What Can We Learn from a Simple Survey? Effectiveness Assessment of Public Integrity Training Courses - A Case Study

Katalin Pallai, PhD, Associate Professor,

National University for Public Service, Budapest, Hungary

katalin@pallai.hu

CleanGovBiz

Declaration of conflicting interests
The author of this article, Katalin Pallai, was the designer of the training curriculum that was examined by the survey. The data processing was conducted by an independent data processing firm and the analysis of the data that this article refers to was conducted by an independent analyst, Aniko Gregor. (Gregor 2015) Some findings have been presented in a joint paper with the analyst at the EGPA Annual Conference in 2015. (Pallai-Gregor 2015) A co-authored paper is under review (awaiting reviewer assessment) for the Teaching Public Administration Journal but will not be printed before the OECD Integrity Forum. Although the content of the co-authored paper overlaps with the content of this paper, the focus of the two papers are different.

Abstract

The OECD Report on Ethics training for public officials (2013), a study that surveyed the practices of 22 countries, concluded that evaluation methodologies focus mostly on outputs. Effectiveness results are usually not available. In 2013-14 an effectiveness assessment survey of a large public ethics and integrity training programme was conducted in Hungary; it examined the knowledge gained and change of attitude among 7,362 participants with the use of pre- and post-training questionnaires. The aim of the survey was to validate the training method and curriculum, collect information for the training providers on how to improve the effectiveness of the training and to demonstrate that even with a simple methodology, applicable even with limited organisational capabilities, information can be produced both to validate and develop the curriculum further. The survey, the data analysis and some results have already been published in more depth than the limits of this paper would allow. The focus of the previous publications was the validation of the training programme. The focus in this paper is on those results of the research that can support the learning of trainers and providers. What can we learn from the detailed analysis of individual participants' opinion changes? How diverse are these changes? What could be the cause of the variety? Who are those participants with whom we failed? What are the differences in trainers' performance? What are the differences between the results of satisfaction and learning impact surveys? Our data will not produce conclusive answers for all questions. However, regarding the points where no definite conclusions can be drawn from this research, important questions can still be raised for further research and suggestions can also be formulated for the design of further research on the effectiveness of Public Ethics and Integrity training.

Key Words: effectiveness assessment of training, training evaluation, trainers' performance, integrity education, integrity training

Introduction

During the "Anti-corruption Policy and Integrity Training" conference organised by OECD and OSCE in Vilnius in 2011, experts from Eastern Europe, Central Asia and several OECD countries agreed that new, more advanced and interactive approaches were needed for the education of public ethics and integrity. (OECD 2013: 8) A project was developed by the OECD-EU SIGMA programme, together with ACN[1] and in cooperation with the OECD Public Integrity Network, to survey the actual practices. Public officials from 22 countries provided information about their training activities. On the basis of the research, the Report "Ethics Trainings for Public Officials" concluded that "evaluation of the effectiveness of ethics training is a challenging task. Current evaluation methodologies focus mostly on outputs such as the number of public officials trained". (OECD 2013: 10) The Report also states that "[it] is important to evaluate the results of ethics training in order to ensure that the use of public money is producing the desired effects, and to look for ways to improve and to modernise the training approaches to make them more effective" (OECD 2013: 24)

In 2013 and 2014 a large "Public Ethics and Integrity" training programme was implemented that reached 7,362 civil servants in Hungary. Although effectiveness assessment had not previously been part of civil servant training in Hungary either, an effectiveness assessment survey was also conducted in the context of this programme. The direct objectives of the survey were to validate the effectiveness of the training by measuring the change in participants' knowledge and attitudes and to collect information in order to improve the training methods. At the same time, the survey had one more, connected objective: it intended to demonstrate that even using a simple methodology, applicable to even limited organisational capabilities, information can be produced for both validating the effectiveness of training and developing training curricula and methods further.

The survey, the data analysis and some results have already been published in more depth than the limits of this paper would allow. (in Gregor 2015; Pallai-Gregor 2015) The focus of the previous publications

---

[1] ACN stands for Anti-Corruption Network for Eastern Europe and Central Asia

was the validation of the training courses, that is, proving that the training produced the desired effects. The focus in this paper is on those results of the research that can support the learning of trainers and providers. For this discussion, only a short summary of the key points from the already available publications will be given and the technical details will be omitted. The question here is what is beyond the positive average figures that were presented for the validation. What can we learn from the detailed analysis of individual participants' opinion changes? How much variety is there among the participants' learning? What could be the cause of the variety? What are the common characteristics of participants whose opinions we could not change in the desired direction? What is the difference in trainers' performance? What is the difference between the results of satisfaction and learning impact surveys?

However, beside the intention to discuss all these questions, we must also note that this research was not initiated with sufficient planning. It was only a last minute decision that took the opportunity to distribute a questionnaire among the participants of a training programme that was already starting. The positive consequence of this situation is that our results indicate how much can be learned even from a simple survey that is implementable even with limited organizational capabilities. This was one of the objectives of the research. By the same token, our data will not produce conclusive answers for all questions. However, regarding the points where no definite conclusions can be drawn from this research, important questions can still be raised for further research and suggestions can also be formulated on how to design similar surveys on the effectiveness of Public Ethics and Integrity training courses.

1. Training evaluation

1.1. Types of training evaluations and their use

The most widely known categorisation of training evaluation is that of Kirkpatrick, who defines four levels of training impact surveys:

- Reaction surveys, where participants are asked their opinion, are Level 1.

- Learning impact surveys, which examine the extent to which participants have improved their knowledge and skills and have changed their attitudes, are Level 2.

- Behavioural impact surveys that examine the extent to which participants have changed their behaviour in their workplace are Level 3.

- Surveys that analyse the organisational benefits of the results are Level 4.

Philips (Philips at al 2007) has added a 5th level to the Kirkpatrick model, Return on Investment (ROI), aimed at quantifying and comparing the investments incurred and the benefits produced by the training activities.

For Level 1 evaluation, a simple post-training survey is enough. In principle, level 2 evaluations can be accomplished during the training sessions with pre- and post-training surveys. The advantage of this solution is that it can be implemented even when the project is restricted to the time of the training; the disadvantage is that only the short term learning impact can be evaluated right after the training. For the analysis of longer term impacts, one more survey is necessary later but this can turn out to be difficult to organise in certain contexts. Level 3 and 4 evaluation surveys need to be conducted not only before and during the training programme but after the training as well, when the trainees are already back in their work environments. These evaluations not only take longer but can also be more complicated, as the impact of the training has to be differentiated from other possible impacts on the trainees during the surveyed period. Level 5 analysis can only be implemented when the investments in the trainings and also the organisational benefits can be expressed in financial terms. This possibility decreases with the

increasing complexity of the theme. Not surprisingly, 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)

1.2. Prevailing practice in training evaluation

The above-cited OECD report concluded that most training providers share only output results. (OECD 2013: 10) At the same time, my experience[2] in some of the surveyed countries is that it is also a general practice to distribute a simple questionnaire at the end of training programmes to survey participants' satisfaction levels: providers ask participants for their opinion of the location and catering and also the trainer and training. The result of the OECD survey indicates the lack of willingness to share the results of such surveys with external parties. Sadly, my own experience and anecdotal evidence from colleagues also proves that the results are often not shared with the staff who were involved in the programme, either. This means that possible learning from the surveys is not exploited and raises the question whether these survey are perceived by organisers as mere formalities.

There is a second problem with the actual practice. While participants' satisfaction is not negligible, satisfaction surveys are necessarily subjective, and as I will discuss later, can produce different results to those of other more objective methods. In view of this it is problematic that in my country, Hungary, similarly to other countries in our region, satisfaction surveys are the highest level of evaluation. The most common explanation training providers give for the absence of higher level effectiveness assessment of training courses appears to be practical: they claim that it would be complicated and costly and it is often beyond their organisational capacity. More sophisticated answers are that larger training programmes are financed by projects that are short-term and cannot accommodate longer impact assessments, and the system of budget allocation does not allow longer term planning either. The key question our research tried to answer is whether the absence of evaluation is really a question of capacity and context, or that it rather depend on motivation and competence.

---

[2] I have worked and participated in training projects in 14 out of the 22 countries.

1.3. The goal of our research

The starting point of our research was that recording the money spent on the training, or satisfaction surveys (level 1) alone do not prove that the use of public money is producing the desired effects. The aim of the survey presented in this paper was to demonstrate that project financing and limited organisational capabilities should not be used as an excuse for the absence of evaluating effectiveness. We wanted to show that information on the impact of training can be produced even with the use of simple questionnaires, the management of which is identical to the usual satisfaction survey questionnaires. By simply changing the questions, the analysis becomes more insightful. Formulating different questions is not prohibitive, whether in cost, time or organisational 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 gathered from the data which can support further development of training or organisational methods.

2. The training programme we evaluated

The survey presented in the paper was carried out concerning a large "Public Ethics and Integrity Training Programme" that was implemented in Hungary in 2013 and 2014. My task was to design introductory training on the integrity approach to corruption prevention for civil servants so far unfamiliar with both the term and the approach. The project parameters were given. Two types of training courses had to be designed, one with 8 contact hours (one day long), the other with 20 contact hours (two and a half days long). The target group was Hungarian civil servants from the whole public administration, with staff-level civil servants for the 8 hours of training and civil servants in higher positions for the 20 hours of training. A similar design had to be developed for the two forms, with standard hand-outs and schedules, and trainers had to be trained to deliver possibly identical curricula.

During the curriculum design, we decided that we wanted to convey not only cognitive knowledge but we equally wanted to affect participants' attitudes towards the fight against corruption. Our objective was to

initiate reflection on corrupt and inappropriate organisational practices and introduce the concept and approach of integrity management. We also wanted to break the learned helplessness of participants, build trust in the possibility of positive changes, show the role of well-organised and functioning public administration in curbing corruption, and with this to encourage participants to understand their own possible role in anti-corruption and take responsibility. Through the change in knowledge and attitudes, we hoped to influence participants' behaviour in their work. The design was based on a participant-centred, experiential method that contained ample room for dialogic interactions, self- and group reflections and facilitated peer learning. We used a few, clear conceptual frames, with visualisations used to support their transmission. Experiential method, focused cognitive content and visuals were aimed at making deep, memorable imprints.[3]

The training sessions were delivered to some 10.000 civil servants by 26 trainers who were trained to deliver the methodology drawn up. The average size of training groups was 18.4 participants /group in the 8-hour training courses and 15.2 participants /group in the 20 hour ones.

3. The research

The survey was neither part of the original project, nor of the training design; it was organised as an additional activity. The idea to change the standard satisfaction survey with more substantive questions was raised only shortly before the pilot training sessions began. The decision was to assemble questions for two research topics in one questionnaire, which was completed by participants at the start and again the end of the training course. 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. The research could be implemented because it did not demand any extra effort or skill from the organisers.

---

[3] For a more detailed description of the approach and method: Pallai 2014 and Pallai 2015

3.1. The objectives of the research

We set three objectives for the research:

- to validate the effectiveness of the training through the measurement of

    o change in participants' knowledge of integrity and anticorruption

    o change in participants' attitudes towards anticorruption

- to collect information for providers to learn how they could improve the method and the curriculum

- to set an example for the viability of effectiveness assessment using simple methods.

3.2. The research method

A quasi-experimental, non-randomised pre-post quantitative evaluation research design (Powell 2006: 110) was implemented. Participants were assigned to the 8 or 20 hour long training based on their organisational position. Three hundred and sixty-three 8-hour and 44 20-hour training courses were held by 26 trainers. Twenty-four trainers were involved in the shorter training courses and 8 in the longer ones; 6 of them led both forms of training. All trainers had to use the same method and materials. Trainers received unusually sound preparation for producing similar results. Acknowledging this fact, one can assume that the main differences between trainers' performances can be attributed both to systematic differences between groups in their initial attitudes and the individual competences of trainers.

Questionnaires were filled out by 6,692 participants in shorter and the 670 participants in longer training courses before the training course and an identical one right after it. Respondents were told to choose and write the same pseudonym on both questionnaires. This technique ensured the possibility of matching precisely the pre and post surveys and tracking the changes in responses, even at an individual level, along each item.

The survey was based on the responses to eight statements about corruption and anticorruption. In both (pre- and post-training) questionnaires these eight statements were included with exactly the same formulation. 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 respondent (e.g. gender, type of organisation, length of employment, etc.) and respondents' organisational commitment. Of the eight relevant statements, four were related to attitudes and four to cognitive knowledge of (anti)corruption (Table 1). Participants were asked to express the level of their agreement or disagreement with each statement on a 1-to-5 Likert scale.

As we discussed above, the training sessions were most participants' first encounters with the concept of public integrity and integrity management. The objective of the training was to initiate reflection on corrupt practices, break learned helplessness and build trust in the possibility of change, and show the role of well-organised and functioning public administration in curbing corruption and, with this, encouraging participants to understand their own role in anticorruption and take responsibility. This is why the attitude statements revolve around apathy, helplessness and trust, and the knowledge statements check whether participants understand better that not only legal instruments but transparency, organisation and effectiveness are also key instruments in fighting corruption.

Table 1: How do you agree with the following statements on a 1-to-5 scale?

Statements related to attitudes towards corruption and anticorruption	Statements related to knowledge of corruption and anticorruption
A1. The corruption experienced in this country is no particular cause for concern, because it is an inherent feature of transformation. A2. Corruption is as old as mankind and not much should be done to fight it. A3. In Hungary, corruption has assumed such proportions that fighting it has become impossible. 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.	K1. Corruption should primarily be fought using legal instruments. K2. Corruption can be fought the most effectively through transparency. K3. The best remedy for corruption is fast and efficient administration. K4. Well organised public administration can significantly reduce external attempts at corruption.

Because the participants in the 8 and 20-hour training courses differed not just in the length of their programme, but also in the hierarchical position in their organisation, it is impossible to separate these two effects on the changes in knowledge and attitudes. Hence, the analysis had to treat the results of the two kinds of training courses independently from each other and present the results separately.

4. The results

It has already been mentioned that two publications have discussed the results of this research. One was a detailed data analysis conducted by Anikó Gregor, a statistician. (Gregor 2015) The other was a paper presented at the EGPA annual Conference in 2015. (Pallai-Gregor 2015) Links to both papers are in the references section of this paper. The charts in this paper are the work of Anikó Gregor and most are selected from the two previous publications. Due to the limits set for the length of this paper, details of the data analysis and technical information will not be discussed here; only references will be given for the other papers. Regarding the part of the research that validated the effectiveness of the training methods, only as much that is needed for the understanding of the new points made in this paper will be repeated here. The omission of the technical data allows more detail of the questions related to the learning produced: what we can learn from the results for training design and what the further directions for research could be.

4.1. Validation through the average of opinions

The results of the evaluation of the data have proved the effectiveness of the training: the comparison of participants' pre- and post-training responses showed statistically significant change in participants' knowledge and attitudes during both versions of the training, with the average changes having occurred almost exclusively in the targeted direction.[4]

Figure 1 and Figure 2 present the scatter plot of before-after average scores of opinion changes, with regard to the 8 and 20 hour designs. Rotated-scale versions of A1, A2, A3 and K1 statements indicated by an additional letter, "R," were used on the charts in order to have all items with a high grade (grade 5) as expected. The blue line represents the hypothetical zero difference between before and after average scores.
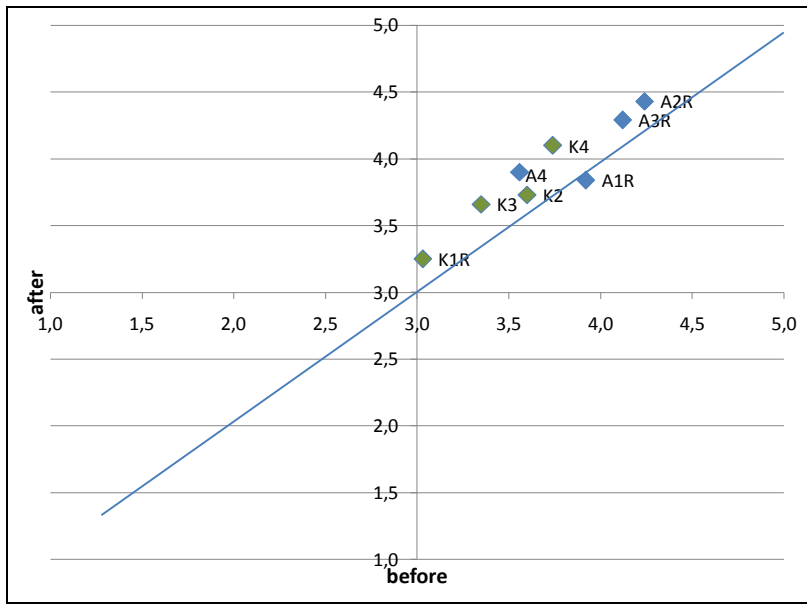
---

[4] for detailed results, see Pallai-Gregor 2015: Annex 3

Figure 1. Average scores of opinion changes on a 1-to-5 scale after 8 hours training
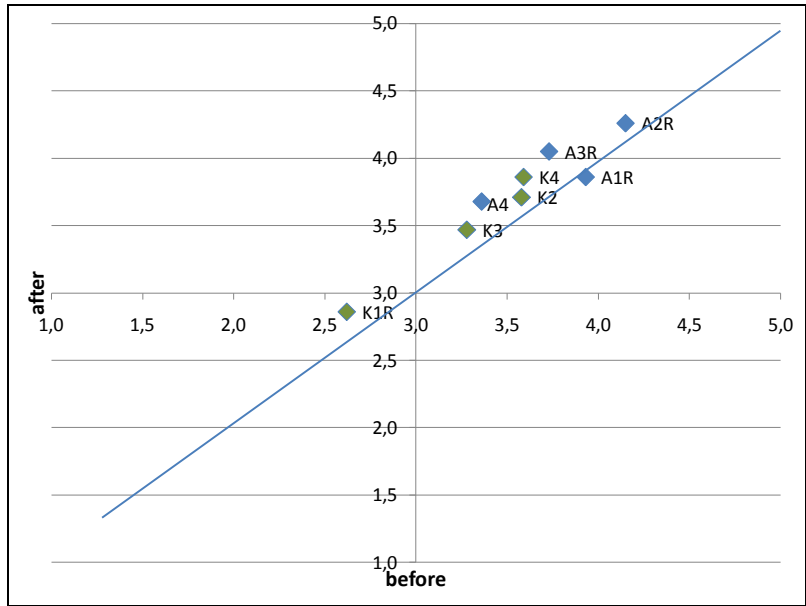
Source: Pallai-Gregor 2015

Figure 2. Average scores of opinion changes on a 1-to-5 scale after 20 hours training

Source: Pallai-Gregor 2015

Figure 1 and 2 show that, in both training types, the strongest changes are regarding the statement 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 statement K4 (Well organised public administration can significantly reduce external attempts at corruption.). These are important results because A4 is the most important attitude statement since it reflects trust in possible positive change, and K4 is the most substantive knowledge statement regarding the concept of integrity management.

Figure 1 and 2 also show that all changes happened in the desired direction except for statement A1, where a very modest average change occurred in the opposite direction. We will examine in details in part 4.2.3. of the paper how could this happen.

Before getting back to the discussion of A1, we discuss another important question. In statistical terms, all changes are moderate but what does this mean? Is it a good result to accomplish a statistically moderate average change regarding these statements in the opinion of civil servants after 8 or 20 hours of training? Does it matter what were the initial positions of participants? The last two questions are the easier to answer. Gregor executed an OLS linear regression analysis on the relationship between the average initial scores and the absolute value of changes, which revealed that, on average, the lower initial average score someone had, the higher amount of absolute change they showed in their opinions (Pallai-Gregor 2015: 25)[5] Although it was already mentioned that, because of the differences in target group and length of training sessions, we must treat the two training courses as separate training programme, it is still interesting to compare their results. Figure 3 shows that, on average, a similar degree of changes occurred in the opinions of participants during the 8 and 20 hour training courses. This result that opinions from lower average levels changed faster is in accordance with the previously mentioned result on the impact of initial scores on the rate of change. However, because the relationship Gregor identified was not as

---

[5] Detailed information is in Pallai-Gregor 2015 Annex 4

strong as the difference here, another question can also be raised: can shorter training courses be more effective? Unfortunately our research cannot give an answer to this question but, because of its importance, further research should deal with it.

Figure 3. Average scores of opinion changes
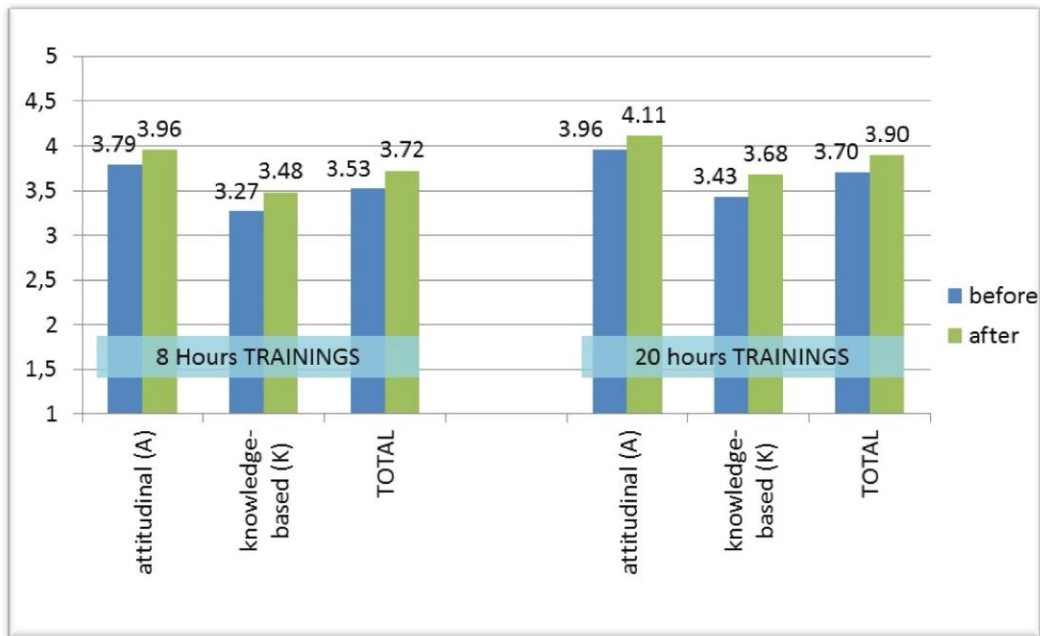
Source: Gregor 2015

Figure 3 also shows that the 8 hour training courses elevated staff-level participants' average attitude and knowledge to the initial level of the leaders. This may indicate that, in our case, the statistically moderate change is a good result. However, for a more definitive claim on this, more research and comparative results would be needed.

4.2. Learning from the analysis at the level of individual participants

The pseudonyms participants used on the questionnaires ensured the possibility of tracking the changes of responses, even at an individual level, along each item.

4.2.3. Analysis of participants' learning

The change of opinion among a randomly selected 10% of participants are depicted in the following scatter plots.

Figure 4. Scatter plots of impacts on individuals and histograms of average impacts after the 20 hour training course
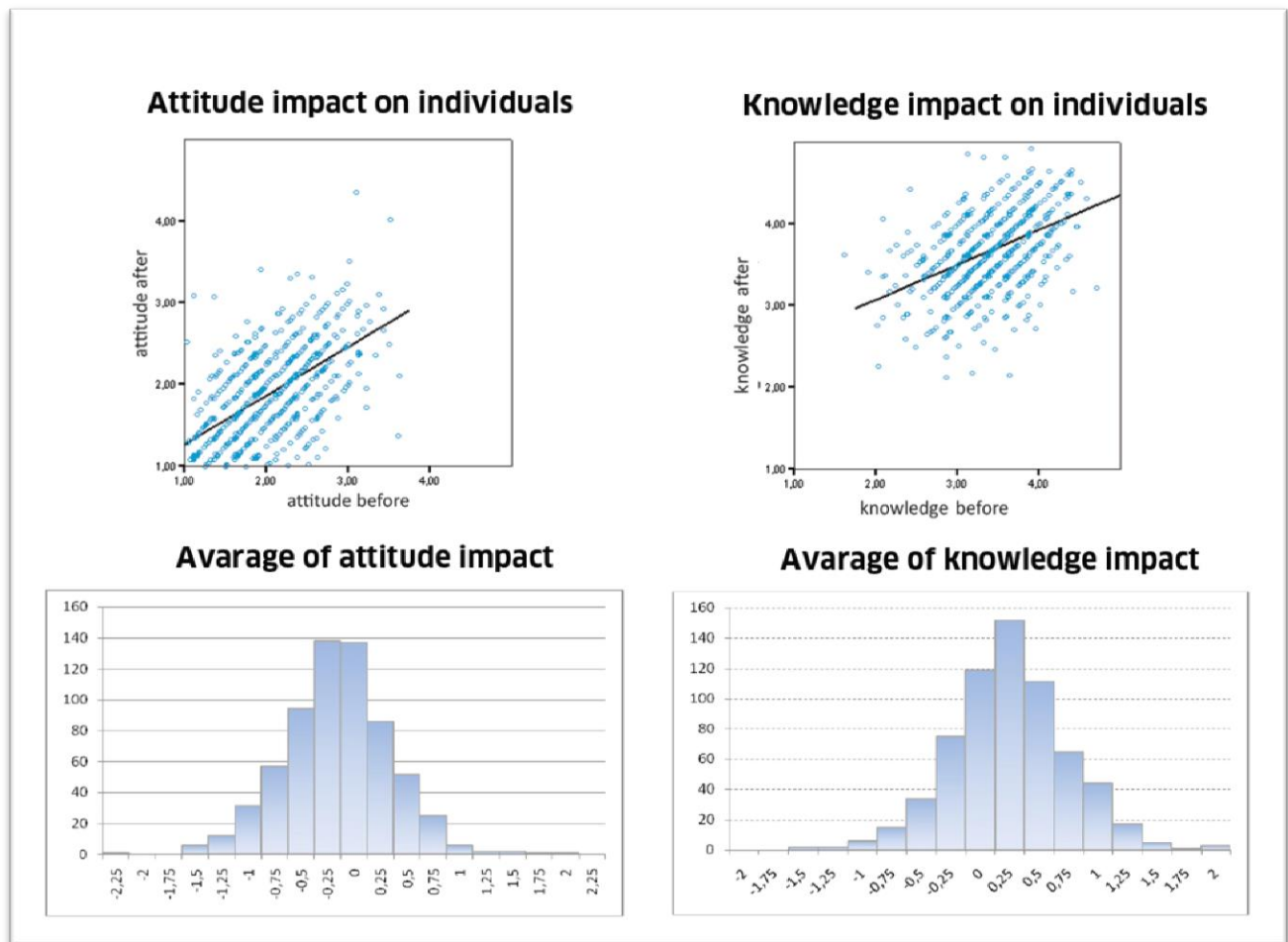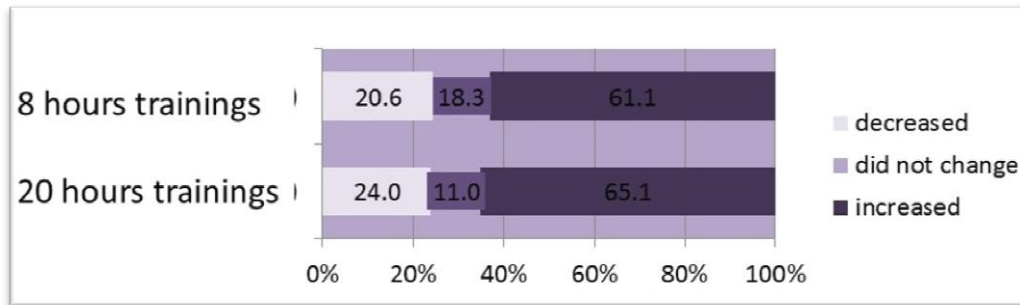
Figure 4 shows that while the majority moved in the desired direction, there were participants who did not change opinion or moved in the opposite direction. Figure 5 summarises the results for both training courses.

Figure 5. Respondent categories based on changes in total achievement (%)

Trainers with a minimal amount of humility know that they cannot reach all participants to the same degree and some adverse reactions can also be produced; still, the share of adverse moves within the positive average results that the survey revealed is remarkable. In order to understand why these moves could happen, similar analysis was carried out for each question. (Gregor 2015) Figure 6 shows the results for the two statements where the average change was the strongest. On average, opinions „nudge" in the desired direction but the histograms on the right shows that, even with regard to these statements, some participants change their opinions in the opposite direction. The positive average results from the larger number of those who change opinion in the desired direction than the number of those ones who change opinion in the opposite direction.
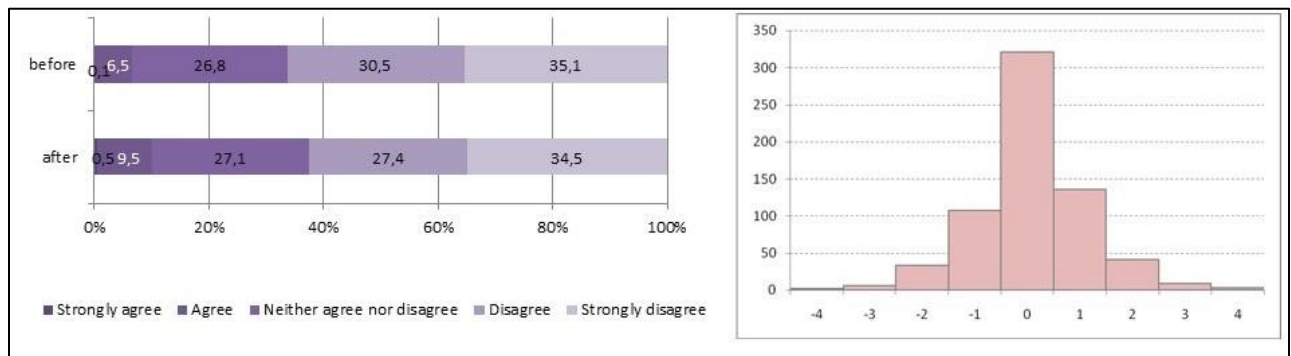
Figure 6. Opinion change on A4 and K4 statements during 20 hour training course

Source: Gregor 2015

Figure 7 shows the same types of charts regarding the A1, where the increasing agreement with the statement is a change in the wrong direction.

Figure 7. Opinion change on A1 statement after 20 hours training

Source: Gregor 2015

Here the histogram shows that the weak average opinion change in the wrong direction is the result of the weaker impact on all participants than is the case of the previously discussed A4 and K4 statements and the wrong impact on a larger portion of participants than in the previous cases.

Because of the relatively large proportion of those who changed their opinion in the wrong direction, we attempted to find out more about this group. We were interested to find some common characteristics of them and see whether this group have special needs. It was also important to see how and why we failed with them, and whether the content, method or delivery of the training courses could be the main source of this problem.

Regarding the first questions, we wanted to identify the common characteristics of this group. Our first hypothesis was that commitment to work and their organisation might positively affect participants' openness to learn about integrity and the change in their opinions. The analysis of the personal data collected by the other research in the first questionnaire included some questions related to general commitment but there was no correlation between the wrong movers and their commitment or the initial level of their knowledge or attitude, either. For the participants in the 20 hours of training, a weak but significant positive correlation was found with leaders who had worked for their organisations for less than 2 years, and for the 8-hour training with male non-executives and non-executives heading for pension.[6] Our research did not give more precise answers but this question is so important that further research should reveal more in this field.

---

[6] 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 to pension (chi-square = 10.491, df=2, p=0.005<0.05, Cramer's V = 0.04); male non executives (chi-square=9.797, df=2, p=0.007<0.05, Cramer's V = 0.04)

4.2.3. Analysis of trainers' effectiveness

The earlier publications on this research have already discussed how the data allowed detailed comparison of trainers' effectiveness to be carried out. (Gregor 2015 and Pallai-Gregor 2015: 26) We measured trainers' effectiveness as the average change in participants' attitudes and knowledge with the possible effect of initial average levels of participants' attitude and knowledge in their groups controlled. Among the trainers of the 8-hour training courses, a significant correlation (r= 0.73) was found between the performance of trainers in trainees' attitude and knowledge change (Figure 8) but no similar relationship was found between the two rankings with regard to trainers of the 20 hours sessions. (Figure 9)

Figure 8. Ranking of trainers on the 8 hour trainings
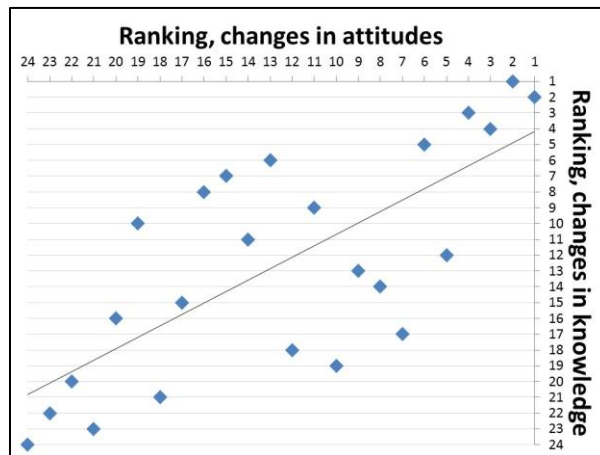
Source: Pallai-Gregor 2015

Figure 9. Ranking of trainers on the 20 hours training
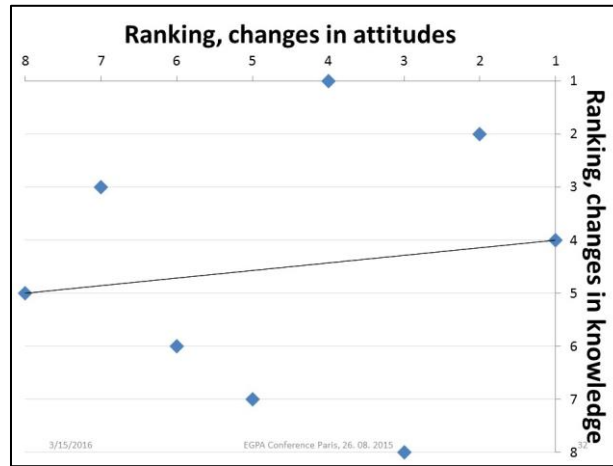

Source: Pallai-Gregor 2015

It is in itself an interesting research result that, during the longer training courses, the strengths of impact of some trainers may shift pronouncedly towards either attitude or knowledge. It is an indication that should be assessed further because, according to the experiential method, a balance of attitude and knowledge change would be the most desirable. However, this result is important not only for the research. It has potential practical use as well. Individual trainers should also reflect on these results as it indicates where they have to improve and providers should also reflect on what kind of further training they should provide for their trainers.

The comparison of trainers' performance brought more practice-relevant results. The research was initiated to show that important information can be gained even from simple effectiveness evaluations: considerably more than from the more prevalent satisfaction surveys. Although a good comparison between the results of satisfaction and effectiveness surveys would be important, during this research, unfortunately, we did not ask participants about their satisfaction with the training and the trainer. Nevertheless another opportunity for such comparison opened because 18 trainers participated in a project right after our survey in which they held nearly identical training sessions for participants from identical

target groups as during the 8 hours training courses. Those participants had to fill out an evaluation questionnaire that surveyed only their satisfaction. An important difference between the results of our survey and the satisfaction survey was that in the satisfaction survey all trainers were graded above 3,85 (on a 0-4 scale) – even the trainers who, by our survey, could produce only minimal or zero learning impact. (The high average figure explains why satisfaction surveys are often called "smiley sheets".)

We calculated the average satisfaction score of each of 18 trainers and we compared their satisfaction ranking and their achievement ranking. (Figure 10) We expected a weak positive correlation between someone's satisfaction ranking and effectiveness ranking. The results proved that trainers having a greater impact on participants' attitude and knowledge (total effectiveness) were evaluated better by respondents from the other study ($r=0.64$, $p=0.002<0.05$). However, if the best 50% of trainers were selected on the basis of satisfaction scores, 30% of the most effective trainers would be lost. This result also speaks for the importance of effectiveness surveys because it shows that "pleasing" is not necessarily the same as "educating" and selection on the basis of satisfaction survey results would lead to a suboptimal choice of trainers regarding effectiveness.

Figure 10. Comparison of satisfaction and effectiveness ranking of trainers during 8 hours training
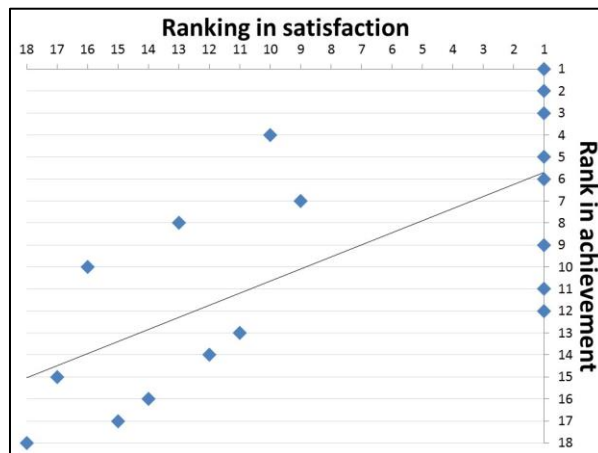
Source: Gregor 2015

At this point we have to turn back to the problematic A1 statement. The second question I raised regarding statement A1 was whether the content, method or delivery of the training could also be the source of the problem. The detailed analysis of trainers' results showed, in the case of A1, that the results of 6 out of the 24 trainers on the shorter training course were positive (it happens in the targeted direction), 6 had a negative change and 12 had no significant effect. With regard to the longer training course, 1 out of 8 trainers achieved the targeted opinion change, 1 did the opposite, while the remaining 6 trainers did not cause any change. This result, together with the result of the previous examination on participants' changes of opinion, indicates that in order to avoid the adverse result regarding the A1 statement probably no major changes in the training content are necessary but only fine tuning of the activities and messages related to this statement. Additionally, because some of the trainers could achieve positive results and a large portion of them could not, trainers' education may also play a role. This brings us back to our starting point: effectiveness analysis is far better guidance for trainer selection and education than satisfaction surveys.

5. Questions for further research

Many questions have already been raised for further research in the previous sections of the paper. When we discussed that a similar degree of average change in opinions happened during the 8 and the 20 hour training courses, we already touched upon the role of length on impact. However, because we had to treat the two training courses as separate (and the participants in each were drawn from different staff grades) we could not draw and infer from this result that longer training courses would be less cost-effective. Nevertheless, this is an important question that further, targeted research should investigate.

It is a remarkable result that this research showed that, within the positive average results, significant groups may change opinion in an adverse way. It was already mentioned that the characteristics of such groups and the causes of the adverse reactions should be investigated with further research. Here I add

one more issue: it could also be investigated whether other teaching methods (e.g. traditional, frontal education) tend to bring more uniform results.

The changes in opinions after our training courses were statistically significant but moderate. Statistical significance is clearly relevant for us. At the same time it is an open question what a statistically moderate change means in our case. Regarding the target groups and length of the training courses, is a statistically moderate change a good result, or is it a weak impact in the given conditions? Can a stronger impact be expected on these target groups in these time frames? The fact that, during the 8 hour training, staff-level participants reached the initial level of knowledge and attitude of senior civil servants is indicative of the strengths of the method. Nevertheless, more definite answers to these questions can only be given on the basis of comparative research. We intend to replicate our inquiry soon on training implemented with curricula and methods that have been updated on the basis of the findings of this research. The results will be the first comparable figures. At the same time, we hope that either other training providers will also implement similar evaluations and will share results or a comparative project can be launched.

6. References

Gregor A 2015: Training effectiveness analysis. Research report prepared for the Centre for Excellence in Integrity at the National University for Public Service, Budapest. http://www.pallai.hu/wp-content/uploads/2016/03/Gregor-A-2015-eng.pdf

Horton S 2007: Evaluation of Leadership Development and Training in the British Senior Civil Service: the Search for the Holy Grail. Paper presented at Leading the Future of the Public Sector: The Third Transatlantic Dialogue University of Delaware, Newark, Delaware, USA

Kirkpatrick D L and Kirkpatrick J L 2006: *Evaluating Training Programs*. (3rd ed) San Francisco: Berrett-Koehler.

OECD 2013: Ethics training for Public Officials. Report prepared by ACN, SIGMA and OECD Public Sector Integrity Network.

Pallai K, Gregor A 2015: Assessment of Effectiveness of Public Integrity Training Workshops for Civil Servants - a case study. paper presented at *EGPA Conference*, Toulouse, 26. August 2015 in the Permanent Study Group IX: Teaching Public Administration. pp. 42. http://www.pallai.hu/wp-content/uploads/2010/11/EGPA-paperafter-conference.pdf

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. http://www.pallai.hu/wp-content/uploads/2016/03/Pallai-2014-Design-considerations-Pro-Publico-Bono.pdf

Pallai K 2015: Integritásoktatás a közszolgálati képzésben – építkező oktatás a fejlődő integritásért. in: Dargay, Eszter –Juhász, Lilla Mária eds: Antikorrupció és Integritás. (pp. 93-108.) Budapest: Nemzeti Közszolgálati Egyetem. http://www.pallai.hu/wp-content/uploads/2010/11/Dr-Pallai-Katalin-integrit%C3%A1soktat%C3%A1s.pdf

Phillips J, Phillips P, Stone, R and Burkett H 2007: *The ROI Field Book: Strategies for Implementing ROI in HR and Training*. Massachusetts: Butterworth-Heinemann/Elsevier

Powell R R 2006: Evaluation Research: An Overview. *Library Trends* 55(1): 102-120.