Article

Teaching Public Administration
1–23

DOI: 10.1177/0144739416650431
tpa.sagepub.com

Assessment of effectiveness of public integrity training workshops for civil servants – A case study

Katalin Pallai
National University for Public Service, Hungary

Aniko Gregor
Eotvos Lorand University (ELTE), Hungary

Abstract

The general practice of civil servant training providers in Hungary is to evaluate their products only through surveying the reaction of participants. The obvious weakness of this practice is that the variance in the level of satisfaction does not necessarily coincide with the positive professional impact that the trainings are aimed to produce. This paper presents the results of an effectiveness assessment survey of a large public ethics and integrity training programme that was delivered to civil servants in Hungary. The trainings examined were delivered by the same methodology but conducted by 26 different trainers for 7362 participants. The assessment was not part either of the original project, or of the training design; it was run as an additional activity. The aim of the survey was to show that even with a simple method, applicable even with weak organizational capacities, information can be produced both for validation and curriculum development. The results prove that the trainings made a moderate but statistically significant impact on participants' knowledge and attitudes, and most changes happened in the targeted direction. Beyond the validation of the training methodology, the survey also produced information on trainers' performance and a relatively differentiated picture on participants' learning that can contribute to the further development of the methodology. Thus, the results prove that even with simple quantitative survey methods, evidence for both validation and learning can be produced. The results also raised some questions for further research.

Corresponding author:
Dr Katalin Pallai, National University for Public Service, Budapest, Hungary.
Email: katalin@pallai.hu; telephone: + 43 676 507 8457

Keywords
Training evaluation, effectiveness assessment, trainer performance, integrity education, integrity training

Introduction

The most common method for training evaluation is that organizers ask participants whether they enjoyed the training or believed that it was valuable for them. This reaction survey is a level 1 training evaluation method, according to the most widely used Kirkpatrick scale.[1] While it is important to know how participants feel about the training, a level 1 evaluation constitutes only part of a sound assessment, because when applied alone it does not prove that the training served its purpose, and its contribution to the improvement of the methodology is also limited. Nevertheless, satisfaction surveys prevail in training assessments because they are the simplest to implement. In some cases, an additional reason for their exclusive use is for organizers who are not motivated to face results: the ''smiley sheets'' are a ''noble way'' to avoid producing more sound performance results. Both reasons are attributed to the fact that, in Hungary, civil service training practice applies only satisfaction surveys. (Pallai, 2015a: 6–8)

The starting point to our research was the conviction that ''smiley sheets'' do not fulfil the prudential obligation of public training providers. Organizers of civil servant trainings who are financed from public resources should be obliged to prove that their trainings have achieved the targeted impact on the participants and, when possible, should also use survey results for improving training activities. For these validation and learning purposes, surveying participants' reactions is not enough. The trainings' impacts on participants' learning should also be surveyed, and if possible even the impact on behaviours and organizational results (i.e. level 2 and, if possible, higher level evaluation methods[2]). This claim, in principle, can hardly be challenged. The most prevalent explanation training providers give for the absence of effectiveness assessment is seemingly practical: they claim that it would be complicated and costly and it is often beyond their organizational capacity. The aim of the survey presented in this paper was to demonstrate that validation is possible even with the use of simple questionnaires whose management is identical to the usual satisfaction survey questionnaires – only with the questions changed the analysis becomes more substantive. Formulating different questions is not prohibitive in either cost or organizational capacity. Through our research, beyond setting an example for the viability of effectiveness assessment, we also wanted to see whether some further information can also be deducted from the data that can support further development of training or organization methods.

The survey presented in this paper was carried out on a large public ethics and integrity training programme launched in 2013 in Hungary. The objectives of the trainings were to impact participants' knowledge, attitude and behaviour towards the fight against corruption. Two types of trainings (a one-day long and a three-day long[3]) were delivered to 6692 and 670 civil servants, respectively. Both the single-day trainings and multi-day trainings had identical content and methodology, and were conducted by 24 and eight different trainers respectively, for a total of 26.

The survey was neither part of the original project, nor the training design; it was organized as an additional activity. The idea to change the "smiley sheet"[4] to more substantive questions was raised only shortly before the pilot trainings began. The decision was to assemble questions for two pieces of research in one questionnaire that is completed by participants at the start and the end of the trainings. Within the questionnaire, eight questions were included to validate the effectiveness of the two training methodologies. Four were used to reveal the change in participants' attitudes towards the fight against corruption, and four to explore the change in their cognitive concepts.

In this paper, we have gathered and analysed the results gained from the responses to these eight questions, and have sought to answer the following two research questions:

1. Did the trainings effectively fulfil their purpose in changing participants' knowledge and attitudes towards anticorruption?
2. Can further inferences also be drawn from the results of such a simple exercise? Can the results also support training methodology or programme development activities, or can they contribute to furthering our knowledge regarding trainings and their evaluation?

The results of the evaluation of the data have given a clear answer to the former research question: the comparison of participants' pre- and post-training responses showed significant change in participants' knowledge and attitudes, with the average changes having almost exclusively occurred in the targeted direction. This result proves that the common methodology was effective. The detailed analysis of the data produced results regarding the second research question as well. It gave information about the difference in trainers' performance and a relatively differentiated picture on participants' learning. As we will discuss later, this information can contribute to the further development of the training methodology and the training programme, and can also lead to some questions for further research.

Training evaluation theory and practice

In training evaluation theory, two methodological directions can be distinguished: the quantitative school with its positivist, quantitative, conclusive analysis, and the qualitative school with a focus on improving the transfer of learning through subjective, qualitative, action-oriented exploration (Horton, 2007: 4–5).

The Kirkpatrick model already introduced in Note 1 is focused on another dimension. It classifies the type of impacts trainings can produce. Although this model has been widely criticized since the 1990s (Holton, 1996), we use its levels as a reference system because it is the most well-known and most widely applied tool, and as such offers a terminology that is accessible not only for academics but also for practitioners, who are an equally important target group for our results.

It was also mentioned before that the model of Kirkpatrick identifies four and one of Johnson's five evaluation levels. Although training theorists often see higher level and more accurate analysis as the key challenge in aligning training programmes with

training goals and objectives, the practice lags far behind this expectation (Hall, 2006). Many practitioners involved in relatively complex fields of civil servant training see level 5 analysis as the search for the "holy grail" of evaluation (Horton, 2007: 4), and the overwhelming majority of organizations – even in countries with a highly developed practice of performance management – use only level 1 and some level 2 evaluations (Horton, 2007: 7). This result is not surprising in light of the fact that higher level evaluations are more costly and complicated, can only be completed after the training when trainees are already back in their work environments, and with the cooperation of many other organizational units beyond the one responsible for trainings (Kirkpatrick and Kirkpatrick, 2006). The most advanced training evaluation practice can be applied where such cooperation is achievable, and where a concerted combination of qualitative and quantitative methods can be applied at different assessment levels. These combined methods can fulfil multiple objectives beyond validation. On the one hand, they support training providers in training design and development (learning of providers). On the other hand, the qualitative inquiries may also enhance the learning of participating individuals and organizations. Complex evaluation schemes need to be designed together with the training programme, because besides the relatively simple reaction and impact surveys, they also often use sophisticated methods that facilitate individual and group reflection of a wide range of stakeholders before, during and after the actual training activity. This stakeholder involvement is a key element of the individual and organizational learning process. Notwithstanding the obvious benefits produced by such schemes, they are rarely applied in practice because they can be implemented only when the partner organizations not only support them in principle but also allocate extra resources for the collaboration of stakeholders.

The training programme we evaluated

The public ethics and integrity trainings we evaluated were implemented in the frame of a large anticorruption programme implemented between 2013 and 2015 in Hungary, in cooperation with the Ministry for Public Administration and the National University for Public Service, and supported by the European Union. The goal of the programme was to strengthen public ethics and to introduce the concept and method of public integrity management into Hungarian civil service. Within the frame of the programme a one-day-long training curriculum had to be designed for staff-level civil servants, and a three-day-long training curriculum was designed for civil servants in leadership positions. The target group was the entire Hungarian public administration at the central government level, which is some 150 organizations with approximately 100,000 civil servants. For the implementation, trainers were trained to deliver the trainings for 6692 staff and 670 leaders.

Public integrity, the approach that was taught, is a preventive and holistic approach to anticorruption. Its key tenet is that rules and sanctions are not enough to curb corruption. The opportunities for corruption that the organizational functioning produces should also be minimized through the development of the integrity of the organization. That is, the development of professional and organizational competencies that can produce and

maintain clear, consistent and applicable rules, transparent functioning and rule-consistent behaviour. In the context where our target group works, in most organizations at least the last two components (transparency and rule-consistent behaviour) are often in short supply and many deeply rooted practices clash with the principles of public integrity. Within this context, the objectives of the trainings were to (1) encourage participants to reflect on the integrity-deficient organizational practices, (2) break their apathy and learned helplessness towards positive action and strengthen their trust in possible change and (3) help them see their potential role in the process, that is, lead them to take responsibility. For all these objectives we had to show that criminalization of corruption alone is not enough, that is, the full responsibility cannot be shifted to the legal system. In addition, we had to show the role of a well-organized and functioning public administration in curbing corruption, and with this encourage participants to understand their own potential role in fighting corruption and to take responsibility. All this was done in the hope that this may lead them to return to their organizations and to encourage others to reflect and to change corrupt practices.

Our goal was to initiate learning on all three levels of Bloom's taxonomy, that is, initiating cognitive, affective and behavioural learning (Bloom et al., 1956). The teaching method best suited for this purpose was a participant-centred interactive, experiential method with a strong argumentative component based on facilitated peer learning. (Pallai, 2014: 142–146, 2015b: 96) In order to secure, to the highest degree possible, identical content and method, a detailed methodology was designed with schedules, learning tools, visuals, PowerPoint presentations and handouts. Each trainer involved had to deliver the given content and use the same training tools and method. The trainers were experienced trainers – mostly university faculty members, with the exception of two out of the 26. They received an intensive training, including individual learning of the integrity approach, and 7 days of ToT (training of trainers) sessions, which included deeper discussions of the substantive issues and methodology, as well as a demonstration of the sequence of activities during the training.

The trainings were organized in mixed groups. An effort was made to diversify groups based on a variety of characteristics, such as professional field, organization, background, position, civil service experience, etc.

The training programme was implemented by a very small management unit established at the National University for Public Service. Organization started only a month earlier than the trainings started. The organizers worked under time pressure not only because of the late start but also because the programme had to be finished before the campaigns for the national elections in 2015. There was neither time nor capacity available to establish substantive communication between the team at the university and the public organizations from which participants were recruited. Information flows were often flawed. Within such conditions the first contact between the designers/researchers and participants could be scheduled only for the time of the training. Furthermore, the approaching elections limited follow-up possibilities once the trainings had concluded. These conditions seriously limited what could be planned.

The research aims and plans

In Hungary, the general practice of civil servant training providers is to use only a level 1 evaluation (reaction survey) that explores participants' opinions about the trainings, in some cases together with trainers' reports. With our research we wanted to raise the issue within this domestic context that a level 1 evaluation alone is not sufficient for civil service trainings. While we definitely do not want to suggest that participant satisfaction or trainer experience is not important, we want to call attention to the fact that they are by nature subjective and as such they can misrepresent information regarding trainers and programmes in relation to training objectives. The lack of information on effectiveness is problematic because public resources are spent on civil servant training programmes.[5] The use of public resources can be legitimized only when the training is a vehicle to implement public goals and benefit public organizations. Therefore, we need to prove that our trainings effectively implement their objectives by collecting reliable information on the implementation in order to learn from it and be able to continuously improve performance of our training activities. This means that we need to use evaluation for validation (i.e. producing sound proof for effective impact) and also as a performance management tool to continuously improve our trainings' effectiveness. With our research, we want to show that even with simple tools we can produce more reliable and substantive results both for training validation and development than the usual "smiley sheet" practice.

Our programme offered introductory sensitizing trainings to integrity management. While the larger-than-usual number of participants trained during the programme provided the opportunity for the application of complex analytical tools, the complexity of the theme (which was discussed above) and the introductory nature of the trainings obviously limited research possibilities. The holistic nature of the integrity approach, the complexity of the activities of integrity management and the dependence of each activity on its individual contextual characteristics and dynamics seriously limit the possibilities for impact assessment. While surveying the learning impact on the individual level of participants (level 2) seemed simple and possible, behavioural (level 3) and organizational impact (level 4) seemed more challenging because of the complexity and mutual dependence of possible variables. The same complexity rendered higher level quantitative analysis hardly implementable and clearly prohibited level 5 evaluation (the calculation of return on investment in trainings).

The goals we set were to validate effectiveness of level 2 (the learning of participants), collect information for further development of the programme and method (learning for providers) and explore short- and medium-term behavioural and organizational impact to the highest degree possible. The full research plan was based on a combination of qualitative and quantitative methods. It had three components that seemed implementable even within the constraints set by the lack of organizational cooperation and the limited capacities of the management team. The first component, a quantitative analysis, is the subject of this paper, and was based on a questionnaire with statements that reflected attitudes towards and concepts (general knowledge) about the fight against corruption. Participants completed identical questionnaires at the start and

end of the trainings. The role of the questionnaires was to see the change in responses and through them validate the effectiveness of learning (level 2: evaluation of learning), and through the detailed analysis of individual opinions collect information for further programme development (learning of providers). The second component was to repeat the questionnaires in 6 months after the trainings to see the longer term effects.[6] The third component was an action research to explore personal (level 1–3) and organizational effects (level 4) 6–10 months after the trainings.[7] In summary, the plan was a combination of qualitative and quantitative methods implementable with very limited management capacities and organizational cooperation.

Only the first quantitative component, conducted during the training sessions, is the subject of this paper. In this component, our objective was to make a level 2 evaluation of the short-term learning impact on participants. Although the training method aimed to produce cognitive, affective and behavioural impact, we focused this short-term impact evaluation on the cognitive and attitudinal impact because these are the two components that occur during the training. The behavioural effect evolves after the training when the individual applies the learning within the organizational context. As such, this effect cannot be evaluated at the time of the trainings. Therefore, no attempt was made to include statements regarding behaviours in the questionnaire; instead, the above-mentioned action research aimed to produce such results.[8]

The research method

A quasi-experimental, nonrandomized pre-post quantitative evaluation research design (Powell, 2006: 110) was implemented. As was presented earlier, participants based on their organizational position were assigned to either one-day (below managerial level) or three-day-long training sessions (managers and leaders). A total of 363 one-day sessions and 44 three-day sessions were held between September 2013 and January 2014 across Hungary. As described earlier, a total of 26 trainers were involved – six of whom held both one-day and three-day courses.

All trainers instructed participants with standard materials. The above detailed description of the preparation of the curriculum and the training of the trainers were meant to show that trainers received sound preparation for producing similar results. With this in mind, one can assume that the main difference among trainer performance can be attributed both to systematic differences among groups in their initial attitudes, and the individual competencies of each trainer.

The number of participants in training groups varied between five and 34, with an average group size of 18.4 in the case of one-day trainings and eight and 24 with an average size of 15.2 in the case of three day trainings (for further description of participants in both types of trainings, see Annex 1). A total of 6692 participants of single-day trainings and 670 participants of the three-day courses completed a questionnaire prior to the training session and immediately after it. Respondents were told to choose and write the same pseudonym for both questionnaires. This technique ensured the possibility of matching precisely the pre- and post-surveys and tracking the change of responses, even at the individual level, for each item.

Table 1. How much do you agree with the following statements on a 1–5 scale (1: totally disagree, 5: totally agree)?

Questions related to **attitudes** towards corruption and anticorruption	Questions related to **knowledge** about corruption and anticorruption
A1. The corruption experienced in this country is no particular cause for concern, because it is an inherent feature of transformation.	K1. Corruption should primarily be fought using legal instruments.
A2. Corruption is as old as mankind and not much should be done to fight it.	K2. Corruption can be fought the most effectively through transparency.
A3. In Hungary, corruption has assumed such proportions that fighting it has become impossible.	K3. The best remedy for corruption is fast and efficient administration.
A4. It is possible to change people's thinking about what's right and wrong, allowing them to apply self-criticism to previously accepted procedures from which they derive personal benefits.	K4. Well organized public administration can significantly reduce external attempts at corruption.

For our survey we included eight questions regarding corruption and anticorruption. In both (pre- and post-training) questionnaires these eight questions were identical in wording. The questionnaires included additional questions that belonged to another research plan, and in the pre-training questionnaire some additional questions were also included that mapped the demographic characteristics of the respondents (e.g. gender, type of organization, length of employment, etc.) and each respondent's organizational commitment. Out of the eight relevant questions, four were related to attitudes and four to cognitive knowledge about (anti)corruption (Table 1).

As discussed previously, the objective of the trainings was to initiate reflection on corrupt practices, and learned helplessness, build trust for the possibility of change, show the role of well-organized and functioning public administration in curbing corruption, and with this encourage participants to understand their own role in anticorruption and to take responsibility. This is why the attitude questions revolved around apathy, helplessness and trust, and knowledge questions determined whether participants understood better that not only legal instruments, but also transparency, organization and effectiveness are key instruments in fighting corruption.

Participants were asked to express the level of their agreement or disagreement with each statement on a 1–5 Likert scale.

Because the participants of one-day and three-day sessions differed not just in the length of their programme but also in their hierarchical position within their various organizations, it is impossible to ignore the impact these factors may have on attitudes towards anticorruption and opinions after having completed the trainings. We must, therefore, treat the results of the two types of trainings independently from each other and present the results separately.

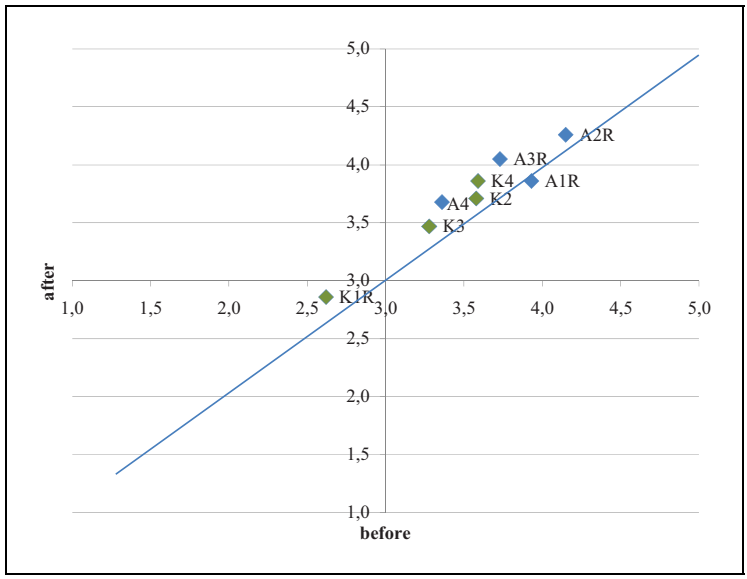


Figure 1. Average scores on a 1–5 scale along the examined variables (before and after training), one-day training sessions. (Colour online only.)

Results

Changes in attitudes and knowledge about corruption after the training sessions

Paired sample *t*-tests were conducted to compare the level of agreement with the four attitudinal and four knowledge-based statements before and after the training. Results confirmed that in both training designs, changes of average agreement along each item were statistically significant; however, these changes – based on Cohen's *d* effect size measure – can be interpreted as moderate at most (for detailed *t*-test results, see Annex 3).

Figures 1 and 2 present the scatter plot of before–after average scores along the examined variables, in the cases of one-day and three-day designs. Rotated-scale versions of A1, A2, A3 and K1 statements indicated by an additional letter, "R", were used in further parts of the analysis in order to have all items with high grade (grade 5) as expected. (The blue line represents the hypothetical zero difference between before and after average scores.)

Based on Cohen's *d* measures in both training types, the strongest overall changes could be detected along item A4 (it is possible to change people's thinking about what's right and wrong, allowing them to apply self-criticism to previously accepted procedures from which they derive personal benefits) and item K4 (well organized public administration can significantly reduce external attempts at corruption), while the weakest changes were along item A1 (the corruption experienced in this country is no particular cause for concern, because it is an inherent feature of transformation) and item K2

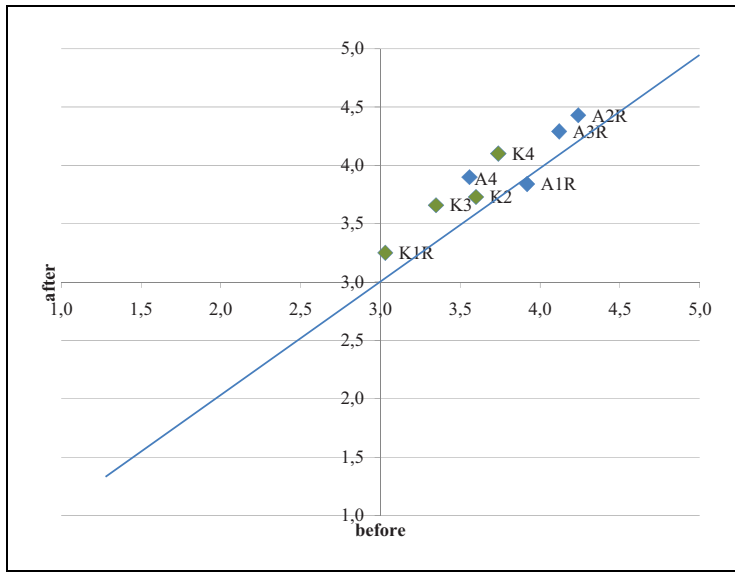


Figure 2. Average scores on a 1–5 scale along the examined variables (before and after training), three-day training sessions. (Colour online only.)

(corruption can be fought the most effectively through transparency). Agreement with all statements changed in the expected way – that is, growing, except in the case of item A1R: after the training sessions participants became, statistically speaking, significantly but only a little less anxious about the level of corruption in Hungary. This result might be surprising at first sight, but taking into consideration that participants were not only told about the worrying prevalence of corruption but also during the training they discussed related topics and gained a deeper understanding of the phenomenon, it is reasonable that their tolerance threshold towards the normality of corruption increased a little. Nevertheless, this result offers a point for reflection for curriculum development (discussed in more depth later in the Conclusions).

Moving further from one-by-one analysis, we also calculated separately the aggregated average agreement with attitudinal (A) and knowledge-based (K) items and finally a general total score of average agreement with both groups of items. Paired sample *t*-tests confirmed that in both cases of training designs, participants on average became significantly ($t_{\text{one-day}} = 24.555$, $df_{\text{one-day}} = 6504$, $p_{\text{one-day}} = 0.000 < 0.05$; $t_{\text{three-day}} = 7.658$, $df_{\text{three-day}} = 650$, $p_{\text{three-day}} = 0.000 < 0.05$) more committed to the fight against corruption and the level of their overall knowledge about this fight also increased ($t_{\text{one-day}} = 33.537$, $df_{\text{one-day}} = 6583$, $p_{\text{one-day}} = 0.000 < 0.05$; $t_{\text{three-day}} = 12.797$, $df_{\text{three-day}} = 650$, $p_{\text{three-day}} = 0.000 < 0.05$). As a result of these, the overall increase in the level of total agreement was also statistically significant regardless of the length of the training programme ($t_{\text{one-day}} = 38.805$, $df_{\text{one-day}} = 6470$, $p_{\text{one-day}} = 0.000 < 0.05$; $t_{\text{three-day}} = 13.978$, $df_{\text{three-day}} = 646$, $p_{\text{three-day}} = 0.000 < 0.05$; Table 2).

Table 2. Average scores on a 1–5 scale before and after training along aggregated attitudinal, knowledge-based and total agreement.

	Before (mean)	After (mean)	Difference (after-before)	*t*-value	*df*	*p*	Cohen's *d*
ONE-DAY SESSIONS							
Attitudinal	3.79	3.96	0.17	24.56	6,504	0.000***	0.28
Knowledge-based	3.27	3.48	0.21	33.54	6,583	0.000***	0.42
Total	3.53	3.72	0.19	38.81	6,470	0.000***	0.45
THREE-DAY SESSIONS							
Attitudinal	3.96	4.11	0.15	7.66	650	0.000***	0.27
Knowledge-based	3.43	3.68	0.25	12.80	650	0.000***	0.51
Total	3.70	3.90	0.20	13.98	646	0.000***	0.51

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.
Paired sample *t*-tests were used for comparing before and after scores.

Whilst changes in the cases of attitudinal variables can be interpreted as small (Cohen's $d_{\text{one-day}} = 0.28$, Cohen's $d_{\text{three-day}} = 0.27$), the general increase of agreement with knowledge-based indicators was found to be closer to moderate (Cohen's $d_{\text{one-day}} = 0.42$, Cohen's $d_{\text{three-day}} = 0.51$).

Figure 3 also shows that after the sessions, the average attitude and knowledge of participants of the one-day training courses increased approximately as high as the average initial scores of three-day participants. In this sense, one-day training elevated staff-level participants' attitude and knowledge to almost exactly to the same as the initial level of leaders.

Comparing the homogeneity of opinions before and after trainings, it can be found that one-day trainings significantly homogenized their participants along with their knowledge about fighting against corruption ($Var_{\text{before}} = 0.52$, $Var_{\text{after}} = 0.49$, $r = 0.484$, $p_{\text{Pitman-Morgan-test}} = 0.000 < 0.05$), while no such change can be detected in the case of three-day training participants. Neither participants of one-day nor three-day sessions became more homogeneous in terms of their attitudes towards fighting against corruption.

All in all, as Table 3 shows, after the one- and three-day training courses five out of 10 participants became more committed towards fighting against corruption, a fifth of them remained on the same level, while a quarter of them became less committed. Approximately six out of 10 respondents gained more knowledge about fighting against corruption, one-fifth of them did not improve their knowledge, while a quarter became less knowledgeable about the possible tools of preventing corruption. Approximately two-thirds of respondents showed signs of improvement, either through attitudes or knowledge, and a quarter changed their opinion into the opposite of what was expected.

This remarkable amount of respondents who changed their attitudes or knowledge in another, unexpected direction raised the question of whether they are different in any

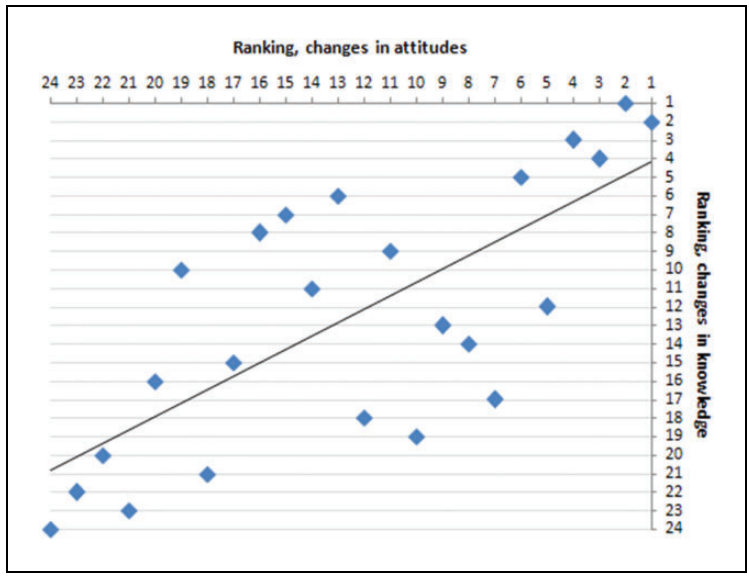


Figure 3. Correlation between trainers' ranking (one-day, $r = 0.73^{***}$).

Table 3. Respondent categories based on changes in the average level of attitudes, knowledge and total achievement (%).

	Attitudinal		Knowledge-based		Total	
ONE-DAY SESSIONS	*N*	%	*N*	%	*N*	%
Decreased	1751	26.9	1516	23.0	1578	24.4
Did not change	1365	21.0	1418	21.5	815	12.6
Increased	3389	52.1	3651	55.5	4078	63.0
Total	6505	100.0	6584	100.0	6471	100.0
THREE-DAY SESSIONS						
Decreased	175	26.9	134	20.6	155	24.0
Did not change	137	21.0	119	18.3	71	11.0
Increased	339	52.1	398	61.1	421	65.0
Total	651	100.0	651	100.0	647	100.0

other sense from the other groups. Firstly, we assumed that general commitment towards making change and progress in the organizational culture might positively affect the outcome of training. Results revealed that those showing progress either along attitudes or knowledge do not differ significantly in the level of general commitment towards making change compared to those showing regression. However, leaders working for their organization for less than 2 years (chi-square = 19.854, $df = 6$, $p = 0.003 < 0.05$, Cramer's $V = 0.13$), non-executives heading towards retirement (chi-square = 10.491,

$df = 2$, $p = 0.005 < 0.05$, Cramer's $V = 0.04$) and male non-executives (chi-square = 9.797, $df = 2$, $p = 0.007 < 0.05$, Cramer's $V = 0.04$) are weakly but significantly more likely to show regression in their level of knowledge. This is a result that is surely worthy of deeper investigation in later research focused on these sub-groups.

Initial response effect on changes in attitudes and knowledge

Ordinary least squares (OLS) linear regression analysis on the relationship between the average initial scores along attitudinal and knowledge-based variables and the absolute value of changes revealed that the lower initial average score someone had, either along attitudinal or knowledge-based dimensions, the higher the amount of absolute change she or he showed in his or her opinions (see Annex 4). This means that those being initially less engaged and less informed in the topic of fighting against corruption were more likely, on average, to shift their opinions, compared to those being more engaged and more informed. Thus, it was proven that initial opinions did have an effect on the outcome. There is also no significant difference between participants of one-day and three-day sessions regarding whether the average absolute value of their opinion changed.

Trainers' effect

Despite all attempts to standardize the context and content of the training courses, trainers may have easily differed in their effectiveness of producing the required outcome in participants at the end of the training. The effect of trainers was analysed in order to shed light on the importance of personal factors, too. Although respondents were not randomly and uniformly assigned to trainers, the organizers attempted to construct relatively mixed groups. Because of the lack of randomization in a statistical sense, first it must be analysed if trainers were significantly different due to the average attitude and knowledge of the participants they worked with. One-way analysis of variance (ANOVA) tests provided evidence that, while trainers of one-day training sessions were significantly different both along the average attitudes of assigned participants ($F = 3.536$, $df_1 = 23$, $df_2 = 6526$, $p = 0.000 < 0.05$) and their average level of knowledge-based elements ($F = 1.913$, $df_1 = 23$, $df_2 = 6582$, $p = 0.005 < 0.05$), trainers of three-day training courses did not differ significantly from each other either along the level of attitudes ($p = 0.353$) or the level of knowledge ($p = 0.173$) of participants assigned.

Our aim was to examine if any kind of ranking of trainers could be detected, in respect to their effectiveness. Effectiveness can be measured as the average change in participants' attitudes and knowledge. We conducted OLS linear regression analysis in which trainers were represented with dummy (0–1) variables. In every model the trainer with the lowest average difference between pre- and post-training scores was assigned to be the reference category. The possible effect of initial average scores was controlled by involving the variable it measured. As a result, the descending order of standardized regression coefficients (beta) belonging to each trainer represented the ranking of effectiveness.

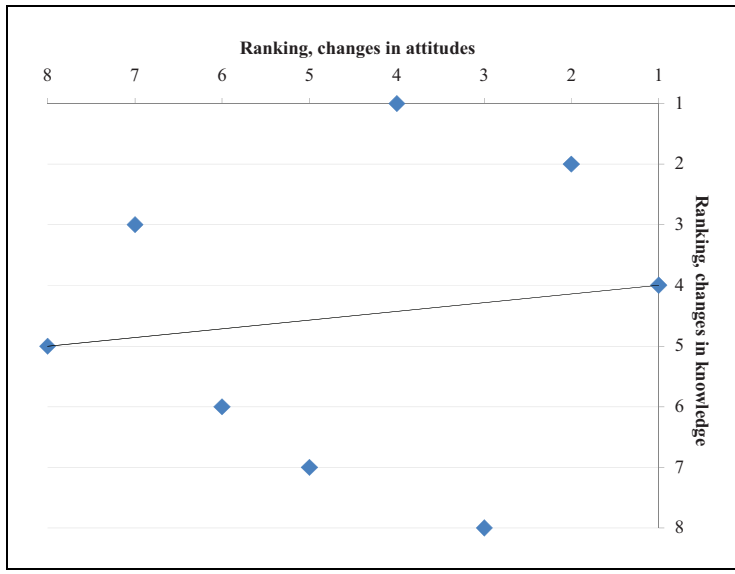


Figure 4. Correlation between trainers' ranking (three-day, $r = 0.14$, $p = 0.74 > 0.05$).

If we take the one-day trainers' rankings (where rank 1 belongs to the most effective trainer) along the attitude and knowledge axes we can find a strong positive correlation ($r = 0.73$, $p = 0.000 < 0.05$) between both achievements. The greater positive shift a trainer could achieve along attitudes, the greater achievement s/he caused along knowledge (Figure 3).

However, there is no similar relationship between the two rankings in the case of trainers of three-day sessions (Figure 4). When the training is longer and executives and managers participate in the session, those few trainers show no clear pattern of effectiveness.[9]

Conclusions

Regarding the first research question, the comparison of participants' pre- and post-training responses showed a significant change in participants' knowledge and attitudes, and the average changes occurred in the targeted direction. This result proves that the common methodology was effective.

The results also show that significant changes happened in the responses to each question, except one. The largest change happened in the two most important questions: "*A4: It is possible to change people's thinking about what's right and wrong, allowing them to apply self-criticism to previously accepted procedures from which they derive personal benefits*" and "*K4: Well organized public administration can significantly reduce external attempts at corruption*". The change in support of attitude question is important because it indicates increasing trust in the possibility to change

corruption-tolerant attitudes while also signalling a decrease of learned helplessness, and the K4 question is important because it reflects that understanding of the key tenet of the integrity approach increased. The only question where average opinion changed in the wrong direction was "*A1: The corruption experienced in this country is no particular cause for concern, because it is an inherent feature of transformation*". This result offers important learning for the further development of the training: it indicates that the difference between sober analysis and understanding, on the one hand, and acceptance on the other hand should be rendered more explicit in the curriculum. At the same time, the detailed analysis of trainers' results showed in the case of this question that the result of six out of the 24 one-day trainers is positive (in the targeted direction), six have negative change and 12 have a non-significant effect. In the cases of three-day trainers, one out of eight trainers achieved the targeted opinion change and one the opposite, while the remaining six trainers did not cause any change. This result reinforced our hypothesis that major changes are not needed, only fine tuning of related activities and messages.

Some results about trainers' performance have already been discussed. Here we mention only two more. Although the questions included in the questionnaire all related to anticorruption and no direct question was raised regarding the trainers, the results give grounds for the comparison of the trainers' performance. It is important to mention that in our case all trainers were obliged to deliver identical content and methodology and they had relatively similar profiles. They were not only experienced trainers but university teachers as well, and all had sound knowledge and credibility in public administration. Probably even more variance could result in performance if trainers without academic backgrounds would have also participated in the programme. This result shows the importance of trainer selection.

It is a result worth mentioning that while there is significant correlation ($r = 0.73$) between the performance of trainers in attitude and knowledge change of trainees during the one-day-long trainings, such a relationship does not exist between the attitude and knowledge change results of trainers of the three-day-long trainings ($r = 0.14$). Moreover, only three trainers produce similar results in the two domains, while others have pronouncedly different performances. It is also an interesting finding that the trainer achieving the second ranking in the three-day trainings has achieved one of the worst rankings (23) in the one-day trainings. This individual result is not strong evidence, but if more similar results show similar patterns they might indicate the importance of trainer and programme-fit.

Potential practical applications and further research

The aim of our research was to demonstrate that even with simple evaluation methods more relevant information can be produced than with level 1 reaction surveys. The research component discussed in this paper was a first attempt to produce a level 2 performance evaluation of an integrity training programme for civil servants. The results show that this type of analysis, when conducted on such large participant groups, can not only validate the effectiveness of the method, but it can also deliver information for further improvement and development of the programme and methodology as follows.

- The results regarding different questions indicate which components of the training worked and which ones need to be further developed or substituted by alternative components, or perhaps their delivery should be improved.
- The results regarding trainers' performance can indicate areas for further preparation and ToT, or can support selection of trainers on a more sound basis than the satisfaction survey results.
- The detailed analysis of the change of individual opinions shows that within the positive aggregate change there are minor groups who changed their opinions in the wrong direction. Further investigation of these groups, including the application of qualitative research methods, could produce interesting insights into the nature and process of failures to avoid.

In our case the training programme was given and the research presented in this paper was only an additional activity. The programme parameters could not be modified for the purpose of better research and limited opportunities were offered for the research. Our research was an attempt to show that even in these conditions, with the application of a simple methodology, some important results can be produced both for validation and for methodological and organizational learning purposes, much more than participants' satisfaction surveys (level 1 reaction survey) could ever produce. This message is aimed to change the practice and direct attention to the importance of the survey of effectiveness.

If a similar programme could be designed together with the aim of efficiency research, researchers should pay attention to a full randomization of participants in order to be able to separate the possible effect of, for instance, length of training as an impact and the effect of hierarchical position in the organization. Another important further step could be if the same respondents were asked to express their satisfaction with their trainers and fill out the attitudinal and knowledge-based questions, too. After the training sessions it is also worth asking trainers about their subjective feeling of achievement in order to analyse the possible match or mismatch between the subjective and objective indicators of their efficiency. Finally, a follow-up questionnaire after a couple of weeks or months with exactly the same attitudinal and knowledge-based questions can show if changes in attitudes and knowledge are lasting or not.

It would also be important to assess how effective the impacts we produced are compared to other training programmes, but as we do not know about any similar impact assessment of integrity trainings, we cannot find any comparative data and therefore we cannot assess this. The source of the problem is not only the lack of similar assessment but the understandable fact, as well, that organizations are reluctant to share programme evaluation data. If more organizations would produce similar performance assessments for their integrity trainings, and would be willing to publish their results, we could compare results, assess our relative performances and learn from them. It would also be an option to launch a comparative project in countries faced with similar corruption phenomena.

Authors' note

This paper has been presented at EGPA Conference, Toulouse, 26 August 2015 in the Permanent Study Group IX: Teaching Public Administration.

Conflict of interest

The author(s) declared the following potential conflicts of interest with respect to the research, authorship and/or publication of this article: One author of the article, Katalin Pallai, was the designer of the training curricula that was examined by the survey, but the survey data processing was conducted by an independent data processing firm and the analysis was conducted by an independent analyst, who is the co-author of the article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: The trainings discussed in the paper and the questionnaire survey whose results have been used for writing this article was financed by State Reform Operative Program 1.1.12. The first analysis of the data, whose results are available in Hungarian, was financed by the Centre for Excellence in Integrity of the National University for Public Service in Budapest. At the time of the analysis the centre was led by Katalin Pallai, the primary author. At the same time, the analysis presented in the article received no funding.

Notes

1. According to the most widely used four-level categorization of training evaluation developed by Donald Kirkpatrick (1998) in the 1950s, and also by the five-level scale by Jack Phillips et al. (2007), level 1 is the Reaction survey where participants are asked about their opinion about the training (sometimes referred to as the "smiley sheets").
2. In both cases the scale of Kirkpatrick's (1998) and Philips et al.'s (2007) level 2 evaluates the learning, level 3 the behavioural impact/application of the new learning and level 4 the results/impact on the organization. Level 5, added by Phillips et al. (2007), is the ROI (Return of Investment).
3. One-day-long training contains eight contact hours and three-day-long 20 contact hours.
4. We find the "smiley sheet" term adequate because the examination of satisfaction surveys conducted on similar integrity trainings for the same target group shows that all trainers are graded above 3.85 (on a 0–4 scale) – even in the case of trainers who, by our survey, could produce only minimal or zero learning impact.
5. Not only the financial resources but the time of civil servants as well.
6. This part could not be implemented.
7. The action research was started but could not be completed and its results are not included in this paper. A short report about the results that the unfinished action research could produce is available in Hungarian in Siklaki (2015).
8. The action research (although unfinished) revealed many important behavioural impacts. One was that the facilitated collaborative peer interaction that participants experienced during the trainings meant for many, especially for participants in leadership position, a model for effective interaction and collaboration and had strong behavioural impact (Siklaki, 2015).
9. Although during this study we did not ask respondents about their satisfaction with the training and the trainer, 18 trainers participated in a project right after our survey in which they held nearly identical trainings for participants from identical target groups as in this study. Those participants had to fill out an evaluation questionnaire that surveyed their satisfaction. We calculated the average satisfaction score for each of the 18 trainers. Later, we compared their satisfaction ranking and their achievement ranking. We expected weak positive correlation

between someone's satisfaction ranking and effectiveness ranking. The results proved that trainers having greater targeted impact on participants' attitude and knowledge (total achievement) were evaluated better by respondents of the other study ($r = 0.64$, $p = 0.002 < 0.05$). However, if trainer selection happened on the basis of satisfaction results, two of the nine most effective trainers would not be selected.

References

Bloom BS, Engelhart MD, Furst EJ, et al. (eds) (1956) *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay Co Inc.

Hall B (2006) Missing Link? *Training Magazine*, February, 22.

Holton E (1996) The flawed four- level evaluation model. *Human Resource Development Quarterly* 7(1): 5–21.

Horton S (2007) Evaluation of leadership development and training in the British senior civil service: The search for the Holy Grail. In: *leading the future of the public sector: the third transatlantic dialogue*, University of Delaware, Newark, Delaware, USA.

Kirkpartick DL (1998) *Evaluating Training Programs: The Four Levels*. San Francisco, CA: Berrett-Koehler.

Kirkpatrick DL and Kirkpatrick JL (2006) *Evaluating Training Programs*. 3rd ed. San Francisco, CA: Berrett-Koehler.

Pallai K (2014) Design considerations and results of a large integrity training program for civil servants in Hungary. *Pro Publico Bono* 2014(3): 137–147.

Pallai K (2015a) Impact of culture on the results of a tender for faculty for a post-graduate education program: the role of sense making, expectations and networks of potential teachers. In: *the IASIA conference*, Paris, 8 July, p.24.

Pallai K (2015b) Integritásoktatása közszolgálati képzésben – építkező oktatás a fejlődő integritásért. In: Dargay E and Juhasz LM (eds) *Integritás és antikorrupció Magyarországon*. Budapest: National University for Public Service.

Phillips J, Phillips P, Stone R, et al. (2007) *The ROI Field Book: Strategies for Implementing ROI in HR and Training*. Burlington, MA: Butterworth-Heinemann/Elsevier.

Powell RR (2006) Evaluation research: An overview. *Library Trends* 55(1): 102–120.

Siklaki I (2015) Az ITK integritás-tréningjeinek hatása a közigazgatási vezetőkmunkájára és gondolkodására. (Impact of integrity trainings on the thinking and work of public leaders). Research report prepared for the Center for Excellence in Integrity at the National University for Public Service, Budapest.

Annexes

Annex 1. Descriptive statistics of participants of one-day and three-day training sessions.

	One-day training sessions		Three-day training sessions	
	N	%	N	%
TYPE OF ORGANIZATION				
Governmental	2917	43.6	171	25.5
Territorial administration	2569	38.4	354	52.8
Other administration	480	7.2	-	-
Other type	358	5.3	65	9.7
No answer	368	5.5	80	11.9
TOTAL	6692	100.0	670	100.0
FOR HOW LONG HAVE YOU BEEN WORKING FOR THIS ORGANIZATION?				
Less than 2 years	1719	25.7	155	23.1
2–5 years	1056	15.8	91	13.6
5–10 years	942	14.1	71	10.6
10 years or more	2549	38.1	276	41.2
No answer	426	6.4	77	11.5
TOTAL	6692	100.0	670	100.0
ARE YOU GOING TO RETIRE IN THE NEXT COUPLE OF YEARS?				
Yes, within 3 years at most	427	6.4	30	4.5
No	5836	87.2	562	83.9
No answer	429	6.4	78	11.6
TOTAL	6692	100.0	670	100.0
GENDER				
Male	1553	23.2	271	40.4
Female	4717	70.5	321	47.9
No answer	422	6.3	78	11.6
TOTAL	6692	100.0	670	100.0

Annex 2. List of the eight statements in the questionnaire that belong to our research.

1. It is possible to change people's thinking about what's right and wrong, allowing them to apply self-criticism to previously accepted procedures from which they derive personal benefits.
2. Well organized public administration can significantly reduce external attempts at corruption.
3. The best remedy for corruption is fast and efficient administration.
4. Corruption could be reduced by openly stating the price of things for which people have been paying secretly.
5. In Hungary, corruption has assumed such proportions that fighting it has become impossible.
6. Corruption is as old as mankind and not much should be done to fight it.
7. Corruption can be fought the most effectively through transparency.
8. The corruption experienced in this country is no particular cause for concern, because it is an inherent feature of transformation.

Source: Integrity training questionnaire, 2013.

Annex 3. Paired sample *t*-test results (one-day and three-day trainings).

ONE-DAY TRAINING	Mean (before)	Mean (after)	Change (after – before)	*df*	*t* (sign.)	Abs (Cohen's *d*)
A1. The corruption experienced in this country is no particular cause for concern, because it is an inherent feature of transformation.	2.07	2.14	+0.07	6603	5.421***	0.07
A2. Corruption is as old as mankind and not much should be done to fight it.	1.85	1.74	–0.11	6644	–8.577***	0.11
A3. In Hungary, corruption has assumed such proportions that fighting it has become impossible.	2.27	1.95	–0.32	6651	–24.525***	0.31
A4. It is possible to change people's thinking about what's right and wrong, allowing them to apply self-criticism to previously accepted procedures from which they derive personal benefits.	3.36	3.68	+0.32	6587	26.769***	0.36
K1. Corruption should primarily be fought using legal instruments.	3.38	3.14	–0.24	6654	–19.374***	0.24
K2. Corruption can be fought the most effectively through transparency.	3.58	3.71	+0.13	6645	10.706***	0.14
K3. The best remedy for corruption is fast and efficient administration.	3.28	3.47	+0.19	6637	15.411***	0.18
K4. Well organized public administration can significantly reduce external attempts at corruption.	3.59	3.86	+0.27	6626	24.074***	0.32

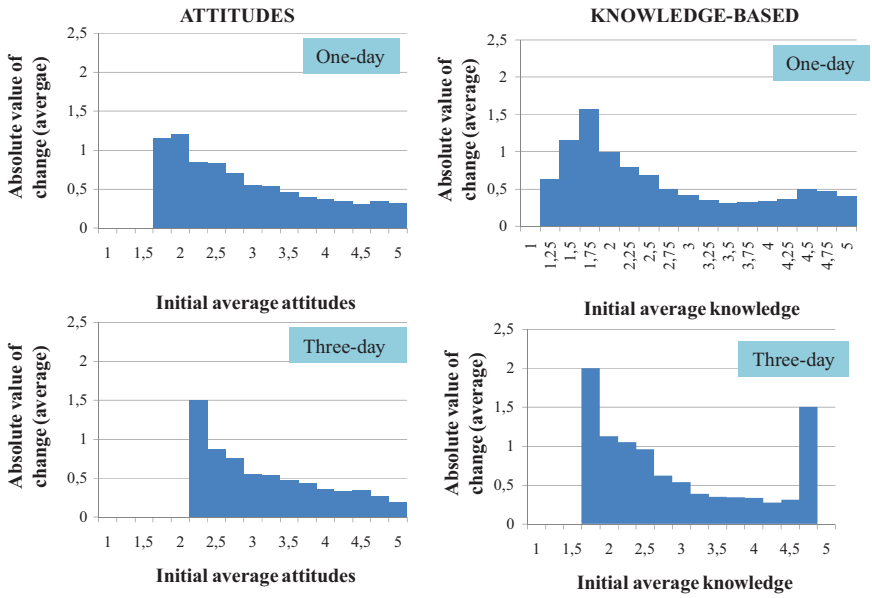
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Annex 3. (Continued)

THREE-DAY TRAINING	Mean (before)	Mean (after)	Change (after – before)	*df*	*t* (sign.)	Abs (Cohen's *d*)
A1. The corruption experienced in this country is no particular cause for concern, because it is an inherent feature of transformation.	2.08	2.16	+0.08	657	2.118*	0.04
A2. Corruption is as old as mankind and not much should be done to fight it.	1.76	1.57	–0.19	659	–5.408***	0.22
A3. In Hungary, corruption has assumed such proportions that fighting it has become impossible.	1.88	1.71	–0.17	659	–5.217***	0.20
A4. It is possible to change people's thinking about what's right and wrong, allowing them to apply self-criticism to previously accepted procedures from which they derive personal benefits.	3.56	3.90	+0.34	656	8.650***	0.40
K1. Corruption should primarily be fought using legal instruments.	2.97	2.75	–0.22	658	–5.640***	0.24
K2. Corruption can be fought the most effectively through transparency.	3.60	3.73	+0.13	655	3.439**	0.14
K3. The best remedy for corruption is fast and efficient administration.	3.35	3.66	+0.31	658	7.722***	0.31
K4. Well organized public administration can significantly reduce external attempts at corruption.	3.74	4.10	+0.36	656	9.670***	0.45

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Annex 4. Relationship between initial average opinion and average absolute change in opinion.



Annex 5. Results of ordinary least squares (OLS) linear regressions.

	Dependent variable: absolute value of difference (OLS)							
	One-day, attitudinal		One-day, knowledge		Three-day, attitudinal		Three-day, knowledge	
	B (S.E.)	beta	*B* (S.E.)	beta	*B* (S.E.)	beta	*B* (S.E.)	beta
Initial value	–0.192*** (0.008)	–0.299	–0.222*** (0.008)	–0.316	–0.191*** (0.024)	–0.298	–0.284*** (0.026)	–0.394
Constant	1.158 (0.029)	-	1.136 (0.027)	-	1.159 (0.096)	-	1.407 (0.090)	-
Observations	6504		6504		650		650	
R^2	0.089		0.100		0.089		0.155	
Adjusted R^2	0.089		0.099		0.087		0.154	
F-statistics	636.706		728.116		63.138		119.343	
df_1/df_2	1/6503		1/6582		1/649		1/649	
p	0.000***		0.000***		0.000***		0.000***	

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.