European University Institute
Robert Schuman Centre for Advanced Studies

Report for the Estonian National Electoral Committee

Internet Voting in Estonia

A Comparative Analysis of Five Elections since 2005

Prof. Alexander H. Trechsel
Director of the European Union Democracy Observatory (EUDO),
Robert Schuman Centre for Advanced Studies,
European University Institute, Florence

and

Kristjan Vassil
PhD researcher at the Department of Political and Social Sciences,
European University Institute, Florence

October 2011

The methodological basis for this study as well as the previous reports on Estonian e-voting have been written in collaboration with Dr. Guido Schwerdt, Ifo Institute for Economic Research, Dr. Fabian Breuer, GPPI Berlin and European University Institute, Prof. Michael Alvarez, California Institute of Technology, Pasadena and Prof. Thad Hall, University of Utah, Salt Lake City. We would also like to thank dr. Anu Realo, Department of Psychology, University of Tartu, for her contribution to the Personality section of this report.

Introduction

This report presents our fourth consecutive analyses on Estonian internet voting since its introduction in 2005. Containing data from the five elections that took place in Estonia since then - 2005 local, 2008 national, 2009 European Parliament, 2009 local and 2011 national elections – this report first looks at overtime trends in internet voting and further scrutinizes the mechanisms that lead voters to decide in favor of internet voting over traditional means of casting a ballot. Furthermore, this report attempts to shed light on a number of new issues that have not been addressed previously and that have gained relevance due to the newly emerging contextual circumstances or innovation within the field of internet voting.

As in the case of the previous studies, this report is based on the joint effort of a research consortium led by the European University Institute's (EUI) "European Union Democracy Observatory" (EUDO). We acknowledge the funding of this study by EUDO and by the Estonian National Electoral Committee (ENEC). Earlier reports led by EUDO were written with researchers from the California Institute of Technology (Caltech), the University of Utah, the Institute for Economic Research in Munich, Germany (IFO) as well as the e-Governance Academy (eGA) in Tallinn. The team also acknowledges the significant support of the initial studies given by the Council of Europe (CoE).

Turnout and e-turnout in five elections

The sixth general election in Estonia since regaining the independence in 1991 was held on March 6, 2011. Out of 913,346 eligible voters, 580,264 decided to participate in these elections amounting to an overall turnout of 63.5 per cent. A somewhat lower number of valid votes were recorded (5,131 votes were invalid).[1]

The number of internet voters rose from 104,313 in the last local elections of 2009 to a staggering 140,846.[2] Internet voters amounted to a 15.4 per cent of the total electorate and to 24.3 per cent of participating voters. Figure 1 compares the overall turnout rates with the growth of e-voters over time.

[1] http://www.vvk.ee/varasemad/?v=rk2011
[2] http://www.vvk.ee/riigikogu-valimised-2011/statistika-2011

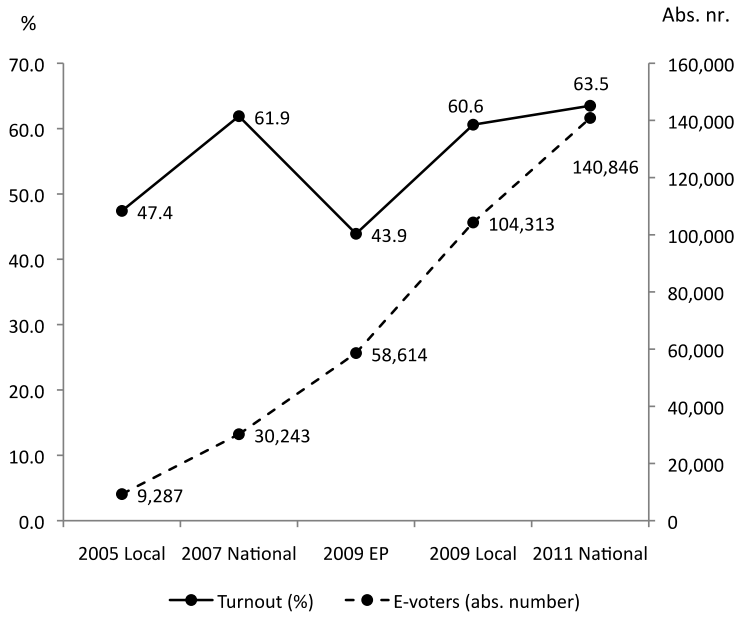


Figure 1. Turnout and the growth of internet voters

As regards the election results, the Reform Party could hold its position of largest party in the Estonian Parliament by obtaining 33 seats (gaining two seats from the last national elections). The Estonian Center Party arrived at 26 seats (losing three seats); Pro Patria and Res Publica Union holds 23 seats in the new Parliament (gaining four seats) and the Social Democratic Party obtained 19 seats, which corresponds to a gain of 9 seats. The Social Democratic Party thus was the largest winner in individual seats as compared to the previous national elections.

1. Questions addressed and design of the study

In this report we follow the basic logic of the previous studies and address the primary question: Why do voters chose to vote over the internet rather than through traditional means? Moreover, by having gathered data on five elections we will look at dynamics over the past six years. In so doing, we achieve an overtime comparison in which we have different types of elections, in different context.

The design of our study is identical to the previous ones that we coordinated. We have conducted five post-elections surveys within the period of one month after each

respective election. Using the method of computer-assisted telephone interviews (CATI) we have collected information from 4,916 respondents, which corresponds to roughly 1,000 respondents per election. The respondents belong to three different groups of equal size: traditional voters voting at the polling place, e-voters voting over the internet and non-voters.

Because the number of e-voters was low during the first e-enabled elections (particularly those of 2005 and 2007) we could not use a random sample. Had we done so, the number of e-voters in the resulting sample would have remained negligible – particularly so in 2005. At the first e-enabled elections, we would have found at best ten internet voters out of a representative sample of 1,000 Estonian residents. Such low figures pose large problems in terms of statistical analyses. Therefore, we decided to boost the number of e-voters we interviewed. We did so by creating these three groups whose differences we were primarily interested in (i.e., traditional voters, e-voters and those who abstained from the elections). We split the samples into three roughly equally large groups. Next, due to their rarity, we had to find a way to obtain a sufficient number of interviews with actual internet voter. The Estonian National Electoral Committee, very kindly provided the polling company "Faktum Ariko" with 1,000 names of e-voters, randomly chosen from the entire pool of e-voters, who did participate in each respective election. The mandated survey company, on the basis of this electoral roll, acquired the phone numbers of the e-voters and contacted them by phone and with the permission of the Estonian National Electoral Committee. The details of each contacted e-voter were later completely anonymized by the survey company.

However, since the number of e-voters rapidly rose from the 2009 elections onward, one could have abandoned this sampling strategy and used a sampling procedure yielding a normal probability sample. Yet, we decided to continue with our initial strategy in order to assure that the gathered time series data are methodologically consistent with each preceding and succeeding wave. Therefore, also for the most recent elections – those of 2011 – we used the electoral roll and quotas for voter types even though we would have found a sufficient proportion of e-voters also in the random probability sample. Table 1 demonstrates the proportion of each voter type in our sample across the five survey waves.

In the next section we turn to our substantial findings. We first deal with the variable we are most interested in (our dependent variable), i.e. type of participation and then identify a set of relevant explanatory variables that will be used in the subsequent analysis.

Table 1. Proportion of voter types in the sample

Voter type	2005 Local	2007 National	2009 EP	2009 Local	2011 National	Total
Normal voter	318	365	278	337	333	1,631
%	33.9	37.3	28.0	33.7	33.1	33.2
E-voter	315	367	400	328	335	1,745
%	33.6	37.5	40.3	32.8	33.3	35.5
Non-voter	305	247	315	334	339	1,540
%	32.5	25.2	31.7	33.4	33.7	31.3
Total	938	979	993	999	1,007	4,916
%	100.0	100.0	100.0	100.0	100.0	100.0

2. Explaining the usage of internet voting

In order to explain usage of internet voting, one has to be explicit about how to operationalize the concept of e-voting. A straightforward way how to look at e-voters is a simple distinction between internet voters and all the others, including the non-voters and traditional voters. This way, however, means that e-voting is effectively compared with traditional voters and those abstaining from elections. Since the profiles of traditional voters and non-voters are, however, markedly different from each other, this way of explaining internet voting does not make much sense. We therefore opt for a second solution by which the reference category only includes traditional voters. Consequently, what we find empirically always reflects the effect that explains internet voting as compared to voting by conventional means (i.e., going to the ballot station)[3].

Explanatory variables

In order to explain why some voters use the internet for casting their vote, rather than going to the polling places, we need to specify a statistical model that contains a number

[3] The main dependent variable used in this section is thus limited to internet voters (taking the value of 1) and traditional voters (taking the value of 0).

of explanatory variables. We deliberately chose to keep the number of factors explaining internet voting at a minimum. The parsimonious model is thus a bit different from the rather larger models, in terms of size, presented in earlier reports. Having learnt a lot from our research on internet voting in Estonia, we now have a clearer picture and can move from an exploratory stage to a more analytical one. We retain six demographic characteristics: age (dichotomized into four groups), gender (coded 1 for males and 0 for females), place of residence (1 for urban, 0 for rural), education (split in three dummy variables "secondary education" and "higher education", with elementary education being the reference category), income (coded in 19 categories), and language (1 for Estonian, 0 for other languages).

Next, we include two variables measuring respondent's engagement with politics. First, we include *political discussions* – a variable measuring how often somebody discusses politics with family or friends.[4] Second, we use *political activity* as a categorical self-assessed measure of past electoral turnout.[5] Finally, we include *trust toward the system of e-voting* (four categories[6]) and computer literacy (five categories[7]) as potential predictors of internet voting.

In specifying a concise model of internet voting, our goal is to demonstrate that we can indeed, with a limited number of variables, explain a large amount of the variance in our dependent variable, i.e., that we can present a clear picture of why people chose one or the other form of voting. Therefore, we refer to this model as our baseline model of internet voting. However, other variables may also be of great interest, both academically and from the policy related point of view. Therefore, we look at further evidence in a second step, after presenting the findings from our baseline model.

[4] The survey question reads as follows: When You are among Your family or friends, do You always discuss politics, discuss politics from time to time, rarely, or never? Response categories include: often, sometimes, rarely, never.

[5] Have you participated in all the elections (category 1), when You have had the right to vote, some elections (category 2), sometimes (category 3) or never (category 4).

[6] How much do You trust the procedure of internet voting? Response categories: totally trust, rather trust, rather do not trust, do not trust at all.

[7] How do You evaluate Your computer skills? Are they: very good, good, average, rather poor, very poor.

Findings

Table 2 reports the findings. Recall that our dependent variable is coded 1 for e-voters and 0 for traditional voters. We have estimated five different models for each respective election in order to compare how the performance of each characteristic varies over time. Note that Table 2 reports so called average marginal effects[8]. To give an example of how to interpret the latter, let's look at the average marginal effect of 28.9, corresponding to the effect of *Estonian language* in 2005. This means that when moving language from its minimum value (0) to its maximum value (1)[9], then the probability of internet voting increases by 28.9 per cents (while keeping all other variables at their means). In other words, in 2005 internet voting was more probable for those speaking Estonian than other languages. However, as the marginal effect captures the effect across the entire scale of the particular characteristic being measured (e.g., in the case on trust toward the e-voting system or PC literacy that contain more categories) it is often difficult to infer at which particular value the probability of internet voting increases substantially. Therefore, we have also plotted so called "predicted probabilities" for some of the key variables of interests in order to show these thresholds. We will explain how to interpret these figures below.

Age

We begin by interpreting the effects of seven demographic variables on the probability of e-voting. Age is often referred to as the most prominent predictor of internet voting. The main argument is that because the young are exposed to the information and communication technologies (ICTs) to a far greater extent than the elderly, they also have the necessary skills to use the internet for voting. Subsequently, internet voting is expected to attract mostly the young. According to our findings in Table 2 this expectation is not entirely correct. If we compare the effects of the three age groups (36-55; 56-65 and 65-95) to the reference category (18-35) we see that in only some of the elections and only some of the age groups come out as statistically significant. This

[8] Average marginal effects show the average of the variation induced in the probability of interest by a marginal change in an independent variable for each individual in the sample.

[9] This move effectively means that the dummy switches from other language, most notably Russian (coded 0) to Estonian (coded 1).

heterogeneity points to the fact that age is not an extremely strong predictor of internet voting when controlling for other characteristics as well. For example, if we estimated a model with demographic variables only (refer to Appendix 1), age came out as an important and a systematic predictor of internet voting providing support for the age-related argument explicated above. In such a non-controlled case e-voting appeared to be indeed a feature of the young voters. However, as we included other measures (so called "controls"), and most notably "language" and "trust toward internet voting", the apparent effect of age became less marked.

What does this mean in practical terms? Does age predict internet voting (as the popular expectation would prescribe) or not? We argue that age is an intrinsic part of the baseline model of internet voting and it predicts internet voting fairly strongly. However, its effect is suppressed by other factors that explain internet voting more effectively. Because of age does have some kind of impact, albeit a weak one, we keep this variable in mind when answering our main question of why people vote over the internet or not. However, it is important to underline that while age may function as a factor impacting on the probability of voting over the internet, it is not the root cause of engaging in internet voting.

Table 2. Explaining internet voting

	2005 Local	2007 National	2009 EP	2009 Local	2011 National
Age group (36-55)	5.19	14.35***	1.36	-5.60	-1.63
Base: age group 18-35	(5.25)	(4.55)	(6.01)	(5.16)	(5.35)
Age group (56-65)	7.58	16.48***	14.55**	-6.59	-12.57**
Base: age group 18-35	(7.18)	(5.10)	(6.99)	(6.49)	(6.10)
Age group (65-96)	5.52	9.07	-8.76	-11.89	-13.56*
Base: age group 18-35	(7.30)	(8.93)	(8.43)	(7.61)	(7.24)
Male	-5.13	-1.65	6.15	-2.46	12.54***
Base: female	(4.18)	(3.97)	(4.44)	(4.30)	(4.23)
Urban	-3.40	6.18	-6.54	-5.52	-4.37
	(4.49)	(4.08)	(4.36)	(4.35)	(4.34)
Secondary education	2.74	4.30	2.47	-8.41	-22.15**
Base: elementary	(10.14)	(11.30)	(9.87)	(15.45)	(10.75)
Higher education	10.58	10.88	15.91	7.19	-3.02
Base: elementary	(10.48)	(12.68)	(11.12)	(15.76)	(9.79)
Income	-5.41	15.71*	17.62	19.95*	51.74***
	(8.63)	(8.38)	(11.16)	(11.35)	(15.75)
Estonian language	28.93***	28.72***	31.29***	22.17***	-0.13
Base: other languages	(5.28)	(4.14)	(5.86)	(5.60)	(6.65)
Political discussions	-17.45***	6.15	-11.12	4.58	-12.16
	(6.75)	(8.06)	(7.67)	(7.29)	(7.44)
Political activity	-4.01	-20.75*	16.20	-15.80	-3.49
	(11.11)	(11.86)	(13.35)	(12.09)	(11.37)
Trust toward e-voting	82.54***	66.82***	67.06***	69.09***	43.67***
	(6.73)	(8.51)	(7.25)	(8.86)	(5.77)
PC literacy	34.99***	37.57***	24.62**	24.62**	10.73
	(9.71)	(8.64)	(9.83)	(9.73)	(8.87)
Constant	-7.39***	-8.44***	-7.78***	-5.90***	-2.19*
	(1.12)	(1.29)	(1.16)	(1.15)	(1.10)
Observations	515	619	553	483	482
Pseudo R2	0.32	0.24	0.29	0.29	0.33
Correctly classified	59.5	48.16	59.94	56.32	57.09
Log Likelihood	-171.3	-236.6	-189.4	-165.9	-161.1

[1] Dependent variable is a dichotomy between internet voters (1) and traditional voters (0)

[2] Model reports average marginal effects. Reading guide: An effect of 28.9 (in the case of *Estonian language* in 2005) means that when moving language from its minimum value (0) to its maximum value (1), the probability of internet voting increases by 28.9 per cents

[3] Standard errors in parentheses

[4] *** p<0.01, ** p<0.05, * p<0.1

Gender

With respect to gender we witness that only in the last elections of 2011 internet voting is about 12.5 per cent more probable among men than among women. Other than in this exceptional case, gender does not exert a significant effect for the explanation of e-voting. This finding may seems surprising, because according to the official statistics there were 54 per cent females participating in the last elections of 2011. If so, then how can one reconcile our findings (were males are more likely to vote online) with those of the official aggregate statistics?

First, it is important to understand that in the Estonian population women form the majority, with about 54 per cent of the total Estonian population.[10] On the aggregate more females also vote online, but within the female population the proportion of e-voters (as compared to other types of voters) may be lower than within the male population. Therefore, irrespective of whether we witness a higher aggregate percentage of women among the e-voters, the effect may still be positive for men, because of a higher number of e-voters among the respective population. Moreover, the gender effect may come out due to the fact that our model controls for a number of other characteristics. This, of course, differs from the descriptive official statistics. The latter offer a description showing that there is no particular gender bias in e-voting. Our results, in return, demonstrate that gender has become a moderate predictor of e-voting. It is difficult to say anything more substantive about the differences in gender and voting – the 2011 elections were so far the only ones in which gender had a limited impact and we refrain from generalizing this effect or to predict any future effects.

Place of residence, education and income

With respect to place of residence we see no significant results. It suggests that the urban-rural cleavage does not play any role in one's probability to use internet voting. Much in the same pattern, education is not significantly related to internet voting apart from the last elections, in which we observe the negative relationship between voting online and those who have elementary education. Note however that in the simple model (restricted to socio-demographic variables - refer to Appendix 1) both place of residence and

[10] Statistics Estonia: http://www.stat.ee/

education come out as significant elements of the analysis. In particular, internet voting is more probable among those with higher levels of education. Again, we interpret these findings in light of our baseline model of internet voting: while both place of residence and education are part of the baseline model, their effects seem to be overtaken by other attitudinal and behavioral characteristics once we control for the latter.

Income has been a moderate predictor of internet voting over the years but it has gained explanatory power in the last elections in 2011. It appears that voting online is 52 per cent more probable for the group of voters earning the highest income as opposed to those earning the lowest income. Again, whether this constitutes an increasing cleavage is at the moment difficult to predict and we refrain from more general statements – in particular in view of future elections - regarding this factor.

Language

In our last report we cited language as one of the most important predictors of internet voting possibly hinting towards a structural self-exclusion of the Russian-speaking minority from this mode of voting. We argued that because the internet voting system itself is offered only in Estonian, it might deprive large parts of Russian speaking minority from using the system.

It is for these reasons that the recent findings on language are of utmost importance. In particular, we have found that language as a predictor of internet voting has entirely lost its statistical significance in the recent 2011 elections. Although with some empirical caution in mind, we infer from this finding that the important language cleavage has evaporated (or at least shows a clear tendency to disappear) due to the large number of users, who also seem to be more heterogeneous language wise. We cannot empirically link this finding to any particular cause, but we are inclined to think that it may have its roots also in the increasing attention that the Estonian National Electoral Committee has given to the inclusion of the Russian speaking minority in the voting process. For example, the website of www.valimised.ee, which contains instructions how to vote online was translated into Russian for the first time for the 2007 election. Advertisements in Russian language, help pages etc. may have shown their effects. This,

in interaction with the growing number of e-voters provides us with reasonable confidence that language may gradually disappear as a characteristic of internet voting.

Political attitudes and behavior

Finally, coming to the characteristics that measure one's political attitudes and behavior, our data show that political discussions and political activity do not predict internet voting (apart from the first elections of 2005 in the case of political discussions and in 2007 in the case of political activity).

At the same time, however, trust toward the system of internet voting has remained the single most important predictor of e-voting. The overtime consistency of this finding shows that trust is by far the most important and most valuable feature that needs to be retained – and fostered - if e-voting is expected to proliferate. Table 2 shows that its effect has slightly diminished over the years, but rather than interpreting this finding as a decreasing level of trust we are confident that it happens because of the higher heterogeneity among the respondents. Namely, as the number of internet voters increases those who trust it *completely*[11] is also diminishing and the most frequently reported levels of trust slowly approach the mean of the trust scale. Therefore, we are confident that trust remains on top of the characteristics that are associated with internet voting and that it only slightly weakens over time due to the increasing number of e-voters. In order to provide a better grip on the performance of trust as the most important characteristic of e-voters, we have also plotted predicted probabilities of internet voting over the continuous trust variable. Results are shown in Figure 2.

[11] I.e., those who express levels of trust toward the e-voting at either of the extremes (completely trust, completely distrust).

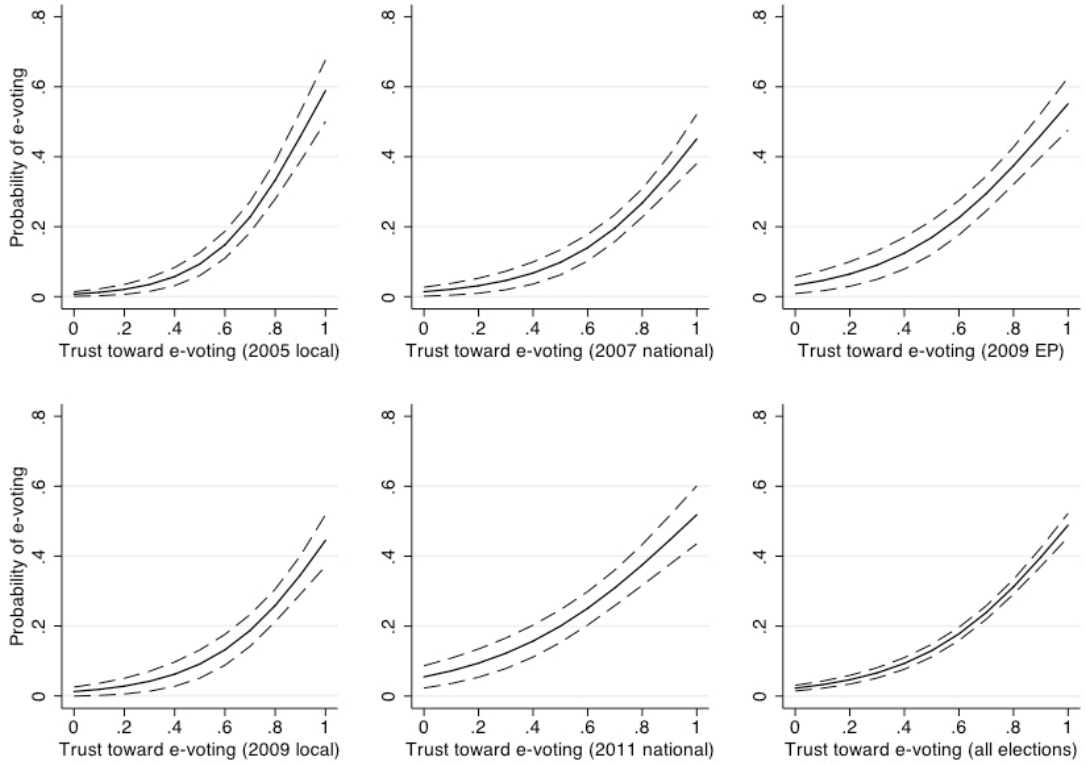


Figure 2. The probability of e-voting over the trust toward the system of e-voting

We conclude by rendering caution with respect to trust related issues. As we have seen in 2011[12] simple attempts to destabilize the system of e-voting and the questioning of its level of security may have a profound impact on the degree to which voters trust the system. With respect to general trust we see no particular changes that could be attributed to this cause. However, should one loose trust in the procedure of e-voting, it is unlikely that this person will fully restore this trust to the previous level.

Finally, we come to PC literacy as a variable positively and significantly associated with e-voting. Much alike the dynamics of the language variable, the effect of PC literacy has remained important throughout the first elections and disappeared in the last ones. These dynamics are graphically shown in Figure 3. We gather from this finding that, over time, other characteristics than technical skills have become more strongly associated with the likelihood of e-voting. This finding points to a potentially diminishing

[12] We refer to the Paavo Pihelgas case.

gap between the tech savvy internet voters and those who are less skilled in using new technologies. We find that this trend is important as it shows evidence of decreasing barriers that deprive the less skilled from using the internet to cast their vote. Clearly, it does not only mean that technical barriers no longer exist. Primarily it shows that in an ever-increasing population of internet voters those with lower levels of PC literacy find their ways to use internet voting, and thereby bringing more heterogeneity into the population. Subsequently, it also points to a more balanced distribution of technically more and less skilled citizens among the e-voters.

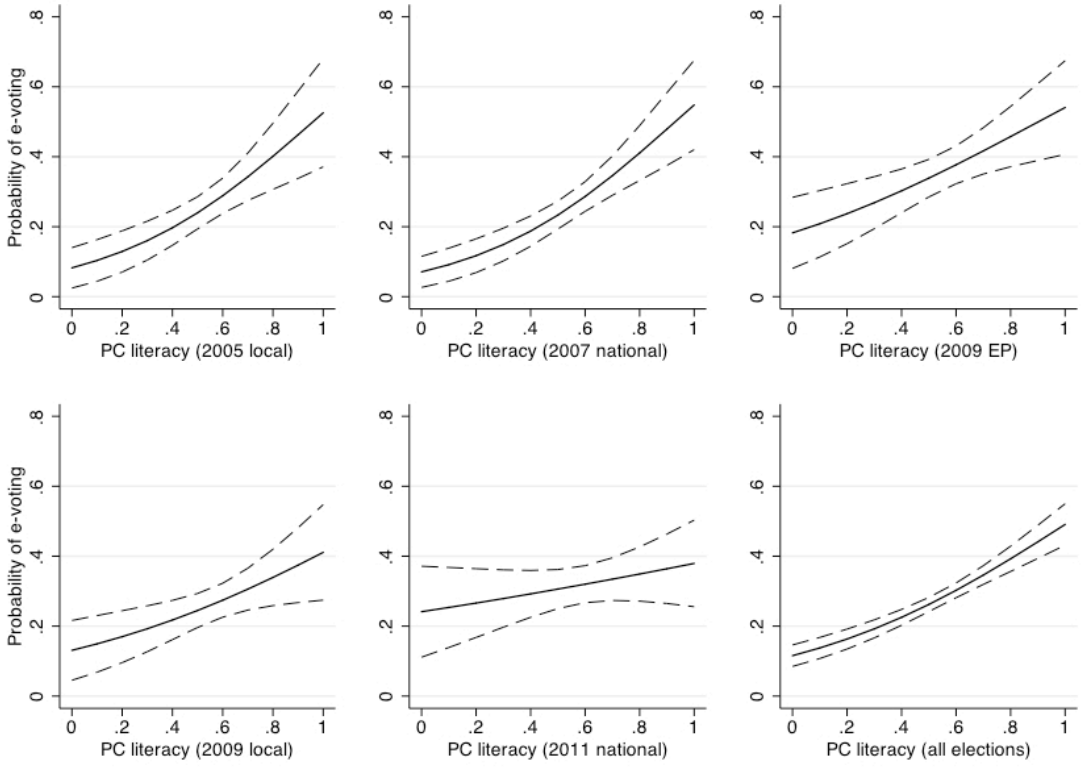


Figure 3. The probability of e-voting over the computer literacy

Concluding remarks on the main findings

When looking at the over time trends in internet voting it is apparent that the performance of some individual level characteristics show mixed patterns. In statistical jargon we witness high levels of heterogeneity. We find that these variables fluctuate with regard to magnitude of the effect (either positively or negatively associated with internet voting) or their statistical significance (i.e., uncertainty related to the effect). Such characteristics

are most notably age, gender, income and political discussions. We gather from this that these variables, even if they play an important role on their own (i.e., in the model reported in Appendix 1), when controlling for attitudinal and behavioral covariates they tend to become less stable and convincing in predicting e-voting. Does it mean that these fluctuating variables should be omitted from the analysis and more focus should be given to those that are stable over time? We believe it is better to keep them included as socio-demographic characteristics are considerably more stable when *not* controlling for attitudes and behavior (see Appendix 1). This finding gives us confidence that internet voting can be conceptually and empirically modeled as a two step process in which basic socio-demographics determine the baseline propensity to vote online. In this first step, indeed, age and education are important predictors of e-voting. However, when moving toward the second step, in which attitudes and behavior are included we see that the latter group of variables suppresses the effect of the socio-demographic characteristics. Does it mean that socio-demographics are no longer explaining internet voting? Quite to the contrary, we argue that the configuration of certain socio-demographic traits provides a basis by which some voters are more likely to vote online and others not. However, this potential is only released (i.e., they actually vote online) if a person has a certain attitudinal and behavioral profile (most notably trust in the system of internet voting).

This brings us to the key elements that explain the patterns of internet voting. First, it appears that trust toward the system of internet voting remains an important predictor of e-voting and consistently so over the years. Second, PC literacy and Estonian language have been substantial explanatory characteristics, but both of them have lost their explanatory power in the last 2011 election. What remains central is trust in internet voting.

This concludes the main part of our analysis. However, as we have mentioned above, we will address a few new issues that have emerged between the time of the last report that we published in 2010 and the current one. We will address these in the following sections.

3. Time period of e-voting

That internet voting is allowed over several days during the advance voting is an important convenience factor for voters. The longer it is, the more likely it is that more of those voters who would not have participated in elections come across e-voting and subsequently vote. Initially, in 2005 and 2007, the period of e-voting was three days. Since the 2009 European Parliament elections the period was extended to seven days. In our last report we wrote (Trechsel, et. al 2010, p. 64[13]):

> The added value of voting over the Internet is, amongst others, the freedom of choice with regard to the moment of voting. Note, that in all our reports *convenience* has been the single most important reason for the choice of e-voting. We were glad to learn that in both elections in 2009 the voting period was extended to seven days. As we report in our current analysis, the length of the voting period is potentially an important determinant of voting activity. In light of this event, we suggest to keep the e- voting period at seven days or even extend it (note that, for example in Switzerland, remote voting is open up to three weeks prior to Election Day).

Recent debates in the media and among policy makers have raised questions of whether the period of internet voting needs to be shortened to three days again. The reason for this is the fact that different voting channels have different durations. Advanced voting can be done during three days and traditional voting at the polling place can only be done on one single day. Some people argue that internet voting, open for seven days, is creating an inequality. We are opposing the shortening of the voting period, because in such a way any advance voting technique can be questioned, most notably postal voting. Recall, that the very reason why advance voting is introduced is to provide additional means to those voters who cannot vote on the Election Day. In other words, as soon as one allows for several forms of voting, these forms will not be equal to each other. And neither do they have to be equal. Finally, we do not believe that allowing voters to gain flexibility in the casting of their vote, to make them used to this flexibility only to take it away again,

[13] http://www.vvk.ee/public/dok/Report_-_E-voting_in_Estonia_2005-2009.pdf

without a sufficiently grounded reason, may have devastating effects on trust in the elections overall, the work of the authorities etc.

However, these are all normative questions, which we cannot address empirically at this stage. However, we have at our disposal aggregate level data about voting frequencies over days, and individual level data from our surveys. Combining the two allows us to provide a rough estimate of what would be the cost of shortening the voting period in terms of potentially lost votes.

Losing votes – an empirical approximation

In order to empirically approximate the number of votes that would be lost due to the shortened voting period we first need to consider the number of e-voters by days (for this exercise we employ data from the recent 2011 election).

According to official statistics 145,230 votes were cast online out of which 4,384 were repeated ones.[14] That is, the total number of registered e-votes was 140,846. That is, three per cent of e-voters voted more than once. We subtract this percentage from official record that shows voting activity by days, so that it corresponds to the actual total number of e-votes that was registered. The resulting e-voting activity is reported in Figure 6.

[14] Recall, that in Estonia an e-voter is allowed to vote repeatedly so that the last vote will dismiss the previous ones and counted as cast.

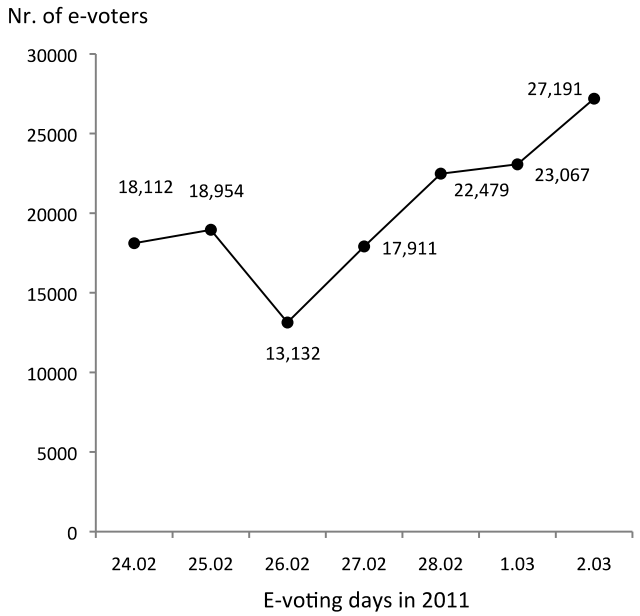


Figure 6. Number of e-voters by days

Next, we calculate a proportion of voters who did cast their vote in the first four days (February 24-27) during the e-voting period, which is the period that is at stake and that certain political forces whish to cut. In those four days 68,109 e-votes were cast (48.4 per cent of total e-votes, i.e. almost half of all internet voters!).

We further turn to our survey evidence and identify the number of those voters who reported that they would most certainly or rather certainly not have voted had there been no internet voting available to them. In our 2011 survey, we found 14.6 per cent of internet voters belonging to this category.[15] Next, assuming that (1) 14.6 per cent is a reasonable approximation of the proportion of e-voters who would not have voted without internet voting at the e-voters' population level and (2) that it is equally distributed across seven days during which e-voting was available, we can calculate the potentially lost e-votes in the event of shortening the voting period from seven days to three.[16]

[15] This proportion remains similar across all of our surveys.
[16] Note that we are reluctant to extrapolate these numbers on the entire population of voters, as e-voters differ both attitudinally and behaviorally from the total electorate.

In order to do so, we calculate the 14.6 per cent out of 68,109, which is 9,943 votes. Then we subtract these potentially lost votes from the total number of e-voters (140,846) and arrive at 130,903. In percentages this yields a potential loss of seven percent of total e-votes.

Clearly, our approximation is based on several assumptions that may not entirely hold in reality. For example, in the event of shortening the period to three days, e-voters may in fact, readjust their habit and vote more intensively over the three days. However, it is likely that some of them will also abstain, simply because the voting period is shorter and the convenience is reduced. After all it , might not only be out of chance that half of all internet voters cast their vote over the first four days of the seven-day period. Our empirical approximation of the likely consequences of going back to a three-day voting period shows the loss of around 10,000 e-votes – a consequence that one should, in our view, avoid and not facilitate.

4. Proximity to the polling station

That participating in elections can be costly also in practical terms, i.e., time lost for getting to and from the polling station, is a well-known fact. Therefore, one of the great benefits of internet based voting certainly is to allow voters to spend less time on the actual act of voting thereby bringing also those to the ballot who otherwise would have abstained, due to their lack of time or unwillingness to spend the time necessary to vote at the polling place. In order to examine whether or not e-voting is actually perceived as a means to save time and thereby provide increasing convenience for electoral participation we employ a survey question that measures time that a respondent needs for getting to and from the polling station.[17]

Since this question was asked only since 2007, we pool the data from four waves and specify a model in exactly the same fashion as we did in Table 2. However, to avoid excessive repetition we do not report the whole model, but instead we plot the probability of internet voting (for each respondent in our data) over the time variable while holding all other covariates at their mean values. Furthermore, we constrain the time not to exceed 125 minutes, because the effect stays close to its maximum throughout the values above two hours. The findings are shown in Figure 5.

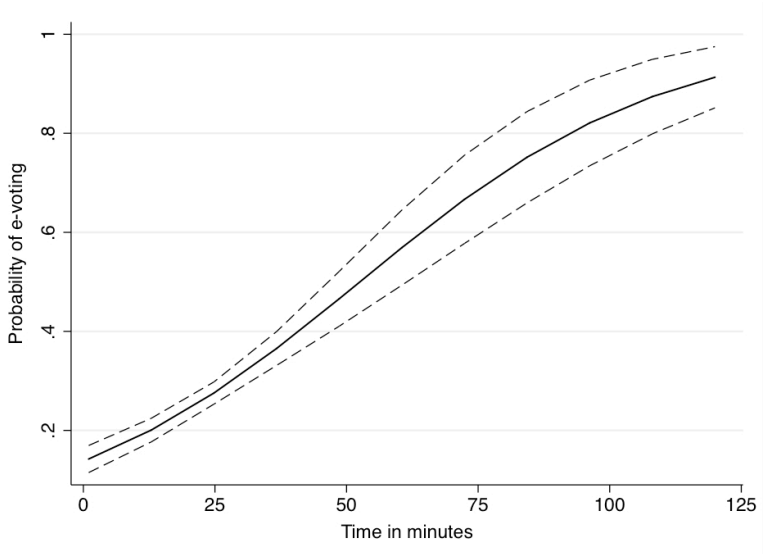


Figure 5. Probability of e-voting and the distance to the polling station measured in minutes (2007/2011 pooled data)

[17] How long did or do you think it would have taken you to go from your home to your polling place, cast your vote and get back? Answers were given in minutes.

The results show that the probability of internet voting reaches beyond the 0.5 probability threshold (which is usually considered the threshold at which an event becomes probable) at around 60 minutes. It means, that *ceteris paribus* for those respondents for whom going to and coming from a polling station takes more than one hour, it is highly probable that they would opt for e-voting. Note that this effect prevails while controlling for other covariates. We infer from this finding that e-voting as a voting channel of convenience should not be underestimated. E-voting clearly lowers the costs of participating in elections and this time-saving effect becomes highly significant the higher this cost is for an individual voter.

This finding echoes our concerns that we explicated in the previous section, too – the time period at which e-voting is available during the advance elections. Both findings highlight the fact that voters perceive e-voting as a matter of convenience. We believe that this convenience should not be disturbed by hectic reforms that change the system to which voters have become increasingly habituated (our previous report showed that if people continue voting in consecutive elections and if they previously have done so over the internet, they stick to e-voting). Failure to provide stability may drastically decrease voters' trust in the system.

5. **Personality traits and Internet voting** (co-authored with dr. Anu Realo[18])

In our previous reports we have always dedicated special attention to certain domains within the realm of internet voting that is not directly linked to our core research questions on e-participation and impact on turnout. Most prominently we have looked at campaign effects and media consumption. In this report we further advance the scope of our research interests and introduce a new interdisciplinary approach to online voting behavior. Namely, we included a battery of ten questions to our 2011 survey that measure five latent dimensions of personality – broadly known as the "Big-Five personality traits". In the following we explain why one should be interested in personality traits when investigating online voting behavior and describe the logic of measurement and the key findings.

Expectations

The likelihood to e-vote is conditioned by a certain set of socio-demographics together with attitudinal characteristics. At the same time, participation in remote online voting involves a number of aspects that are not directly present in traditional voting. For example, the counting of e-votes is not directly observable and it is less straightforward for any ordinary citizen than counting of the traditional votes. Similarly, by casting a conventional paper based ballot a voter is certain that her vote has reached at least the ballot box, which again cannot be directly observed in the event of e-voting. We argue that irrespective of whether these threats are relevant in reality, with respect to online voting a voter relies on trust toward the system of internet voting that may, to a great extent, be mediated by one's openness to new experiences and willingness to try new things for reasons of curiosity.

We theorize that additionally to the traditional predictors of internet voting a certain configuration of one's personality traits may provide further insights into the

[18] Anu Realo is a Senior Research Fellow of Personality Psychology at the Department of Psychology, University of Tartu, Estonia. She received her *PhD* degree in psychology from the University of Tartu in 1999. From 2000 to 2002 she worked as a postdoctoral research fellow at the University of Leuven, Belgium and at the Swedish Collegium for Advanced Study in Social Sciences, University of Uppsala, Sweden.

nature of a typical e-voter. Several researchers have proposed that political attitudes and voting behavior depend to a large extent on people's basic personality traits,[19] [20] which have been defined as enduring tendencies to think, feel, and behave in consistent ways.[21]

Over the past five decades, personality researchers have reached a consensus that personality traits can be best organized by five broad factors, which „represent the most basic dimensions underlying the traits identified in both natural languages and psychological questionnaires".[22] In the following we briefly present the basic characteristics of these five dimensions.

Neuroticism is a general tendency to experience negative emotions such as sadness, anger, and fear. People who score high on Neuroticism are prone to psychological distress whereas people scoring low on Neuroticism are usually emotionally stable and well-adjusted. *Extraversion*, on the other hand, is a tendency to experience joy and other positive emotions, to seek out stimulation, to be active and sociable. People who score high on *Openness to Experience* are usually intellectually curious, attentive to their inner feelings, open to new ideas and opportunities. On the other hand, people scoring low on Openness tend to be conservative and conventional, preferring the "old way" of doing things rather than the "new". *Agreeableness* is mostly about inter-personal tendencies referring to people's altruistic, sympathetic, cooperative, and trustful nature. Finally, *Conscientiousness* is a tendency to manage self-control, to be effective, purposeful, and hard-working.[22]

We argue that by employing a Big-Five personality instrument we gain a powerful predictor of internet voting as a distinct model from the traditional socio-demographic one. In particular, we expect *Openness to Experience* to be a strong predictor of e-voting. In the following we empirically test this hypothetical link.

---

[19] Caprara, G. V., & Zimbardo, P. G. (2004). Personalizing politics: A congruency model of political preference. *American Psychologist,* 59, 581–594.

[20] Gosling, Samuel D., Peter J. Rentfrow and William B. Swann, Jr. (2003). A very brief measure of the Big Five personality domains. *Journal of Research in Personality, 37*, 504-528.

[21] Allik, J., & McCrae, R. R. (2002) A Five-Factor Theory perspective. In R. R. McCrae ja J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 303-321). New York: Kluwer Academic/Plenum Publishers.

[22] Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, Fl: Psychological Assessment Resources.

Measuring personality traits

We used a Ten-Item Personality Inventory (TIPI)[20] to measure the Big Five personality dimensions – Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Each item consists of two descriptors, separated by a comma, using the common stem (e.g., "I see myself as extraverted, enthusiastic"). The items were rated on a 5-point scale (1 – *disagree strongly* to 5 – *agree strongly*). Each scale is constructed by calculating the mean of the two items per dimension.[23]

In order to estimate the effect of personality traits on the probability of e-voting we use the same dependent variable as before. It is coded 1 if the respondent voted online in the 2011 national elections and 0 if she voted by conventional means (i.e., during advance voting or on Election Day at the ballot station). Those abstaining from elections are omitted from the analysis. Independent variables are constructed as scales (ranging from one to five) using the operationalization provided by Gosling et al.[20]

Findings

Table 3 reports the results from the normal model (logit model) where the five trait variables are used as predictors of e-voting. As in the previous report, we report average marginal effects (refer to the interpretation guide above). Notice, that our model specification also includes age, as personality traits are sensitive to one's age. E.g., young people tend to be more open to new experiences than the elderly. Therefore, by controlling for age effects we also maintain that the effects that we find with regard to personality traits are not spurious ones.

First, consider the goodness of the fit of the model. Together with age, personality traits achieve a Pseudo R^2 of 0.18. Clearly, the largest proportion of this 'variance explained' measure is provided by age. When we estimated a model without age, we saw a Pseudo R^2 of 0.05 (refer to Appendix 2). This points to the fact that even if personality traits *per se* do not explain a lot of variance, they still substantially contribute to the baseline model of internet voting.

[23] The Estonian version of the TIPI has been previously used by Kalmus, Realo, and Siibak (2011) to measure the motives for Internet use and their relationships with personality traits among the general Estonian population.

As regards the substantial findings, Table 3 confirms our main expectation. Openness to new experiences appears to be indeed positively associated with e-voting. More specifically, we see that when moving the 'openness' variable from its minimum value (1) to maximum (5), the chances of e-voting increase by about 7.2 per cent while holding all other variables at their means (including age).

Table 3. Personality traits and internet voting

Variables	Model 1
Age	-1.14***
	(0.08)
Extraversion	-0.88
	(2.08)
Agreeableness	-2.15
	(1.49)
Conscientiousness	-5.79**
	(2.85)
Emotional stability	2.11
	(1.96)
Openness to experiences	7.21***
	(2.17)
Constant	3.26***
	(0.83)
Observations	667
Pseudo R2	0.18
Log Likelihood	-380.02

[1] Dependent variable is a dichotomy between internet voters (1) and traditional voters (0)
[2] Model reports average marginal effects. Reading guide: An effect of 7.2 (in the case of *Openness to experiences*) means that when moving this scale from its minimum value (1) to its maximum value (5), the probability of internet voting increases by 7.2 per cents
[3] Standard errors in parentheses
[4] *** p<0.01, ** p<0.05, * p<0.1

However, since the personality items are constructed as scales and as the marginal effect captures the effect across the entire scale of the particular characteristic being measured it is difficult to infer at which particular value the probability of internet voting goes beyond 50 percent, the traditional threshold. Therefore, we have also plotted predicted probabilities for the 'Openness to experience' variable in order to show these thresholds. Figure 6 presents these probabilities graphically. Notice, that with regard to openness the chances of e-voting substantially increase (exceed the threshold of 0.5 on Y axis) at around the mean value of the scale.

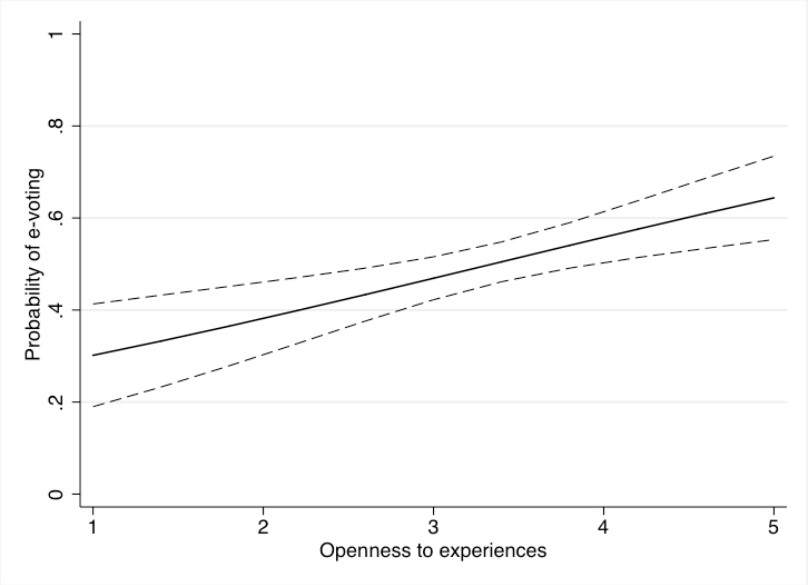
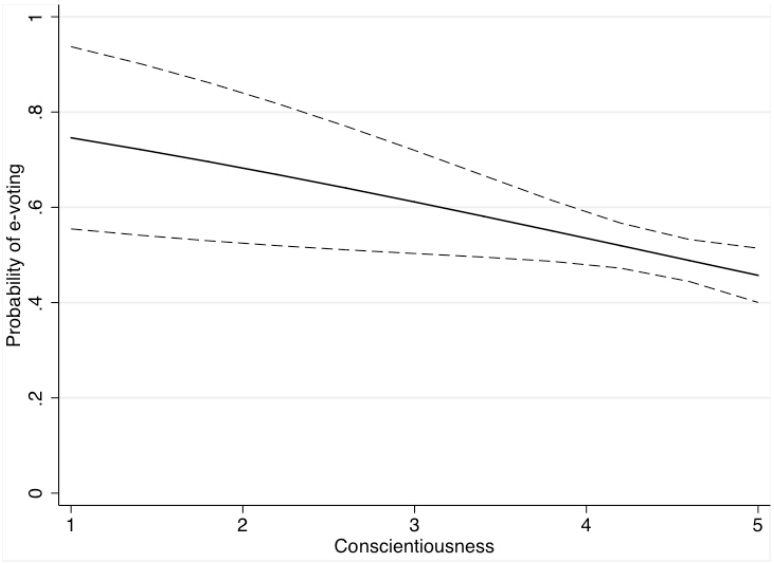


Figure 6. Personality traits and the probability of internet usage

A somewhat puzzling effect appears with regard to *Conscientiousness*. We find that those who score high on the *Conscientiousness* scale are 5.8 per cent less likely to use e-voting as opposed to those who score low on the same scale. We suggest that those who are effective and maintain strong self-control (as this personality trait prescribes) may, at the same time, be likely to avoid mistakes and tend to be cautious toward the new and the unknown. If so, people who score high on *Conscientiousness* are likely to avoid risks and prefer means of voting that they have already used before. Consequently, the probability of internet usage decreases as one's *Conscientiousness* increases. Figure 7 displays the probability of internet usage by *Conscientiousness*.



Taken together, these findings provide us with sufficient evidence to conclude that those using e-voting are, from the personality point of view, intellectually curious, open to new ideas and opportunities. At the same time, those who maintain self-discipline and tend to be well organized are more likely to refrain from e-voting, because they also likely to avoid mistakes and prefer means of voting that they are experienced with.

One can theorize that if these traits go together with political engagement, one's probability to use internet voting increases. On the other hand, the inverse interpretation of this finding is that people who are conservative and conventional in their behavior are less likely to participate in e-voting. Clearly, this finding is not surprising in its own right. However, this is the first empirical evidence of personality related mechanism that affects the probability of internet voting.

Appendix 1

Table N. Socio-demographic model of internet voting

	2005 Local	2007 National	2009 EP	2009 Local	2011 National
Age group (36-55)	-0.83	6.00	-3.36	-9.40**	-8.08**
Base: age group 18-35	(3.98)	(3.76)	(5.37)	(3.75)	(4.05)
Age group (56-65)	-10.34**	-3.26	-5.02	-16.40***	-21.87***
Base: age group 18-35	(4.20)	(4.07)	(5.79)	(3.57)	(3.78)
Age group (65-96)	-18.30***	-15.84***	-28.81***	-27.49***	-32.10***
Base: age group 18-35	(3.61)	(3.54)	(4.56)	(2.96)	(3.65)
Male	1.08	2.84	9.26**	-0.68	11.15***
Base: female	(3.25)	(3.09)	(3.89)	(3.19)	(3.44)
Urban	1.58	4.65	-5.32	-5.24	-1.10
	(3.45)	(3.17)	(4.06)	(3.36)	(3.25)
Secondary education	15.85**	7.31	16.41**	4.25	-8.45
Base: elementary	(6.30)	(7.08)	(7.10)	(7.45)	(6.36)
Higher education	28.26***	18.22**	35.69***	21.06**	11.08*
Base: elementary	(7.11)	(8.30)	(8.48)	(8.35)	(6.67)
Income	8.56	29.29***	30.70***	28.36***	45.27***
	(6.49)	(6.69)	(9.61)	(8.93)	(13.25)
Estonian language	23.04***	24.55***	30.89***	19.16***	9.68**
Base: other languages	(2.63)	(2.35)	(3.64)	(3.42)	(4.68)
Constant	-4.57***	-5.34***	-4.63***	-3.10***	-1.57**
	(0.66)	(0.70)	(0.76)	(0.65)	(0.57)
Observations	609	730	628	589	597
Pseudo R2	0.134	0.141	0.222	0.227	0.282
Log Likelihood	-251.2	-309.9	-238.7	-212.1	-206.2

[1] Dependent variable is a dichotomy between internet voters (1) and traditional voters (0)

[2] Model reports average marginal effects. Reading guide: An effect of 23 (in the case of *Estonian language* in 2005) means that when moving language from its minimum value (0) to its maximum value (1), then the probability of internet voting increases by 23 per cents

[3] Standard errors in parentheses

[4] *** p<0.01, ** p<0.05, * p<0.1

Appendix 2

Table N. Personality traits and internet voting (only personality traits)

Variables	Model 1
Extraversion	3.38
	(2.22)
Agreeableness	-2.36
	(1.65)
Conscientiousness	-9.22***
	(3.07)
Emotional stability	1.55
	(2.15)
Openness to experiences	12.88***
	(2.25)
Constant	0.52
	(0.66)
Observations	667
Pseudo R2	0.05
Log Likelihood	-437.5

[1] Dependent variable is a dichotomy between internet voters (1) and traditional voters (0)
[2] Model reports average marginal effects. Reading guide: An effect of 7.2 (in the case of *Openness to experiences*) means that when moving this scale from its minimum value (0) to its maximum value (5), the probability of internet voting increases by 7.2 per cents
[3] Standard errors in parentheses
[4] *** p<0.01, ** p<0.05, * p<0.1