

## More Than Meets the Eyes: Complementing Surveys With Mobile Phone Digital Data Trail

Ivano Bison, Giuseppe A. Veltri, Matteo Busso, Mattia Zeni, Enrico Bignotti, Fausto Giunchiglia



[giuseppe.veltri@unitn.it](mailto:giuseppe.veltri@unitn.it)  
[ivano.bison@unitn.it](mailto:ivano.bison@unitn.it)





Density of residences

Where the students are at 1 PM from Monday to Saturday

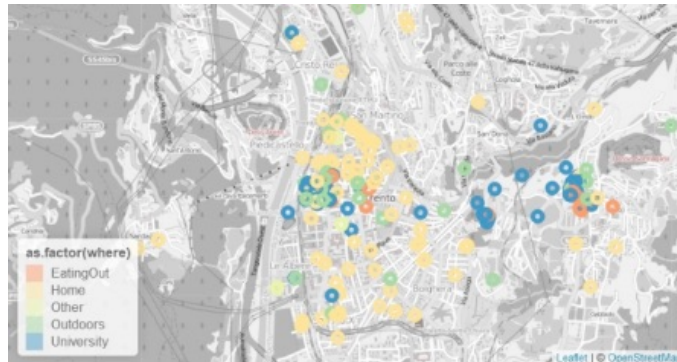
Monday



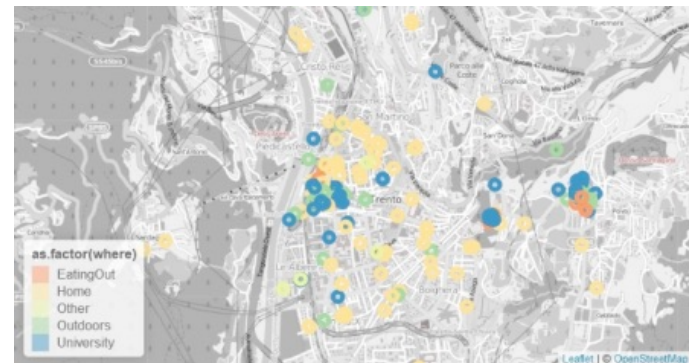
Tuesday



Wednesday



Thursday



Friday



Saturday



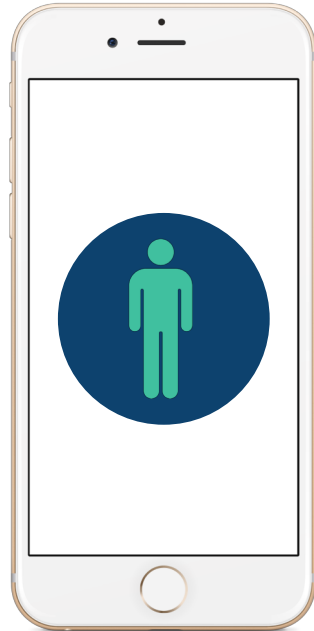


i-Log

---

# The user as source of information

The philosophy: Human centered computing



The user is the most important **source** of information



The user needs to **collaborate** and **help** the machine



For doing so, we need to leverage on the **tools and approaches** used in Sociology to understand **human behavior**



The most important tool are **Time Diaries** used as annotations for the sensor data

# i-Log

... new and specific possibilities ...

i-Log presents new and specific possibilities, that concern the sociological and urban field in three ways:



The first consists in the ability to investigate the space through the phone sensors, such as that related to the geolocation.



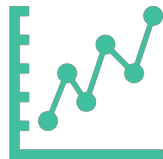
The second aspect is an improvement of the time diaries especially structured one, where respondents are asked to indicate three main dimensions of their everyday life: the activities performed, the locations they visit and the people around them.



The third innovative aspect of the app is to produce highly standardized and comparable information. This is highlighted by a fundamental sub-aspect of the app, that is the longitudinal form of the surveys, that are either internal to the experiment - repeating the observations for weeks, each day can be considered a unit of analysis – or external, replicating the experiment either with different samples - for comparative purposes - or with the same in different moments of time.

# i-Log

How can we collect data from the users?



## Collect streaming sensor data

The application can collect up to 39 streams of sensor data from the user smartphone, generating up to 1GB/day + audio, photo, bluetooth ...



## Collect user annotations as Time Diaries

The user is the only one who can correctly annotate her day life situations so that the machine can learn how to recognize them in the future



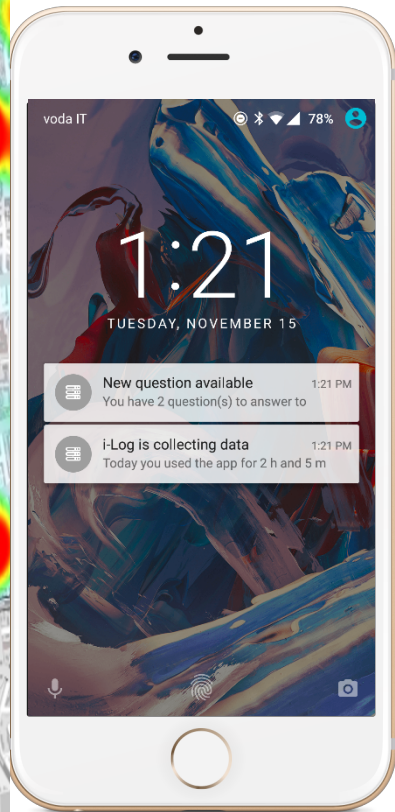
## Configurability and adaptability

The application adapts to different use cases and to different smartphones



## Provide services

i-Log is also the platform that will provide service back to the users, generated from the knowledge extracted from her data





TUESDAY, NOVEMBER 8



**New question available**

9:31 PM

You have 2 question(s) to answer to



**iLog is collecting data**

9:31 PM

Today you used the app for 21 h and 5 m



## What did you do at 04:59?

Lesson

Study

Eating

Self care

En route

Social life

Social media & Internet

Cultural activity

Sport

Shopping


Hobbies



# The survey

# SmartUnitn One

The last real life adoption on UniTrento students

 Understand how students' time allocation affects their academic performances



Last real life  
experiment  
9 May-7 June 2018



Sampled 308 participants  
35 quit the first day  
185 followed the tasks for  
at least 2 weeks



4 weeks asked  
2 weeks every ½ hour  
2 weeks every 2 hour



1TB of data  
Small data:  
7.400.000 records  
2212 observation days



## The main questionnaire.

Before the begin of the experiment an closed-ended questionnaire was administrated to investigate their habits and routine of participants.

The first part contains general information about students live: kind of domicile and address, parent's residence, and information about flatmates.

The second part is about student's habit and routines, concerning how and how much they travel for study and which kind of job they had during the semester (if they had one).

The third part is about student's psychological-social profiling:

- Big Five Inventory (10-item);
- Irrational Procrastination Scale (IPS);
- Smartphone Addiction Scale (SAV-SV)
- Perceived Stress Scale.



## Sample distribution by Department

Department	Population	Sample
Economics and Management	17.6	16.3
Physics	2.9	6.0
Civil, Environmental and Mechanical Engineering	8.8	7.6
Engineering and Information Science	6.4	6.5
Industrial Engineering	8.9	9.8
Humanities and Philosophy	14.9	13.6
Mathematics	2.9	6.5
Psychology and Cognitive Sciences	7.2	7.1
Sociology and Social Research	10.8	10.9
Faculty of Law	20.4	15.8
N.	10006	184



## Systematic vs. Random errors

# Answer timing and accuracy



## **Do they follow the task?**

Answer the following multiple-choice questions every half hour:

What are you doing?

Where are you?

With whom are you?

How do you feel? (a five-point scale on emotional status.)

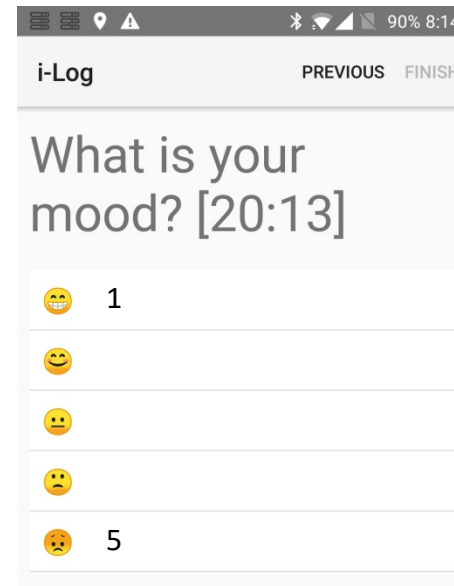


## **Do they answer correctly?**

Where are you?

- At Home (GPS vs. Human)

Multilevel model: Time from notification to answer: 2 levels, random intercept.



Var.Name	B
Intercept	4.40
Mood	-4.13 ***
Procrastination	1.45 ***
Men	7.09
Age	0.80
Sleeping	79.67 ***
Selfcare	-7.55 ***
Study	2.80 ***
Lesson	-0.31
Social Life	3.04 ***
Tv, Youtube	-4.30 ***
Social Media, Chat, etc.	-14.53 ***
Free Time	6.96 ***
Work	8.92 ***
Travel	-5.23 ***
Povo (hard science)	-10.91 **
Rovereto (psychology)	-15.50
Monday	-3.24 ***
Tuesday	-4.17 ***
Wednesday	0.02
Thursday	-10.77 ***
Friday	-3.96 ***
Saturday	0.25
R2	0.20
Level 2 Intercept Var.	724.66 ***
Level 1 Intercept Var.	3622.19 ***

What is your mood (1-5)?

- The more unhappy I am, the faster my answer will be.

Irrational procrastination scale (9 Items)

- The higher the level procrastination, the more I delay the answer

## Multilevel model: Distance from their own House. 2 levels, random intercept.

Var.Name	B
Intercept	11.55
Delta answer (in min.) <sup>+</sup>	0.67***
Time.notification.GPS <sup>*+</sup>	24.43***
Accuracy <sup>+</sup> (in meter)**	-0.51
Time answer (in sec.)	
0-3	88.32***
4-6	ref.
7-8	40.27
9-11	87.03***
12-14	103.53***
16-18	117.38***
19+	129.44***
R2	0.12
Level 2 Intecept Var.	59809.26***
Level 1 Intecept Var.	506378.34***

Note:

(+) zero on grand mean

(\*) Time elapse from notification to the first available GPS < 10 min.

(\*\*) Accuracy < 100 m.



### Accuracy:

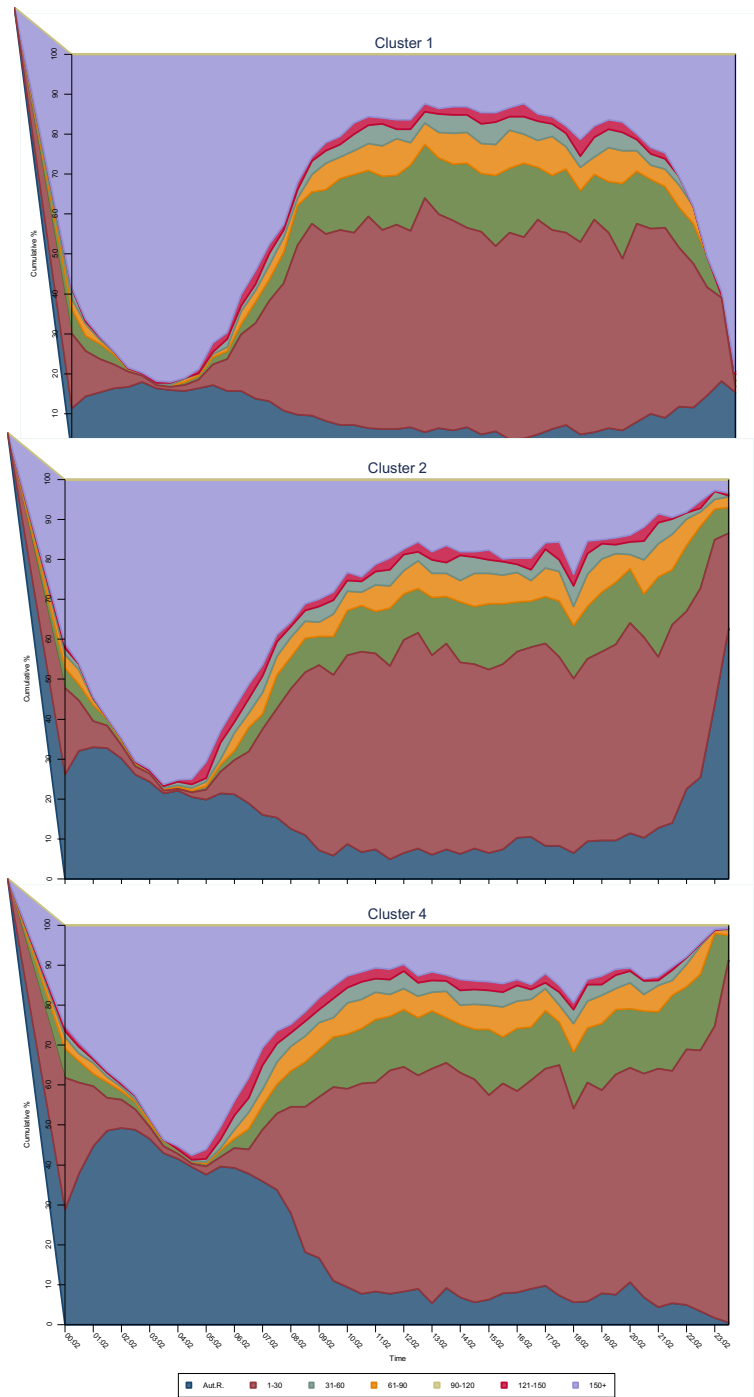
- Every minute of delay increase of 70 cm. the inaccuracy. (memory effect)
- Fill faster or too slowly the questionnaire increase the inaccuracy. (attention effect)



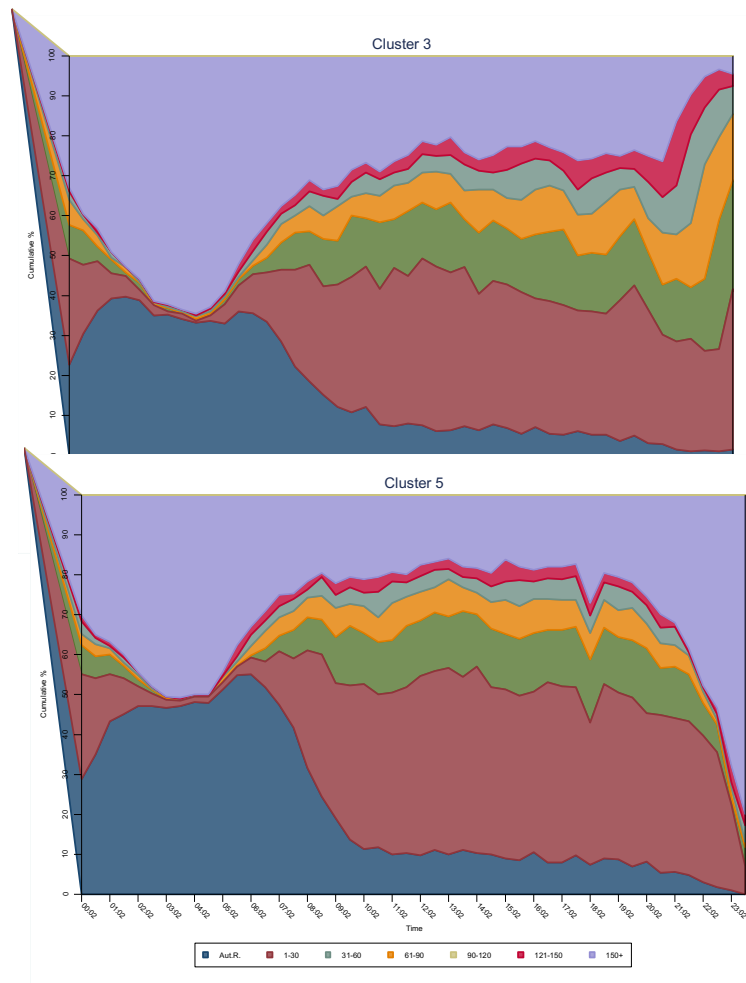
Do they follows pattern of answers?

---

# Sequence Analysis



State distribution plot



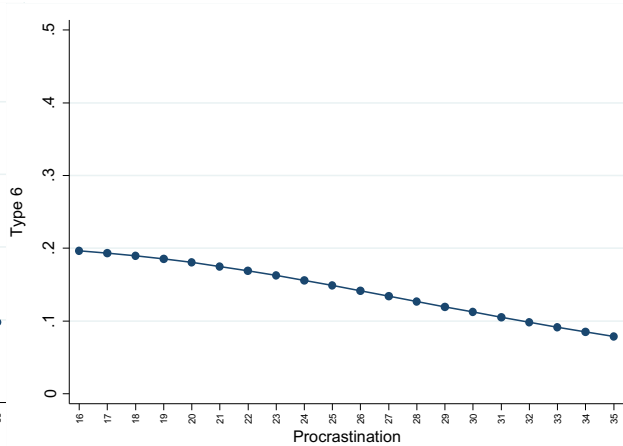
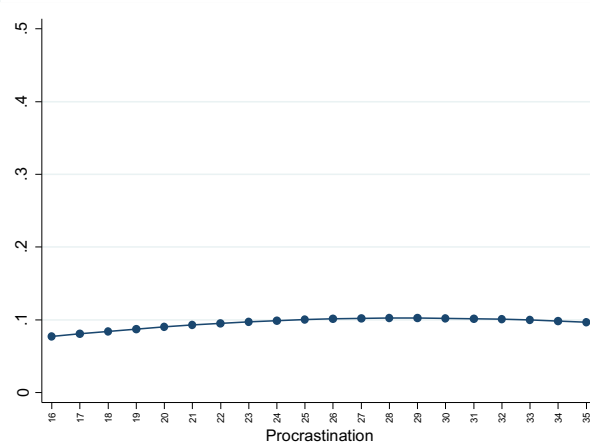
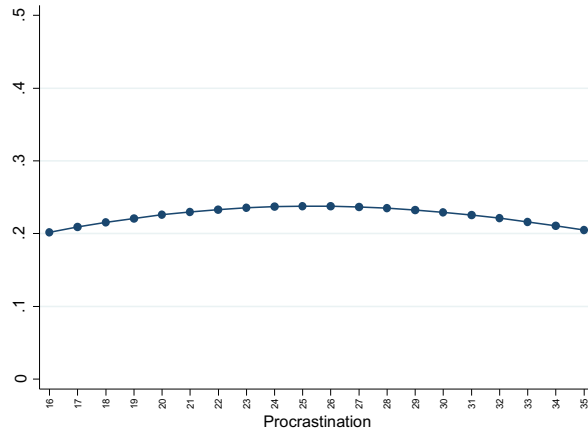
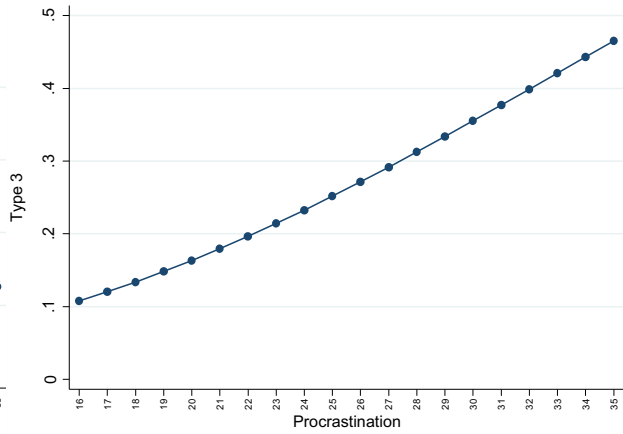
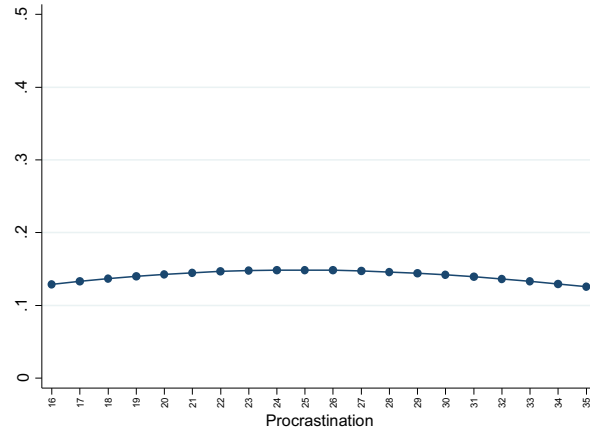
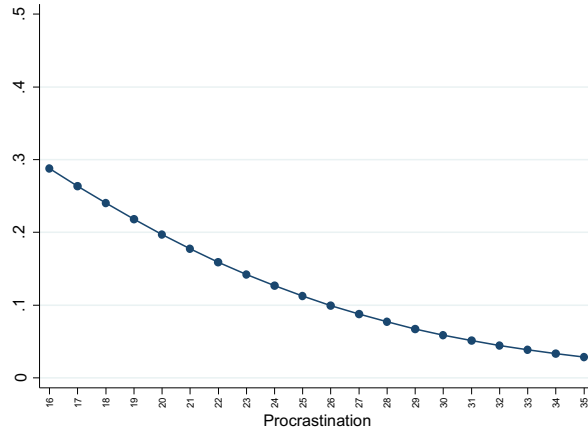
## Time response patterns by gender

	1	2	3	4	5	6	N.
F	20.8	15.6	14.3	15.6	9.1	24.7	77
M	8.6	11.1	34.6	28.4	8.6	8.6	81
<i>Total</i>	<i>14.6</i>	<i>13.3</i>	<i>24.7</i>	<i>22.2</i>	<i>8.9</i>	<i>16.5</i>	
<i>N.</i>	<i>23</i>	<i>21</i>	<i>39</i>	<i>35</i>	<i>14</i>	<i>26</i>	<i>158</i>

## Multinomial logistic regression of user profile by gender, place, course year and procrastination syndrom

	1	2	4	5	6
<i>Gender</i>					
Women(ref.)					
Men	-2.32***	-1.33**	-0.39	-1.03	-2.41***
<i>Univ.place</i>					
Trento(ref.)					
Povo	1.36**	0.13	0.31	0.01	1.17**
Rovereto	-0.69	-0.81	0.04	-0.45	-0.13
<i>Course Year</i>	-0.49**	-0.16	-0.01	-0.3	-0.38*
<i>Procrastination</i>	-0.20***	-0.08	-0.08	-0.07	-0.12*
Constant	6.16***	2.54	1.96	2.05	4.39**

# Margins: Procrastination syndrom effect by Cluster type



# Take-home message

- Introducing to new modes of data collection inevitably introduces new sources of systematic errors
- Source of biases:
  1. *Mood*: negative moods associated with faster responses
  2. *Individual tendency to procrastinate* lead to delayed answers
  3. *Very fast or delayed answers* lead in reduced accuracy for certain self-reported geographical disclosure
  4. *Modalities of uses* of the app are consistent across time and identifiable in groups of participants and the latter have discriminant characteristics