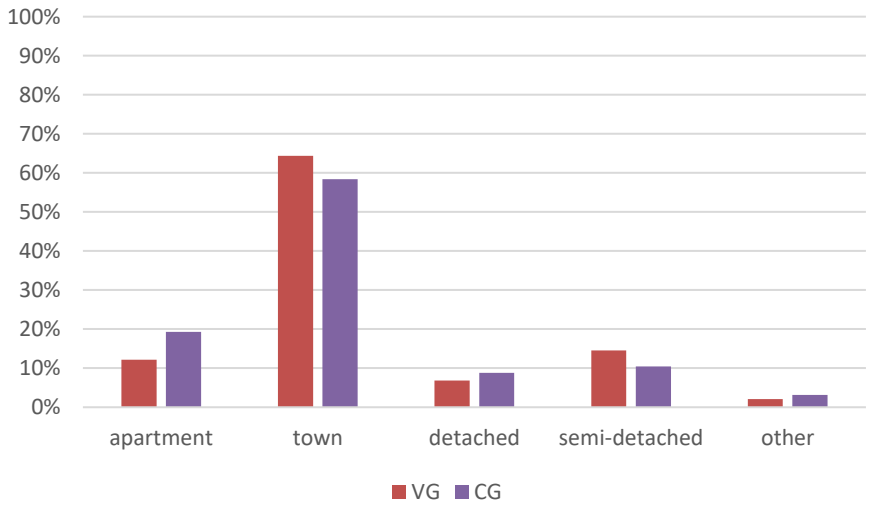

Figure 49 House type (in %)

And finally, in both groups around 70% of the respondents were home owners, while around 30% were renters (see Figure 50).

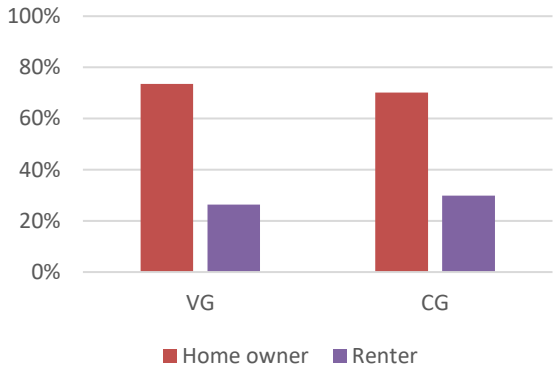

Figure 50 Home ownership (in %)

B Data collection per use case

The flags the Verspillingschecker provides to its users are based on energy data Quby collects from the households. The operationalisation per use case is described in the table below.

UC number	UC name	UC operationalization
1	hot_water_overall	Gas aggregate for hot water (based on 30 most recent non-heating days) in [m3]
3	Shower head	hot_water_gas_rate = Gas rate while using hot water in [l/minute] and gas_hot_water = Gas aggregate for hot water (based on 30 most recent non-heating days) in [m3]
5	heating_home_overall	Gas aggregate for heating (based on heating days in latest 30 day period and 30 most recent non-heating days &) in [m3]
8b	Toon Night	lowest_night_setpoint = Most common thermostat setpoint in last 14 days during the night in [Degrees]
9	electricity_overall	Electricity used over the last full calendar month for which data is available [kWh]
10	Refrigerator	load duration: the time that the compressor is running [minutes] 3 month period
11	Washing machine	Duration of the heating element in the washing cycle (duration detected with highest confidence above threshold e.g. 0.2) in [minutes] 3 month period
12	Dishwasher	Energy use in kWh. Sum over 3 month period
13	Dryer	The frequency of the detected appliance behaviour - number of detected heating blocks.
14	Stand-by electricity	Lowest electricity consumption in 5 minutes in [W] of last day

C Operationalization of factors

We defined a number of factors we wanted to research in the questionnaire. In the table below we describe how these factors were operationalized in questions.

Hypotheses en assumpties	Definition	Q nr
H5. Attention changes significantly after subjects in the test group used the VC app. H6. The VC group will on average show a stronger change in attention, than the CG.	Attention: Attention for energy waste and saving Do VC users pay more attention to information on energy use, saving than before? Do VC users pay more attention than the control group?	Q1a, Q1d
H5. Interest changes significantly after subjects in the test group used the VC app. H6. The VC group will on average show a stronger change in interest than the CG.	Interest: Interest in waste and saving possibilities. Do VC users have more interest in saving possibilities than before? Do VC users have more interest than the control group?	Q1b, Q13a
H5. Desire to take measures changes significantly after subjects in the test group used the VC app. H6. The VC group will on average show a stronger change in desire to take measures than the CG.	Desire: wish to waste less energy or to save more energy Do VC users have a stronger desire than before? Do VC users have a stronger desire than the control group?	Q1c, Q1e
H5. The amount of measures taken changes significantly after subjects in the test group used the VC app. H6. The VC group will on average show a stronger change in the amount of measures taken, than the CG.	Action: saving energy through investments, different settings, and new habits Do VC users act more than before? Do VC users act more than the control group?	Q2, Q3, Q4, Q5
H5. Attitude towards measures changes significantly after subjects in the test group used the VC app.	Attitude towards energy saving in home (perception* evaluation) regarding investments, settings, and habits. Investments: Do VC users have a more positive attitude towards investments in	Q6

Hypotheses en assumpties	Definition	Q nr
H6. The VC group will on average show a stronger change in attitude towards measures than the CG.	energy efficient appliances and their home (insulation, heating, solar panels) than before? Is their attitude more positive than the control group? Setting changes: Do VC use have more positive attitudes towards setting changes than before? Is their attitude more positive than the control group? Daily behaviour: Do VC users have a more positive attitude towards daily energy savings than before? Is their attitude more positive than the control group?	Q1a Q7, Q8, Q12 Q1a Q9, Q10, 11 Q1a
Assumption 2: The generated personal advice offers home owners sufficient insight in their energy conservation potential.	Insight in energy conservation: Do VC users have more insight than before? Do they have more insight than the control group?	Q1f
Assumption 3: The generated personal advice provides home owners insight in what energy conservation measures are the most interesting for them.	Insight in energy measures: Do VC users have more insight than before? Do they have more insight than the control group?	Q1g
Assumption 4: The generated personal advice helps home owners in taking investment decisions concerning energy conservation measures.	Help in taking investments; Do VC users experience being helped more than before? Do they experience this more than the control group?	Q13b, Q13c
Specific questions for VC users.	Satisfaction with and opinions on the app	Q14 – Q21
Personal characteristics	Gender, age, house type, rental vs owner, household size, home living surface, building year, solar panels, income level	Q22 – Q32
Permission to link datasets		Q33

D Questionnaires Verspillingschecker

Two versions of the questionnaire were distributed: one for the group of Verspillingschecker users and one for the control group (Eneco clients without Toon). The two versions were kept alike as much as possible.

VERSION 1: VERSPILLINGSCHECKER USERS (IN DUTCH)

Introductie
Bedankt dat u wilt meedoen aan dit onderzoek. We willen u een aantal vragen stellen over hoe u omgaat met energie in huis. Het invullen van de vragenlijst duurt ongeveer 10 minuten.

Allereerst willen we u een aantal algemene vragen over energieverbruik voorleggen. In hoeverre bent u het eens met de volgende stellingen? *NB Stellingen voorleggen in random volgorde*

Q1d. Ik zie steeds meer informatie over energie verbruiken, verspillen en besparen in huis.
Q1e. Een energiezuinig huis staat hoger op mijn 'verlanglijst' dan een jaar geleden.
Q1f. Ik weet nu beter hoeveel energie ik in mijn huis kan besparen dan een jaar geleden.
Q1g. Ik weet niet wat ik nog meer in mijn huis kan doen om energie te besparen.

Q1a. Ik heb nu meer aandacht voor verspilling van energie in huis dan een jaar geleden.
Q1b. Ik heb meer nu meer interesse in mogelijkheden om energieverspilling in huis tegen te gaan dan een jaar geleden.
Q1c. Ik heb nu een sterkere wil dan een jaar geleden om energieverspilling in mijn huis tegen te gaan.

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens,

Q2. Heeft u **in 2018** één of meer van de volgende dingen in huis gedaan om energieverspilling tegen te gaan? *(in random volgorde voorleggen)*

- De vaatwasser op 50 graden (of Eco-mode) zetten
- Temperatuur van de koelkast instellen tussen 3 en 4 graden C.
- Met schakelklok instellen wanneer je lampen of apparatuur uit moeten
- Sluipverbruik opspeuren
- Weekprogramma voor de thermostaat instellen
- Radiatoren op sommige kamers uitzetten
- Radiatorfolie tussen de verwarming en muur stoppen
- Geen van deze

Q3. Heeft u **in 2018** één of meer van de volgende dingen in huis gedaan om energieverspilling tegen te gaan? *(in random volgorde voorleggen)*

- De vaatwasser pas aanzetten als hij helemaal vol is
- Kleren drogen aan de waslijn/ een wasrek

- Kleding op 30 graden of lager wassen
- De warme kraan van de douche wat minder ver opendraaien
- Stekkers op één stekkerblok groeperen en dan allemaal tegelijk écht uitzetten.
- Waterbesparende douchekop aanschaffen
- Geen van deze

Q4. Heeft u **in 2018** één of meer van de volgende apparaten aangeschaft?
- Vaatwasser
- Warmtepompdroger
- Koelkast
- Wasmachine
- Geen van deze

Toon alleen bij Q4 genoemde apparaten, vraag overslaan indien niets gekocht
Q5. Wat is het energielabel van dit nieuwe apparaat?
- Vaatwasser
- Warmtepompdroger
- Koelkast
- Wasmachine

A+++, A++, A+, een lager label, weet ik niet

Q6. In hoeverre bent u het eens met de volgende stellingen?
Een **nog werkende** koelkast, droger, wasmachine, of afwasmachine **vervangen** voor een zuinigere versie…
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan
b. vind ik de investering niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel
e. vind ik normaal

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens,

Vraag Q7-Q12 in random volgorde voorleggen (om volgorde effecten te voorkomen)

Nu volgen er zes vragen over hoe u denkt over zes verschillende soorten energiezuinig gedrag in uw huis.

Q7. In hoeverre bent u het eens met de volgende stellingen?
Apparaten **standaard** op de ecostand (of het zuinige programma) zetten……
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan
b. vind ik de moeite niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel

e. vind ik normaal

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

Q8. In hoeverre bent u het eens met de volgende stellingen?
Op lage temperaturen wassen…
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan
b. vind ik de moeite niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel
e. vind ik normaal

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

Q9. In hoeverre bent u het eens met de volgende stellingen?
Apparaten niet op stand-by laten staan, maar écht uit zetten…
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan
b. vind ik de moeite niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel
e. vind ik normaal

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

Q10. In hoeverre bent u het eens met de volgende stellingen?
Onder de douche zuinig omgaan met warm water…
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan
b. vind ik de moeite niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel
e. vind ik normaal

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

Q11. In hoeverre bent u het eens met de volgende stellingen?
De verwarming verlagen als ik de deur uit ga…
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan
b. vind ik de moeite niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel
e. vind ik normaal

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

Q12. In hoeverre bent u het eens met de volgende stellingen?
De verwarming 's nachts lager zetten…
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan
b. vind ik de moeite niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel
e. vind ik normaal

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

TOON VERSPILLINGSCHECKER
Nu willen we u een aantal vragen stellen over Eneco, Toon en de Verspillingschecker.

In hoeverre bent u het eens met de volgende stellingen? *NB Stellingen voorleggen in random volgorde*

Q13a. Eneco heeft mij de afgelopen tijd nieuwsgierig gemaakt naar wat ik kan doen om verspilling in mijn huis tegen te gaan.
Q13b. Eneco helpt mij om energieverspilling tegen te gaan.
Q13c. Eneco geeft mij inzicht in wat energiebesparende acties of maatregelen mij opleveren.

Q14. Hoe tevreden bent u over Toon®? Schuifje
Rapportcijfer tussen 1 en 10 (1 = zeer ontevreden, 10 = zeer tevreden)

Q15. Sinds december 2017 is de nieuwe Toon app met daarin de Verspillingchecker beschikbaar. Heeft u deze app gedownload?
- Ja
- Ja maar ik heb hem er weer afgehaald
- Nee
- Weet ik niet

Volgende vragen alleen indien nieuwe app gedownload
Q16. De Verspillingschecker is een nieuwe functie waarmee je energieslurpers in huis opspoort. De Verspillingschecker biedt persoonlijke bespaartips en slimme inzichten per apparaat of situatie.
Hoe tevreden bent u over de Verspillingschecker? Schuifje
Rapportcijfer tussen 1 en 10 (1 = zeer ontevreden, 10 = zeer tevreden)
+optie "weet ik niet, (nog) geen gebruik van gemaakt"
→ indien geen gebruik, door naar achtergrondvragen

Q17. Ruimte voor toelichting
Open vraag, niet verplicht

Q18. In hoeverre bent u het eens met de volgende stellingen? (voorleggen in random volgorde)

De adviezen die ik van de Verspillingschecker kreeg …
a. Zijn geloofwaardig
b. Zijn voor mij uitvoerbaar
c. Vind ik interessant

Q19. In hoeverre sluit de Verspillingschecker aan bij uw eerdere vermoedens/eigen ideeën over hoe u energieverspilling in huis tegen kunt gaan?
- Mijn vermoedens zijn bevestigd: de adviezen sluiten aan bij wat ik al dacht.
- Mijn vermoedens zijn ontkracht: de adviezen zijn anders dan ik dacht.
- Ik had geen vermoedens.

Q20. De Verspillingschecker is gemakkelijk in het gebruik.
- Eens, licht toe
- Oneens, licht toe.

Q21. Heeft u suggesties hoe Eneco de Verspillingschecker (nog) interessanter voor u kan maken?
OPEN vraag + optie ”nee, geen suggesties”

Persoonlijke kenmerken
Tot slot willen we u nog enkele vragen stellen over u zelf

Q22. Bent u man of vrouw? [mogelijkheid om over te slaan]

Q23. Wat is uw leeftijd?
---- jaar / wil ik niet zeggen

Q24. Woont u in een koopwoning of een huurwoning?
Koopwoning
Huurwoning

Q25. In welke type woning woont u?
Appartement
Rijwoning (tussen of hoek)
Vrijstaand
Twee-onder-één-kap
Anders

Q26. Wat is het bouwjaar van uw woning? (als u dit niet exact weet, probeert u dan een inschatting te maken)
…
Weet ik niet

Q27. Wat is de totale vloeroppervlakte van uw woning in vierkante meters (m2)? (als u dit niet exact weet, probeert u dan een inschatting te maken)
...
Weet ik niet

Q28. Uit hoeveel personen bestaat uw huishouden?
…

Q29. Heeft u zonnepanelen voor elektriciteit op uw eigen dak?
Ja
Nee

Q30. Welke apparaten, apps of websites gebruikt u bij uw slimme meter om inzicht te krijgen in uw energieverbruik?
- Toon
- Eneco app
- Nog een andere, namelijk …
- Geen

Q31. Wat is het maandelijks netto inkomen van uw hele huishouden?
Minder dan €1000
Tussen de €1000 en €1500
Tussen de €1500 en €2000
Tussen de €2000 en €2500
Tussen de €2500 en €3000
Tussen de €3000 en €3500
Tussen de €3500 en €4000
Tussen de €4000 en €4500
Meer dan €4500
Weet niet/ Wil niet zeggen

Q32. In hoeverre bent u het eens of oneens met de volgende stellingen
Energie besparen is een belangrijk deel van wie ik ben
Ik ben het type persoon dat energie bespaart
Ik zie mijzelf als een energiebesparend persoon

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

Toestemming:

Q33. Wij zouden graag uw antwoorden op deze vragenlijst willen koppelen aan uw energieverbruiksgegevens. Wij kunnen Toon en de Verspillingschecker functie daarmee nog verder verbeteren. Uw gegevens worden anoniem verwerkt. Geeft u hier toestemming voor?

- Uw gegevens worden anoniem verwerkt: U bent op geen enkele manier in de resultaten als persoon terug te vinden.
- Het gaat uitsluitend om uw wekelijkse gas- en elektriciteitsverbruik in de periode van september 2017 tot en met april 2018.
- De geanonimiseerde dataset wordt beschikbaar gesteld aan onderzoeksbureau TNO, dat het onderzoek uitvoert.
- Uw gegevens worden niet doorgespeeld aan andere derden en niet gebruikt voor commerciële doeleinden.

- Ja ik geef toestemming om mijn gegevens uit deze vragenlijst te koppelen aan mijn energieverbruiksgegevens van bovengenoemde periode.
- Nee

Hartelijk bedankt voor uw medewerking!

VERSION 2: CONTROL GROUP (IN DUTCH)

Introductie
Bedankt dat u wilt meedoen aan dit onderzoek. We willen u een aantal vragen stellen over hoe u omgaat met energie in huis. Het invullen van de vragenlijst duurt ongeveer 10 minuten.

Allereerst willen we u een aantal algemene vragen over energieverbruik voorleggen. In hoeverre bent u het eens met de volgende stellingen?

Q1d. Ik zie steeds meer informatie over energie verbruiken, verspillen en besparen in huis.
Q1e. Een energiezuinig huis staat hoger op mijn verlanglijst dan een jaar geleden.
Q1f. Ik weet nu beter hoeveel energie ik in mijn huis kan besparen dan een jaar geleden.
Q1g. Ik weet niet wat ik nog meer in mijn huis kan doen om energie te besparen.

Q1a. Ik heb nu meer aandacht voor verspilling van energie in huis dan een jaar geleden.
Q1b. Ik heb meer nu meer interesse in mogelijkheden om energieverspilling in huis tegen te gaan dan een jaar geleden.
Q1c. Ik heb nu een sterkere wil dan een jaar geleden om energieverspilling in mijn huis tegen te gaan.

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

Q2. Heeft u **in 2018** één of meer van de volgende dingen in huis gedaan om energieverspilling tegen te gaan? *(in random volgorde voorleggen)*

- De vaatwasser op 50 graden (of Eco-mode) zetten
- Temperatuur van de koelkast instellen tussen 3 en 4 graden C.
- Met schakelklok instellen wanneer je lampen of apparatuur uit moeten
- Sluipverbruik opspeuren
- Weekprogramma voor de thermostaat instellen
- Radiatoren op sommige kamers uitzetten
- Radiatorfolie tussen de verwarming en muur stoppen
- Geen van deze

Q3. Heeft u **in 2018** één of meer van de volgende dingen in huis gedaan om energieverspilling tegen te gaan? *(in random volgorde voorleggen)*

- De vaatwasser pas aanzetten als hij helemaal vol is
- Kleren drogen aan de waslijn/ een wasrek
- Kleding op 30 graden of lager wassen
- De warme kraan van de douche wat minder ver opendraaien
- Stekkers op één stekkerblok groeperen en dan allemaal tegelijk écht uitzetten.
- Waterbesparende douchekop aanschaffen
- Geen van deze

Q4. Heeft u **in 2018** één of meer van de volgende apparaten aangeschaft?

- Vaatwasser
- Warmtepompdroger
- Koelkast
- Wasmachine
- Geen van deze

Toon alleen bij Q4 genoemde apparaten, vraag overslaan indien niets gekocht
Q5. Wat is het energielabel van dit nieuwe apparaat?

- Vaatwasser
- Warmtepompdroger
- Koelkast
- Wasmachine

A+++, A++, A+, een lager label, weet ik niet

Q6. In hoeverre bent u het eens met de volgende stellingen?
Een **nog werkende** koelkast, droger, wasmachine, of afwasmachine **vervangen** voor een zuinigere versie…
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan (oneens…. Eens)
b. vind ik de investering niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel
e. vind ik normaal

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

Vraag Q7-Q12 in random volgorde voorleggen (om volgorde effecten te voorkomen)

Q7. In hoeverre bent u het eens met de volgende stellingen?
Apparaten **standaard** op de ecostand (of het zuinige programma) zetten…
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan (oneens…. Eens)
b. vind ik de moeite niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel
e. vind ik normaal

Q8. In hoeverre bent u het eens met de volgende stellingen?
Op lage temperaturen wassen…
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan (oneens…. Eens)
b. vind ik de moeite niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel

e. vind ik normaal

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

Q9. In hoeverre bent u het eens met de volgende stellingen?
Apparaten niet op stand-by laten staan, maar écht uit zetten…
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan (oneens…. Eens)
b. vind ik de moeite niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel
e. vind ik normaal

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

Q10. In hoeverre bent u het eens met de volgende stellingen?
Onder de douche zuinig omgaan met warm water…
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan (oneens…. Eens)
b. vind ik de moeite niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel
e. vind ik normaal

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

Q11. In hoeverre bent u het eens met de volgende stellingen?
De verwarming verlagen als ik de deur uit ga…
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan (oneens…. Eens)
b. vind ik de moeite niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel
e. vind ik normaal

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

Q12. In hoeverre bent u het eens met de volgende stellingen?
De verwarming ‘s nachts lager zetten…
(NB voorleggen in random volgorde)

a. helpt verspilling tegen te gaan (oneens…. Eens)
b. vind ik de moeite niet waard
c. vind ik gedoe
d. geeft mij een goed gevoel
e. vind ik normaal

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

ENECO
Nu willen we u een aantal vragen stellen over Eneco.

In hoeverre bent u het eens met de volgende stellingen? *NB Stellingen voorleggen in random volgorde*

Q13a. Eneco heeft mij de afgelopen tijd nieuwsgierig gemaakt naar wat ik kan doen om verspilling in mijn huis tegen te gaan.
Q13b. Eneco helpt mij om energieverspilling tegen te gaan.
Q13c. Eneco geeft mij inzicht in wat energiebesparende acties of maatregelen mij opleveren.

Persoonlijke kenmerken
Tot slot willen we u nog enkele vragen stellen over u zelf

Q22. Bent u man of vrouw? [mogelijkheid om over te slaan]

Q23. Wat is uw leeftijd?
---- jaar / wil ik niet zeggen

Q24. Woont u in een koopwoning of een huurwoning?
Koopwoning
Huurwoning

Q25. In welke type woning woont u?
Appartement
Rijwoning (tussen of hoek)
Vrijstaand
Twee-onder-één-kap
Anders

Q26. Wat is het bouwjaar van uw woning? (als u dit niet exact weet, probeert u dan een inschatting te maken)
…
Weet ik niet

Q27. Wat is de totale vloeroppervlakte van uw woning in vierkante meters (m2)? (als u dit niet exact weet, probeert u dan een inschatting te maken)
…
Weet ik niet

Q28. Uit hoeveel personen bestaat uw huishouden?
…

Q29. Heeft u zonnepanelen voor elektriciteit op uw eigen dak?
Ja
Nee

Q30. Welke apparaten, apps of websites gebruikt u bij uw slimme meter om inzicht te krijgen in uw energieverbruik?
- Toon
- Eneco app
- Nog een andere, namelijk …
- Geen

Q31. Wat is het maandelijks netto inkomen van uw hele huishouden?
Minder dan €1000
Tussen de €1000 en €1500
Tussen de €1500 en €2000
Tussen de €2000 en €2500
Tussen de €2500 en €3000
Tussen de €3000 en €3500
Tussen de €3500 en €4000
Tussen de €4000 en €4500
Meer dan €4500
Weet niet/ Wil niet zeggen

Q32. In hoeverre bent u het eens of oneens met de volgende stellingen
Energie besparen is een belangrijk deel van wie ik ben
Ik ben het type persoon dat energie bespaart
Ik zie mijzelf als een energiebesparend persoon

7 punt schaal: 1 = helemaal niet mee eens, 7 = helemaal mee eens

Toestemming:

Q33. Wij zouden graag uw antwoorden op deze vragenlijst willen koppelen aan uw energieverbruiksgegevens. Wij kunnen onze dienstverlening en Toon, onze slimme thermostaat, daarmee nog verder verbeteren. Uw gegevens worden anoniem verwerkt. Geeft u hier toestemming voor?

- Uw gegevens worden anoniem verwerkt: U bent op geen enkele manier in de resultaten als persoon terug te vinden.
- Het gaat uitsluitend om uw wekelijkse gas- en elektriciteitsverbruik in de periode van september 2017 tot en met april 2018.
- De geanonimiseerde dataset wordt beschikbaar gesteld aan onderzoeksbureau TNO, dat het onderzoek uitvoert.
- Uw gegevens worden niet doorgespeeld aan andere derden en niet gebruikt voor commerciële doeleinden.

• Ja ik geef toestemming om mijn gegevens uit deze vragenlijst te koppelen aan mijn energieverbruiksgegevens van bovengenoemde periode.
• Nee

Hartelijk bedankt voor uw medewerking!

E Validation and reliability checks for the Verspillingschecker (provided by Quby)

The Verspillingschecker (Waste Checker) was launched as part of the Toon app in December 2017. It generates automated individual insights and advices about where energy is wasted at a user's home. There are seven specific and three overall topics (use cases) spread over three domains, being gas for heating the home, gas for heating water and electricity.

The algorithms and approaches taken across the varying use cases have been validated scientifically by the development teams at Quby in collaboration with experts from ECN, TU Delft, University of Bonn, Milieu Centraal and Ecofys.

The algorithms to detect inefficiency of white goods have been validated extensively using over 150 homes fitted with smart plugs. The smart plugs give precise details of the energy consumption of individual white good appliances which can be compared to the detections obtained via disaggregation algorithms based on meter data. This approach has allowed the researchers at Quby to quantify accuracy and adapt the formulation of the algorithms.

For each use case Quby performed up to five levels of validation test:
- Lab based signature test
- Field test 10 users – detection test
- Field test 150 users – detection test
- Field test 150 users – use case advice test
- Production - user feedback tests

All tests for the launched 7 use cases confirmed that the reliability of inefficiency measurements was sufficient.

Since the app has been launched, to ensure scalability and reliability we have enabled CI/CD pipelines and have automated monitoring running on the data pipelines.