

EVALUATION OF THE HUNGARIAN
GENERAL CIVIL SERVICE TRAINING PROGRAM

György Gajduschek and György Hajnal

Discussion Papers, No. 16
Local Government and Public Service Reform Initiative

OPEN SOCIETY INSTITUTE

**EVALUATION OF THE HUNGARIAN
GENERAL CIVIL SERVICE TRAINING PROGRAM**

György Gajdusчек and György Hajnal

Discussion Papers, No.16

Published in 2000 by the Local Government and Public Service Reform Initiative

Open Society Institute

Nádor utca 11

1051 Budapest

Hungary

tel: (36-1) 327-3104

fax: (36-1) 327-3105

e-mail: lgprog@osi.hu

listserve: lgj-announce-subscribe@egroups.com

www.osi.hu/ligi

ISSN 1417-4855

Distributed by the Network of Institutes and Schools of Public Administration
in Central and Eastern Europe

Hanulova 5/B

840 02 Bratislava 42

Slovakia

tel/fax: (421-7) 785-357

e-mail: nispa@nispa.sk

www.nispa.sk

Produced by

Arktisz Studio

Budapest

INTRODUCTION TO THE SERIES

The development of democratic and effective government at subnational levels remains one of the central tasks of transition in Central and Eastern Europe and the former Soviet Union. The sharing of expertise between countries can contribute significantly to the reform process in the region. Pursuing this goal, the Local Government and Public Service Reform Initiative (LGI) has launched a series of discussion papers, which will be distributed widely throughout Central and Eastern Europe.

The series will report the findings of projects supported by LGI and will include papers written by authors who are not LGI grant recipients. LGI offers assistance for the translation of the papers into the national languages of the region. The opinions presented in the papers are those of the authors and do not necessarily represent the views of the Local Government and Public Service Reform Initiative.

György Gajdusчек is an MA in Public Administration (Rockefeller College, State University of New York) and an MA in Sociology (ELTE University, Budapest). He has been working for the Hungarian Institute of Public Administration as a researcher since 1987. As a researcher he took a significant part in the modernization program of the Hungarian public administration both on local and central level, actively participated in various comparative research projects and published several papers in Hungarian and English. The main fields of his academic interest include the theory of public organizations, decision-making in government and public program evaluation. Comments can be addressed to: gajdusчек@hotmail.com.

György Hajnal is an MA in Economics, Public Finance concentration (Budapest University of Economics). He has been working for the Hungarian Institute of Public Administration as a researcher since 1996. He took a significant part in the study and introduction of the methods of new public management into the Hungarian public administration. He took part in various international research projects and published several papers in Hungarian and English. The main fields of his academic interest include public policy, and the theory of public management. Comments can be addressed to: gyorgy.hajnal@ella.hu.

Abstract

From 1996 to 1998 the authors conducted a research project with the generous support of the Hungarian Institute of Public Administration. The project's goal was to measure the efficacy of the most important element of Hungarian civil service training: the Basic Public Administration Examination System. The project utilized the public program evaluation method, a technique widely used to assess quantitatively the implementation of public administration policies.

The argumentation of this report runs in two parallel lines. First, it describes the broader framework and the operational details of the Basic Public Administration Examination System and provides an account of the project's findings and recommendations. Second, the report introduces the basic concepts and working methodology of program evaluation and summarizes theoretical and practical problems that emerged in the course of the evaluation.

1. The Program

Civil Service Act No. XXIII (1992) introduced the Basic Public Administration Examination System (BPAES) in Hungary. The BPAES was implemented in practice only after the Cabinet issued Government Decree No. 51/1993 (III.31), which contains the detailed regulations of the system.

Legislation refers to the BPAES in the framework of the civil service career system. According to regulations, all category I and II civil servants except those who hold a degree in law, economics or public administration are obliged to pass the basic exam during the first two years of their employment. Category I and II officials include those who may make legally binding decisions as opposed to category III civil servants, who are administrative employees (typists, etc.), and category IV employees, who are physical laborers. Of the 106 thousand civil servants in Hungary, it is estimated that approximately 65 thousand have passed the BPAES. The number of examinations carried out annually has varied between 4.5 thousand and 6.9 thousand over the last several years.

Exams for civil servants in the central government are organized by the Hungarian Institute of Public Administration, and the twenty county administration offices are responsible for the exams for civil servants employed by the local and territorial administration. The Hungarian Institute of Public Administration and the county administration offices typically offer two- to three-day preparation courses for the exam; although attendance is not mandatory, approximately ninety percent of all examinees participate. The employers of most examinees require attendance at the preparatory courses, pay the course and exam fees and provide study leaves for the training period and the day of the exam.¹ Courses usually are held from three days to two weeks before the exams.

There is a general opinion that participating in a training course is not sufficient preparation for passing the exam; independent study is also necessary, for which examinees are provided a course book. The one hundred fifty-page publication contains information that the exam-

inee is expected to know in order to pass the exam. In addition, candidates must be thoroughly familiar with relevant laws (the Constitution, the Law on Administrative Procedure, the Civil Service Act and the Act on Local Self-Governments, among others).

Exams are held in Budapest and in the nineteen county center cities five times per year. The examination consists of written and verbal sections, both of which are evaluated and graded by a three-member examination committee. Its content concentrates on the current laws of Hungary and does not test specific skills (e.g., communication, foreign language, management, etc.) or customer service orientation.

2. Method of Investigation

2.1. Public Program Evaluation

BPAES managers and designers possessed some of the results of the program (for both training and the exam) before the current evaluation took place. The Ministry of the Interior collected examination results, and BPAES managers had an impression of trainee evaluation regarding the program's usefulness. However, the aim of this report was to gather more reliable and more detailed information on a wider range of issues. While the managers were interested mainly in the *process* of the BPAES, this project also focused on its long-term effects, including questions such as (1) if the BPAES made a positive difference in Hungarian public administration and for the public, (2) and if so, the extent of the difference and "size" of these positive effects. As this research project was concerned with the *real impact* of the program, reliable in-depth information was needed.

The project also attempted to minimize subjectivity through two means:

- controlling subjective opinion by balancing the views of BPAES managers and trainers with those of trainees and the trainees' supervisors;
- collecting information on facts rather than on opinions whenever possible.

Both the philosophy and the methodology of *public program evaluation* fit the project's goals. Public program evaluation started to spread in the United States and other developed countries in the 1960s. According to this approach,² a public program is a set of rational activities carried out by public organizations to achieve specific social goals. The primary aim of this method is to *dissect* the program into simpler—and thus empirically measurable—elements. These features are joined in a series of logical and causal relationships collectively known as the *program model*. Modeling a public program in this manner requires:

- *definition of program objective(s)* (i.e., the broader social goals the program serves);
- *identification of the program's substantial elements* (typically the activities of the program provider) affecting the outcomes of the program;
- *determination of the relationships* between these program elements and the manner in which they collectively achieve the stated program objectives.

Public program evaluation is designed to measure if program objectives have been met, and if so, to what extent. Evaluation research tests two hypotheses: (1) if there are any effects potentially attributable to the program and (2) if identifiable effects are indeed consequences of the program or of other factors.

In order to define the program objectives and model, the legal texts regulating BPAES were analyzed. In addition, several interviews were conducted with officials of the Hungarian Institute of Public Administration and of the twenty county administrative offices who took part in formulating BPAES regulations or served as program managers. The *overall objective* of the program was identified as increasing the administrative and legal performance of public administration. The basic elements of the *program model* are as follows:

- individual study and noncompulsory training, assisted by the provision of a course book;

- examination of each participant; failing to pass the exam three times leads to termination of employment;
- increase in the relevant administrative and legal skills levels of examinees;
- retention of some portion of familiarity with subject matter after the examination;
- utilization of increased familiarity with material covered on the exam in the course of daily administrative work;
- improvement of legal and administrative performance of public service as a whole.³

A simplified BPAES program model is depicted in figure 1.

Once the program model was (re)constructed, the next step was to determine the method by which the overall impact of the program could be ascertained. It was difficult to *measure* if the legal and administrative performance of public service as a whole had increased. Even if improved performance was identified, it was difficult to determine to what extent this was due to the impact of BPAES.

A clear program model assisted with this goal a great deal by depicting a cause-effect chain. If one piece of the chain was missing, the final intended outcomes of the program would not be achieved (at least not as a result of the program). The increased legal and administrative performance of public service would not be an effect of the program if general familiarity with the exam material (1) did not increase as a result of the training process, (2) was irrelevant or (3) was not retained or utilized in the workplace.

The “relevance of knowledge” and the “workplace environment” categories seem difficult to measure. However, the increase in familiarity with material covered on the examination and its retention is relatively easy to measure. This may provide relevant information on program effectiveness.

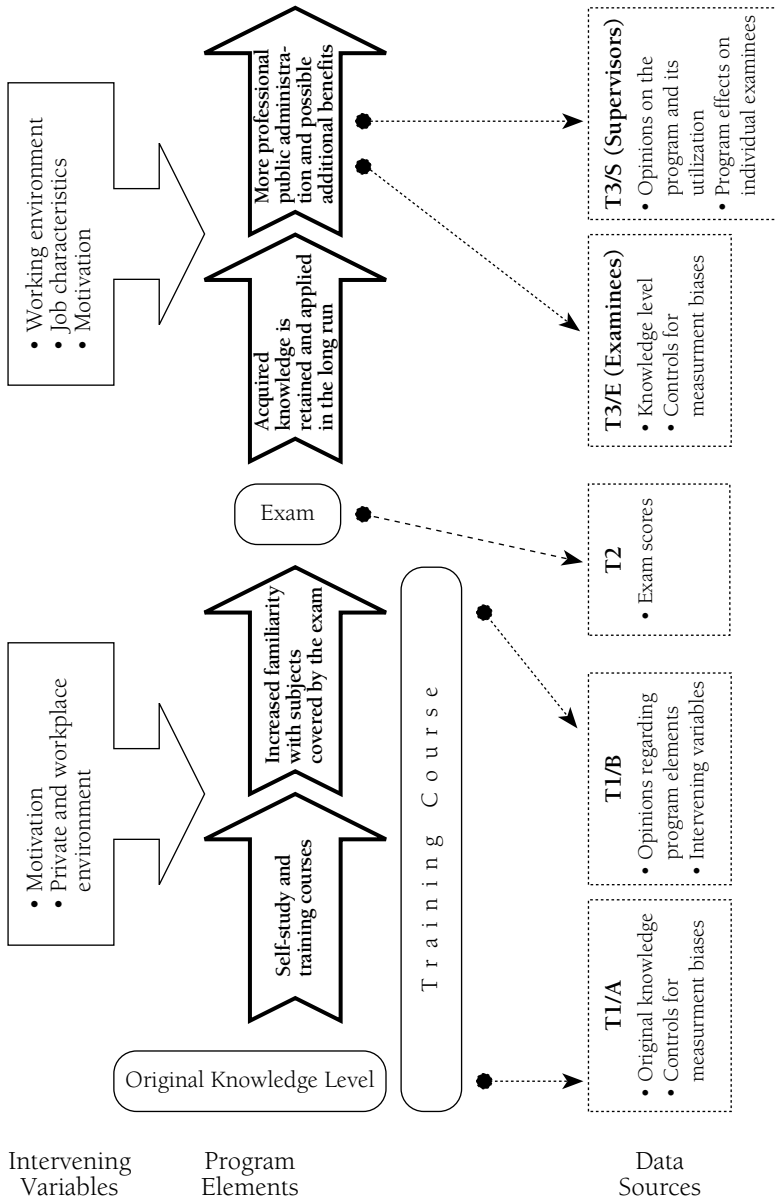


Figure 1. Program Model and the Measurement Process

2.2. Measurement

The fact that some elements are more measurable than others influenced the methodology applied in evaluating the efficacy of BPAES. Other circumstances, however, had to be taken into consideration when formulating the research project, such as the fact that the exam was obligatory, requiring a “quasi-experimental” design due to the lack of a potential control group. Furthermore, limited funding was available; the research budget was approximately USD 2,500 plus an estimated USD 6,000 for administrative support the salaries of two researchers. On the other hand, a relatively longer period of time was available to carry out the research.

These conditions led the research team to adopt a *longitudinal, interrupted time series research design*. In other words, a set of randomly selected examinees was tracked. The first survey took place before officials began preparation for the exam, and the last survey was conducted approximately fifteen months afterwards.⁴

The research team chose the May 1997 exam period and randomly selected ten of the twenty counties in Hungary. All training course participants were asked to complete a test identical to the written exam at the beginning of the training.⁵ In addition, questions regarding previous educational activities were asked. Through evaluation of these written tests, data was gathered on the original level of familiarity with the test material before any significant training took place. A comparison of the original score and the examination results provided information on the immediate increase in familiarity with the test material as a result of the BPAES.

At the end of the training courses, participants completed another questionnaire on the subjective assessment of the utility of BPAES, individual learning strategies and techniques and attitudes towards BPAES, including some control questions regarding the consistency and honesty of the responses.

Subsequently, information on exam results was collected from the official program records compiled and provided by the exam organizers. Approximately fifteen months after the exam, the examinees completed a written test and a brief questionnaire. The test was similar to the one offered at the time of the first measurement. The questionnaire examined issues such as whether or not respondents used the material learned through the BPAES training process and if they had used the course book since the exam. At the same time the examinees' supervisors were asked for their views on the utility of BPAES.

Table 1. Information Sources and the Number of Respondents

Reference	Timing	Content	Number of Respondents
T1/A	Beginning of training course	Written test identical to the written portion of the exam plus additional questions (e.g., previous training individual study for the BPAES)	401
TA/B	End of training course	Questionnaire on the usefulness of BPAES, individual study strategies and methods	434
T2/Exam	Date of exam	Exam results obtained from the organizers on all examinees (not only those who participated in the training course)	622
T3/E	Fifteen months after exam	Written test identical to the written portion of the exam, plus a brief questionnaire on the utility and utilization of material learned during exam preparation	189 (71) ⁶
T3/S	Fifteen months after exam	Structured interview with supervisors on the subjective assessment of the utility and utilization of BPAES material in everyday operations	105

By the time of the T3 testing and surveys, the number of respondents had decreased significantly. As indicated above this is an inher-

ent characteristic of longitudinal surveys. The number of respondents at the T1 stage was higher than at T3, as several groups of former respondents were not approached during the latter stage, including:

- respondents from two counties (the largest) in which widespread cheating was noted in taking the T1/A tests, resulting in a loss of one hundred nineteen respondents;
- those who were not accessible through their previous employers, leading to the loss of another ninety-three respondents.⁷

As a result, the T3/E stage included one hundred eighty-nine respondents.

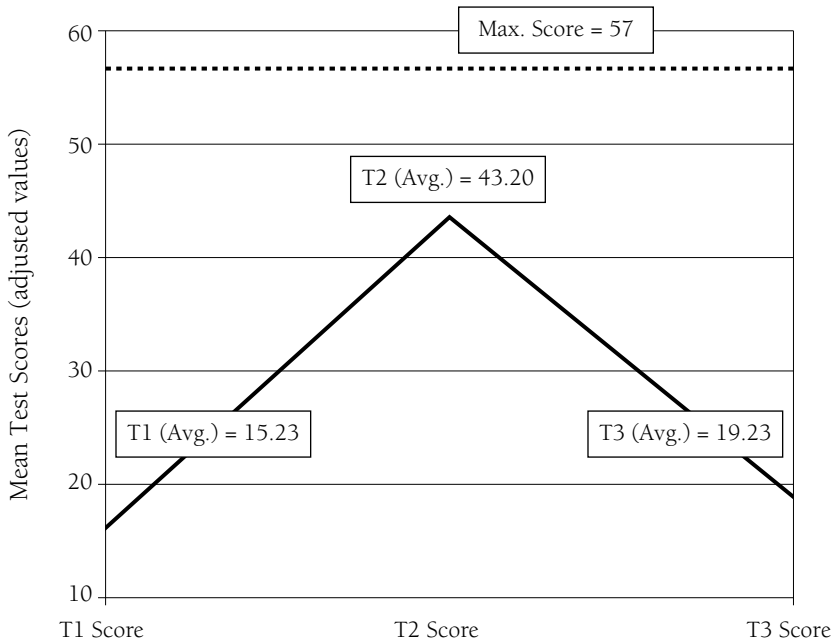
3. Quantitative Analysis of the Program's Effectiveness

3.3. Test Results at the T1, T2 and T3 Stages

As indicated earlier, written exams similar to the official version were used to measure examinees' familiarity with test material at the T1 and T3 stages (in addition to the official "measurement" at T2). Figure 2 depicts the average levels of familiarity with test material at the T1, T2 and T3 stages. Between the T1 (pretraining) and T2 (actual exam) stages the average level increased significantly. Nevertheless, in the fifteen months between the exam and the T3 stage most of the information learned for the examination was forgotten. The displayed values are adjusted so that measurement biases are removed from the data (for further details, see annex).

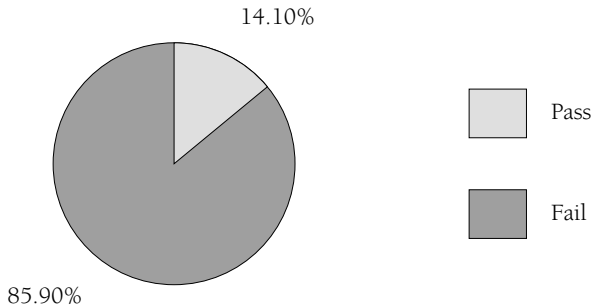
The long-term effect of the BPAES program on examinees is reflected by the positive difference between the T1 and T3 scores. On average this increase compared to the original scores was +39.10 percent. In comparison to the one hundred percent mark, this change represents 6.81 percentage points.⁸

Figure 2. Test Scores at T1, T2 and T3 Stages



Another means of assessing the long-term effect of BPAES is to calculate the percentage of examinees whose scores were above the minimum requirement set for the exam at the time of T3: i.e., the proportion that would fail if examinees unexpectedly had to complete an exam similar to the official (T2) version at the T3 stage. Figure 3 displays the hypothetical exam results at the T3 stage.

Figure 3. Hypothetical Results of Testing at the T3 Stage



3.4. Interpretation of Long-term Change in Familiarity with Test Material

A significant question is whether *the detected change of 6.81 percent is small or large*. The ultimate judgement of the program and its long-term effects is up to the decision makers, as the “masters” of the program must compare the results with their expectations. This is especially pertinent as no quantitative *ex ante* program objectives or program success criteria are specified. Thus, the ultimate interpretation and evaluation is outside the scope of this investigation.

Nevertheless, some reference points may be useful. Indeed, one of the major drawbacks of the current research project is that it does not provide a sound basis on which the magnitude of the program’s effects can be judged. The fact that it has not been possible to apply a more robust research design follows mainly from the universal and uniform coverage of the program and the lack of comparable preprogram data (i.e., pre-1993) on the level of employee familiarity with material currently tested by the BPAES.

In order to decide if the value of *KC2* is indeed high or low, alternative—although substantially less satisfying—bases of comparison may be used.

1. One option is to analyze the retention of an individual's familiarity with the examination material as a result of *other* training and examination programs. Although comparable longitudinal survey data could not be identified, the researchers' intuitive opinion is that this level of long-term program effect is moderate.
2. Another possibility is to *compare the total average effect of the program to that of the best-performing counties*. Within the framework of uniform legal regulations, each county has a certain degree of freedom regarding preparatory training (e.g., the selection of instructors, the timing and organization of courses, etc.). Due to these differences the long-term program effect varies significantly among counties (p value of F-test: 0.009). The two best-performing counties demonstrated a *KC2* value above ten percent—almost double the average program effect. Thus, even within the current general framework of the program, there is substantial potential for improvement.
3. A third point of reference is the *normative expectation* of the program's effect. The most explicit statement in relevant legislation and related documents on the goal of the BPAES program is the minimum requirement to pass the examination. The evaluation of the measured value of *KC2* set forth above indicates that nearly eighty-six percent of examinees would fail if they had to pass an official exam fifteen months after the program (see figure 3).

In summary, the program effect as reflected by *KC2* seems to be modest.

4. Ultimate Effects of BPAES: Practical Utility

The discussion on program effects thus far has been concerned with the necessary but insufficient conditions for the entire program's effec-

tiveness. The program model introduced in the second section indicated that the presence or absence of increased examinee familiarity with test material was expected. If, and only if, there is any long-term increase in the average knowledge of the participants, it is reasonable to analyze the ultimate effect of the program: the *improved professionalism* of public administration. Thus, even a very significant increase in competency in related subjects would not mean that the BPAES program had achieved its actual goal. The increase in familiarity with the relevant material is only a precondition to the utilization of this material.

Due to the practical impossibility of direct measurement because of data, resource and other problems, this report is restricted to proxy measures and qualitative considerations in reviewing the ultimate effectiveness of the program. There are two main sources of information regarding the program's practical effects:

- opinions ("soft" variables) of the examinees and supervisors;
- other related and somewhat "harder" measures outlined in this section.

The most easily accessible information sources were the respondents' opinions. Several questions were directed to both the examinees and their supervisors regarding various indicators of practical program effects. Generally their responses were congruent with what was expected on more theoretical grounds. The more general and commonplace the effect in question, the more positive and "polite" the answers were. Conversely, answers to more specific questions tended to be significantly less reassuring, especially regarding the utilization of knowledge obtained through the training program.

Seventy percent of the examinees answered positively when asked to evaluate the program's effectiveness in improving professional skills. However, slightly more than one year after the exam sixty percent of the examinees could not remember a single case when they had used

material learned in the training program in daily professional practice, and only six percent mentioned more than two such cases.

Supervisors observed a similar tendency. When asked if there was any change in their subordinates' tasks and responsibilities after the exam, one-third of the supervisors responded positively. When, however, questions about more tangible indicators were asked (e.g., if subordinates were trusted to make independent legally binding decisions of lower significance), less than five percent answered positively.

A more reliable and less biased variable may be applied to assess the practical utilization of skills obtained during the training program. Examinees were asked if their positions had changed since passing the exam, and if so, whether such change was a promotion, a horizontal move or a demotion. There appeared to be no relation between improved knowledge (KC2) and the upward mobility of examinees. In fact an adverse—although statistically insignificant—phenomenon was noted: those who moved to higher positions demonstrated relatively less competence (measured either by the absolute level or by the positive change in knowledge). This result seems to be in accordance with the findings of a SIGMA report, according to which “there is no organized and institutionalized connection between performance in training courses and promotion. Positions are at times filled with applicants with the worst training results” (OECD/Sigma 1997, 92).

5. Costs and Efficiency of the Program

In section 4, the *effectiveness* of the program was outlined; in this section, its *efficiency* is explored. In this context “efficiency” is related to comparison of the identifiable effects of the program with estimated costs. It is somewhat surprising that no one appears to have conducted such an evaluation in the past.

As the quantitative estimate of the program's effect has been presented, there is only one element missing in evaluating the efficiency of the program: *costs*. In assessing relevant program costs, a decision must be made as to whether to use the cost for the public in general (social costs) or the financial cost for the various public administration organizations (budget expenditures). As this study is concerned with the social effect of the program, the former option—*social cost*—was selected. This is not simply a theoretical issue; from a practical point of view, social costs and budget expenditures differ in several ways. For example, the efforts of examinees (e.g., spare time spent studying) are part of the social cost but not of the budgetary cost. On the other hand, transfer payments from one party to another do not constitute real social expenditures; e.g., tax and social security fees appear as costs for one public organization and income for another.

There are other complications in providing a relatively precise social cost estimate for the program. Any such estimate must include the program's impact on the general market, yet this only insignificantly increases or decreases the total social cost. In summary, the sophisticated mathematical methodology of cost-benefit analysis has not been utilized in this report. Rather, only the major measurable social costs are considered. When selecting the *sources of cost estimates*, this research predominantly relied on expert assessments based on personal interviews, as data on various cost elements was not readily available or would be misleading.

Four main sources of costs were considered.

1. *Civil servants' working capacity spent on various activities related to the BPAES*. These consist of three days for training courses, one day for the exam, 1.47 days of study leave on average and an average of 2.27 days leave used by examinees from their personal vacation days. Thus, a total average loss of 7.74 working days was projected. Based on the civil service salary tables and the Ministry of the Interior's database, the net daily salary of a civil servant of the sample group was determined as USD 6.16.⁹

2. *Fees paid for trainers.* Based on data from the county administration offices, a USD 5 hourly net wage was paid to trainers with eighteen training hours per course and 42.5 participants per course. Thus, this cost element is USD 2.12 per examinee.
3. *Examination fees.* A net fee of USD 3.5 is paid per examinee.
4. *Additional fixed costs* (mail, administration, room rental, printing, etc.). These were estimated at USD 7.

The total cost per examinee, therefore, is approximately USD 60.3.

Following the calculation of this estimate the research team made a deliberately *overestimated* calculation of the more ambiguous social cost elements. In doing so, higher cost figures for some items and additional minutiae (e.g., energy consumption, cost of transportation to the training centers, etc.) were utilized. The difference between the “best valuation” and the overestimated cost figure was less than forty percent.

The overwhelming portion of costs is wages: lost working days, trainers’ fees and the salaries of examiners and managers. Since civil service wages are quite low, *the cost of the BPAES is surprisingly low*. One reasonably may argue that with such an insignificant cost, even the least effective training program would be quite efficient.

6. Side Effects

According to most evaluation handbooks (Wholey 1994; Carey and Posavac 1992) it is worthwhile to check not only for intended effects but also for side effects or, borrowing the concept from economics, externalities of the program. This section explores such effects. A differentiation is drawn between positive and negative side effects. This report explores only the potential positive effects, as they may justify the existence of programs that otherwise are not able to achieve their intended purpose. In

the case of BPAES, it is especially important to look for positive side effects, since the ultimate goal of the program is not outlined in laws or regulations. As a result, one may argue that the research team misinterpreted the goal of the program, and issues that were interpreted as side effects may have been in fact the original intention of the program.

There appear to be *three primary potential positive effects*.

1. By providing general knowledge of/for public administration the program *fosters horizontal mobility* within and across public administrative organizations.
2. The program *levels* public administration knowledge of civil servants.
3. The program *selects out* civil servants who either do not possess necessary knowledge or are not able to acquire it.

A SIGMA report identifies *horizontal mobility* as an intended effect of BPAES.¹⁰ Unfortunately, there are no data on the change in horizontal mobility within the civil service since BPAES was introduced. However, anecdotal evidence indicates that (1) this rate is still quite low in comparison with the best international practice and (2) the introduction of the BPAES in 1993 did not have any significant impact on the rate itself or on its determining factors. Between the T1 and T3 stages less than two percent of the sample group moved within public administration from one organ to another, whereas—for the same period—the fluctuation in civil service positions was estimated at ten percent.

Limited resources prevented the research team from reaching those who had moved from one public organization to another by the T3 stage. However there was an opportunity to measure the change in knowledge of those who were horizontally mobile within the same organization. If general knowledge fosters horizontal mobility, those who obtained and/or retained more should be more frequently subject to horizontal mobility. However, ANOVA results show that no relationship exists between mobility and knowledge change (*KC2*) or the

absolute knowledge level (T3, p values: 0.194 and 0.763, respectively). In sum, the data do not support the hypothesis that a higher KC2 value increases horizontal mobility.

It may be useful for a training program to *level the general knowledge of civil servants* to a higher-than-previous level. This would indicate that the gap or variation among civil servants, at least with regard to their general standard knowledge, would decrease. If this hypothesis was true, all other conditions being equal, a lower variance of individual scores at the time of the T3 stage should be noted. However, the opposite was discovered. The standard deviation of the participants' knowledge level was *higher* after the program.

Large differences in the knowledge of examinees at the T1 stage were discovered according to educational levels. These systematic differences increased as an immediate effect of the program (ANOVA significance: $p = 0.003$ for T1 scores and still higher, $p \leq 0.0001$, for the exam results). The data seem to indicate that the program does not decrease the variance in scores caused by differences in educational backgrounds.

According to the law, those civil servants who do not pass the exam within the first two years of their careers or who fail to pass the exam three times are terminated from employment. The Ministry of the Interior collects nationwide exam data, but the database does not provide information on how many civil servants have been *dismissed* from the civil service due to BPAES results. Twenty-three percent of the respondents reported that they spent more than two years in the civil service before they sat for the exam. In sum, quantitative data indicates that the legal regulations seem to be stricter than their implementation. After interviews with over a hundred supervisors, hundreds of civil servants and dozens of training and exam organizers, only two individuals had *heard about* someone who was laid off due to failing the exam. A few cases (altogether, less than twenty out of tens of thousands) were identified in which civil servants allegedly left public administration because of their inability or lack of desire to pass the exam.

7. Weaknesses and Possible Improvements of BPAES

This paper focuses on the analysis of program efficacy and provides quantitative, reliable and valid data on the effectiveness and efficiency of BPAES. Measuring program effectiveness may not be enough, however, and there are few—if any—public programs that perform perfectly. Thus, *providing recommendations* to BPAES managers on how to improve program performance is of great importance (Carter 1994). The statistics required to formulate recommendations differ from the data on which this assessment of general program efficacy is based. The latter refer to the actual effects of the program, while the former assist in identifying and analyzing the factors that determine or influence the program's effects.

Recommendations must be based on identified weaknesses and their causes. Some of these causes are beyond the control of the stakeholders. Others, however, may be influenced easily, and by doing so, the program's performance can be increased. The research team made significant efforts to collect information on weaknesses, their causes and potential remedies. Approximately one-quarter of the official evaluation report contained recommendations that were based on a thorough analysis of the current situation. Some substantial *weaknesses and recommendations* to address them are outlined below.

7.1. The Exam¹¹

One result of the research project was discovering that the functions of the current BPAES examination are unclear. By ensuring that examinees learn the required material, BPAES could serve as the stick of a somewhat unbalanced “carrot and stick” system. Or exam results could be indicators of the general knowledge of public administration issues on which promotion and horizontal transfers may be based. Or the exam

could be an objective, reliable feedback mechanism on the level of knowledge of examinees. While answers obtained during several interviews were quite vague, it appeared that program designers and managers rely most on the first function; i.e., the threat presented by the exam and sanctions for failing may encourage examinees to take studying seriously and to do their best in the learning process.

Indeed, the exam seems to fulfill this and only this potential function. The relationship between promotion or horizontal transfers and exam results (ANOVA significance: $p = 0.749$) is insignificant. In addition, the exam results appear to be relatively poor indicators of long-term levels of knowledge.

Exam results largely depend on subjective decisions of the examiners. Analyzing exam results using ANOVA, a significant ($p = 0.002$) difference among counties was noted. This variance may stem from objective factors, such as better training, more qualified examinees, etc., but anecdotal evidence indicates that equal performance on the exam may result in failure in one county and a “fair” or “good” rating in another.

Subjectivity is a major drawback of the examination. The only countermeasure applied in the current BPAES is the examination committee system (three examiners must agree on grades). Multiple choice tests also are used, but only in a limited manner. In addition to the subjectivity of the examiners, another obstacle hindering objective measurement is relatively widespread cheating. Four out of five respondents reported that they witnessed at least some people cheating (either by using the course book or other written documents or by sharing answers with one another) during the examination.

Subjectivity (by the examiners) and cheating (by the examinees) are internal program problems that also have external effects, as they impact the general morale of public administration. Proposed measures to overcome these negative effects include the following.

- *Greater reliance on multiple choice tests.* Currently, multiple choice questions refer to only one chapter of the course book, and less

than one-third of the written exam scores is based on this question format. The written exam could be based on multiple choice questions, while the verbal exam could measure the examinees' understanding of general issues and principles.

- *Shortening the time permitted for the test.* Currently two hours is allowed for the written test. Experience shows that one hour is sufficient for those who are familiar with the topics, and the additional time serves those who cheat.
- *Stricter control and sanctioning on cheating* by exam organizers.

7.2. The BPAES Syllabus

An analysis of the content of the course book and exam questions shows that the exam concentrates almost exclusively on familiarity with current Hungarian laws, especially the Constitution, the Act on Local Self-governments and the Administrative Procedures Act.

The present system can be critiqued from several points of view.

1. The BPAES focuses on teaching individuals about legal *knowledge* rather than improving *skills or attitudes*.
2. Even if one accepts teaching individuals about legal knowledge as a goal, the *choice of the subject matter emphasized* (i.e., valid Hungarian laws) may be criticized.
3. A certain degree of inconsistency can be found between the program's *timing*, its *once-in-a-lifetime character* and its *content*.

One may argue that improving *skills and attitude* is at least as legitimate a goal for training and examinations as increasing general *knowledge* (Shafritz et al. 1992, 182-197, 457-484). Interpersonal and communication skills, computer literacy, foreign languages and a change in service towards citizens and clients are perhaps the most important fields

where training is needed in Hungary and throughout the region (OECD/SIGMA 1997, 13-25).

While improving knowledge—as opposed to skills or attitude—can be induced quickly, it also may disappear rapidly. Program designers and managers tend to forget that the real effect of the program lies in *long-term* change and that this change is less significant than the immediate improvement observed at the exam. Furthermore, it is a widely shared, although false, belief among BPAES designers and managers that only knowledge can be measured by exams.

Another potential point of criticism is the teaching of valid Hungarian laws. This mostly stems from Germanic administrative culture (OECD/SIGMA 1997, 18). *Teaching currently valid legal texts may lead to several difficulties.* First, obscure *legal language* frequently hinders learning and prevents understanding in cases where the content could be taught more efficiently using nonlegal terminology and frames of reference. Indeed, more than half of the respondents complained that the course book was written in legal language that hindered their learning. Widespread anecdotal evidence reinforced by personal observations demonstrates that examinees frequently memorize legal texts but know nothing about their meaning. Basic principles, procedural and democratic structural issues can be explained easily. However, if these issues are “encoded” in the formal legal language of the Constitution and constitutional law, many examinees may not understand the most basic principles of the system. In addition to the “official” exam, questions about the very basics of democratic public administration were asked at the T3 stage. These required the ability to differentiate between political and civil service, between executive and judicial functions, etc.¹² On average, the respondents could answer only slightly more than half (5.69 of 10) of these questions.

Another problem stems from the teaching of *current* Hungarian legal texts. General administrative knowledge should be taught to examinees, but in fact the content of BPAES is very specific both in space (Hun-

garian) and time (current). This may be an overly philosophical problem, but it has practical consequences. The course book, or at least several chapters, must be rewritten and republished repeatedly. Between 1993 and 1997, two crucial chapters and several paragraphs in other chapters became outdated. By 1999, the changes in Hungarian law were so significant that a completely new course book had to be published. Furthermore, examinees frequently have difficulty obtaining photocopies of revised chapters, and despite their efforts, they sometimes study from outdated books containing obsolete laws.

It is important to note that BPAES is related to a certain point in a civil servant's career; the exam must be passed only once, and no retesting is required. Strictly speaking, civil servants *learn laws that become outdated rapidly*. At the same time, this outdated knowledge—at least according to the inherent logic of the program—is retained throughout the active career of a civil servant. It can be stated, somewhat ironically, that the more examinees forget about obsolete laws, the better for Hungarian public administration.

The third point regarding the program syllabus concerns the overall (in)consistency among the program's content, timing and uniqueness. OECD countries utilize a wide variety of syllabi, timing, regularity, etc. in their practice of public administration in-service training (see OECD/SIGMA 1997 and 1999). This variety includes skills improvement, general and specific public administration training, revising and updating courses in various legal subfields, regular and entry training and examinations, etc. However, both OECD experiences and common sense indicate that the current use of the three main program features (content/syllabus, timing and unique character) is less than consistent.

Finally, *various suggestions* were made to address the problems mentioned above. In the short term, the project team recommended that obscure legal language be changed to "normal" language wherever possible. In the long term, the researchers recommended wide-ranging discussion on the syllabus and the examination. In terms of knowledge, more

general material and basic computer and communication segments should be included. Furthermore, the research group suggested that training and the exam be required at an earlier point in an individual's career.

7.3. Lack of Reflectivity

It stems from legal administrative culture that once a law is adopted neither its applicability nor its utility is analyzed systematically in implementation. This does not seem to be specific to Hungary but rather is true for Central and East European administrative systems in general. This administrative culture and, to a large extent, the vested interests of several influential groups in maintaining the present system might be the primary causes of the rigidity of BPAES.

Most of the recommendations made by the research team did not meet the expectations or opinions of supervisors or program managers. Most supervisors (fifty-five percent) prefer the present timing of the examination in an employee's career. Only a few months after submitting the research results, including the suggestion on extending the role of multiple choice test questions, these questions were eliminated completely from the written exam.

8. Conclusions

This paper describes attempts to measure the efficacy of BPAES by following the logical cause-and-effect chain of the program model and identifying the effects of its various elements. Although a relatively high immediate increase in knowledge was noted, the long-term effects appear to be relatively modest.

On the other hand, even if the results seem to be minor, the efficiency of the program may not be low, since the overwhelming majority of the costs consist of civil servants' labor time and, thus, the costs of the

program—especially according to international standards—are minimal. Since efficiency is calculated as outcome/input cost, *the efficiency of the BPAES program seems quite high.*

A great effort was made to control potential “mistakes.” A measure may have been taken incorrectly; data may not be reliable and/or invalid. A detected knowledge increase may have been attributed to the program that is, rather, due to the fact that employees learn while at work. Tests were administered to determine if some effects of the program were missed, either due to inappropriate measurements or overlooked side effects. Such issues were embraced not simply because of the time and energy spent to avoid these potential mistakes but also because they are crucial to the reliability of the findings.

When speaking about reliability, not only its scientific meaning but also the objectivity and robustness of the findings in the eyes of the stakeholders is considered. It is natural that debate be initiated over the evaluation findings. It is also natural that in many cases such debate will be overly political in nature, as various parties dispute the evaluation results, using them to serve their interests. An evaluation report should be “bullet proof” in this regard to as great an extent as possible; all statements either should be appropriately proven or should openly state if this is not so.

The role, applicability and use of public program evaluations in Central and Eastern Europe are crucial issues. It is necessary to identify which factors impede evaluation efforts and other public policy methods and their results. Is it the overly legal approach and an administrative culture that does not favor quantitative, efficacy-oriented methods? Is it the inherent logic of continental law that prevents the application of policy methods and, more specifically, of public program evaluation? Is it the political system and the coalition government that hinder the definition of clear policy goals and thus policy analysis? Or is it the legacy of the communist political system that blocks the application of these techniques? Answering these questions may assist the region in improv-

ing the efficacy of its public administration systems. The authors of this paper plan to launch a metaevaluation research project to try to address these issues.

It is necessary to recognize the fact that this project has not provided information on an international *comparative* basis, and it did not compare BPAES to similar training and/or examination systems. Somewhat unconsciously the research team relied on international best practice (e.g., OECD/SIGMA 1999) for public service training systems in OECD countries, especially when formulating recommendations for improving the Hungarian system.

Hungary has undoubtedly pioneered civil service reform in the region. The stability and efficacy of the country's civil service system generally and its training and examination system specifically seem to be more convincing than those of the majority of postcommunist countries. While few countries in the region had civil service laws even in 1996, in 1992 the Hungarian Parliament adopted the Civil Service Act, which established the basic principles of the training system. Since 1993, a unified, transparent and stable system of basic examinations has existed.

Somewhat similar systems exist in the two countries most often used for comparison with Hungary: the Czech Republic and Poland. However, in the Czech Republic the framework is only valid for a segment of the body of civil servants.¹³ In Poland, administrative apprenticeships introduced by a 1982 government decree are very similar to BPAES. Apparently the training program content, the form of the exam and its role with regard to career paths are quite similar to the Hungarian system.

Some stakeholders plan to support an apprenticeship component of the current BPAES program, requiring daily work in an office and monthly or bimonthly participation in a one- to two-day training course. This proposal could solve the present problem of the examination's timing, introduce motivational and adaptability factors while maintaining the predictability and stability of the current system, which are necessary attributes for an effective training system (OECD/SIGMA 1997, 15-16).

In conclusion, BPAES represents a relatively well-defined, stable and predictable, above-average civil service training system in the region. However, further improvements are necessary in the design and management of the program. First, program objectives must be clearly defined. Second, appropriate methods by which the objectives can be reached should be determined. Third, the program must be adequately implemented. Finally, implementation must be controlled and evaluated on a regular basis.

Annex: Possible Biases in the Evaluation

The findings and conclusions of an evaluation of this nature must be as valid and well founded as possible. At the same time, there are a number of issues that may lead to inaccurate conclusions regarding a program's long-term effectiveness. Some important possibilities for bias in the *KC2* indicator are:

- the *KC2* of long-term program effects does not measure what it is intended to (internal validity problems);
- the identified effects are due to random variations in the sample;
- the identified effects are neither generally invalid nor do they reflect only random effects but rather are somehow biased due to measurement problems.

The first two potential problems can be addressed relatively briefly. First, the most concrete (indeed, the only meaningful) specification is precisely what type of knowledge the BPAES is supposed to provide, an issue that is addressed by the training and examination materials. Thus, it is reasonable to expect that these materials contain the type of information examinees should learn and the program should impart. Apart from this, no significant set of additional skills is introduced by the training course or is required on the exam.

Second, the difference between T1 and T3 scores seems to be systematic. The paired-sample one-tailed *t*-test for the mean difference is highly significant ($p < 0.001$). Thus, the random-effect interpretation must be abandoned.

The third possibility of either positive or negative biases of the *KC2* measurement is discussed below in more detail. *Countermeasures that were applied in order to minimize the various incidents of bias* also are presented.

Possible biases leading to a smaller-than-actual program effect

As the KC2 program effect was calculated on the basis of the difference between T3 and T1 scores, a smaller-than-actual KC2 value (i.e., a negative bias) could be the result of one or both of the following causes:

- the measured values of the T1 results are biased upwards; or
- the measured values of the T3 results are biased downwards.

There are two reasons that T1 scores might be biased upwards: examinees began their preparation for the exam before T1, which would lead to exclusion of some part of knowledge in the KC2 indicator.¹⁴ As a result of various measurements and analyses not discussed here, it was determined that *the effect of early preparation is significant*. When discussing T1 results, “change in knowledge,” “program effects,” etc. in this report, the T1 figure is decontaminated from the effect of early preparation. The other possibility is that examinees obtained better-than-actual results due to cheating during the T1 test. This problem was dealt with in a somewhat simple manner. Care was taken to ensure that examinees did not know that a survey and test would take place prior to T1. As a result, they supposedly were unprepared for effective cheating. In addition, the research assistants conducting and administering the survey were instructed to ensure that the examinees did not cooperate or use any printed material when completing the test.¹⁵

The survey design considers one possible cause that could lead to scores lower than actual knowledge at the T3 stage, i.e., the possibility of negative attitudes: laziness, ignorance and hostility (the following paragraphs refer to this as the “attitude problem”).¹⁶ Analysis showed that *attitude had an effect on T3 scores* and, consequently, on KC2. Respondents with positive attitudes obtained significantly higher scores (ANOVA significance: $p = 0.03$). Furthermore, if the effects of *preparing in advance and a positive attitude are combined, the respective KC2 value rises to as high as 6.81* (see table 2).¹⁷

Table 2. Effect of “Attitude” (at T3) and “Early-Preparation”(at T1) Variables

	No Early Preparation	Early Preparation	Total
Attitude: Negative	-1.76%	-0.72%	-1.54%
Attitude: Positive	6.81%	-0.65%	4.91%
Total	3.56%	-1.32%	2.40%

In summary, *the program effect determined the long-term increase of knowledge of examinees as +6.81 percentage points.* Please note that this is the highest value that is possible to identify. Supposedly the real long-term change in knowledge is somewhat lower than this data indicates but is above 3.5, which is the case if controls for only the early preparation effect and not for negative attitudes are implemented.

Possible biases leading to a higher-than-actual detected program effect

One potential source of bias that can lead to a higher-than-actual detected program effect is the *maturation effect*. This means that although the knowledge of the examinees shows a long-term positive change, this increase is not (entirely) the effect of the program but is due to other, time-related factors: e.g., acquiring more work experience, spontaneous learning processes, etc.¹⁸ Based on various analyses, the research team concluded that *the hypothesis on the presence of the maturation effect can be rejected with high probability.*

Notes

1. Forty-six percent of employers provide additional study leave.
2. Although in the course of developing the field of evaluation, substantially different methodological approaches—the qualitative approaches—have evolved, this report is based on the more “classical” quantitative approach.
3. It is necessary to emphasize that neither the program goals nor the program model is outlined explicitly in any legal or other official document. A SIGMA report

(OECD/SIGMA 1997, 84) identifies the lack of clearly defined objectives as a general attribute of the Hungarian civil service training system.

4. This research design is widely used in public program evaluation when a control group is not available (Wholey et al. 1994). One technical problem with this design is that it makes the identification of respondents necessary, which may lead to dishonesty in answering questions. Researchers for this project used a technique that allowed respondents to submit their personal data and response sheets separately; hence, no significant resistance resulted. A second problem with the longitudinal approach is that a significant portion of the original respondents are “lost” by the end of the research project. The researchers encountered this problem and determined the original sample size accordingly.

5. Please note that this research design excluded those examinees who did not attend the training courses. The group was relatively small (approximately ten percent of all examinees). Choosing a research design that embraced this segment of examinees would have required a much higher budget.

6. In this entry, seventy-one denotes those respondents of the one hundred eighty-nine whose knowledge change data seemed to be reliable beyond reasonable doubt (for more details, annex).

7. In these cases information on the reason for the individuals’ inaccessibility was collected: if the respondent “left the civil service,” “found another position within the civil service,” was “on study leave,” “on sick leave,” “on holiday,” etc.

8. The *average* knowledge change (KC_1) was calculated using the formula

$$KC_1 = \frac{1}{n} \sum_{i=1}^n \frac{T3_i}{T1_i}$$

whereas the *average absolute* knowledge change (KC_2) was based on the formula

$$KC_2 = \frac{1}{n} \sum_{i=1}^n \frac{T3_i - T1_i}{T_{Max}}$$

In both cases $T3$ and $T1$ denote the individuals’ test scores at the $T3$ and $T1$ measurements, respectively, T_{Max} is the maximum test score (57), i is the index of individuals (from 1 to n) and n is the number of individuals.

The first question is the difference between the two indicators of a change in familiarity with the material covered by the examination. The main difference is due to the various bases of reference to which the change in test scores is compared. In the case of KC_1 , the increase (or decrease) in the individuals’ test scores is compared to the original $T1$ score. This formula gives greater weight to individuals starting with lower test scores, even if their $T3$ results are relatively low. In the KC_2 formula the change is com-

pared to the maximum available test score. Thus, instead of relative change, *KC2* is based on absolute ($T3 - T1$) change.

The range of possible values is different as well. *KC1* can range between 0 and +8, whereas *KC2* can range from -1 to +1 (the values of *KC1* and *KC2* are expressed as percentages throughout the text for reasons of simplicity).

In summary, *KC2* seems to have more benefits than *KC1*. Thus, although in practice the two indicators “behave” similarly, this report uses *KC2* as the measure of change in knowledge.

9. Monetary figures are calculated in U.S. dollars.

10. “This examination is an entry pass to civil service, but it also ensures that civil servants with similar basic knowledge are mobile and can easily be transferred among different sectors and areas of civil service” (OECD/Sigma 1997, 85).

11. Although the program is entitled the Basic Public Administration Examination System, the examination itself is only one element. This section analyzes the examination and related issues.

12. These questions were compiled after several consultations with training designers and managers.

13. It is necessary to emphasize that the authors relied on a report in which country profiles were prepared prior to April 1996. The situation may have changed greatly since that time.

14. The research team recognized prepreparation as a serious threat to the validity of the *T1* measure early in the project. As a result, considerable attention was devoted to dealing with this possibility before the *T1* measurement. This precaution was justified as examinees, theoretically, are expected to start their preparation prior to the training course (and, thus, the *T1* measurement). In both the *T1/a* and the *T1/b* questionnaires several partly cross-controlled questions were asked to determine if the examinee had started preparing for the exam before the training course (and, thus, *T1*). It is reasonable to suppose that these measures are effective in controlling for “early” preparation for the exam, especially as even those examinees who did not prepare in advance are more likely to state that they had, since they are officially expected to do so.

15. As indicated in section 2, cheating could not be prevented in two counties. Consequently, these respondents were excluded from the analysis.

16. In order to control this effect two measures were introduced. First, research assistants were asked to score examinees’ attitudes during the completion of the *T3* test on a three-point scale. Second, they measured the time (in minutes) each examinee devoted to completing the test and the questionnaire. The experience, supported by the test completion time, was that examinees generally devoted sufficient time to complete the test. Furthermore, the completion time did not seem to be correlated to the *T3* results.

17. The significance of univariate ANOVA tests for attitude and early preparation variables is less than $p = 0.05$. When testing the two effects simultaneously (two-way

ANOVA) the effect of the attitude variable is significant at 0.05, and the interaction effect between the two variables is significant at 0.055.

18. Before presenting countermeasures to limit biases resulting from a possible maturation effect it is necessary to emphasize that (1) the general character of the program, especially its universal and uniform coverage, does not allow the application of anything truly experimental, and thus, the most effective tools of controlling the maturation effect cannot be applied here; and (2) the overall program effect and the resulting danger posed by the maturation effect is not very large.

In summation, the presence of a maturation effect is unlikely. Prior to the preparatory training course examinees were employed for varying periods of time (mean = 529 days, standard deviation = 374 days). If a maturation effect takes place, it is expected that the longer the time in public administration before the measurement of any one individual's knowledge, the better the test score. On the other hand, if the length of time spent in the examinee's current job is unrelated to test scores, proof has been obtained that spontaneous (self-) development does not have any significant effect on test performance.

Regression and correlation analyses showed that the length of time spent in the current position and T1 test scores are entirely independent. Furthermore, long-run knowledge change resulted in the bivariate correlation between time elapsed and adjusted T1 scores yielding 0.000. This indicates that the hypothesis of maturation as a cause of an increase in knowledge can be rejected with relatively high probability.

References

- Carey, Raymond G. and Emil J. Posavac. 1992. *Program Evaluation: Methods and Case Studies*. Englewood Cliffs, N.J.: Prentice-Hall.
- Carter, Reginald. 1994. "Maximizing the Use of Evaluation Results." In *Handbook of Practical Program Evaluation*, ed. Wholey, et al. San Francisco: Jossey-Bass, 1994.
- Hungarian Institute of Public Administration. *Közisztatási alapvizsga* (Course Book for Basic Public Administration Examination).
- OECD/SIGMA. 1997. "Country Profiles of Civil Service Training Systems." *SIGMA Papers* 12 (Paris: OECD) GD(97)122.
- OECD/SIGMA. 1999. "Public Service Training Systems in OECD Countries." *SIGMA Papers* 16 (Paris: OECD).
- Shadish, Reichardt. 1987. "The Intellectual Foundations of Social Program Evaluation: The Development of Evaluation Theory." *Evaluation Studies Review Annual* 12.
- Shafritz, J. M., N. M. Ricucci, D. H. Rosenbloom, and A. C. Hyde. 1992. *Personnel Management in Government*. New York: Marcel Dekker.

- Sonnichsen, R. C. 1994. "Evaluators as Change Agents." In *Handbook of Practical Program Evaluation*, ed. Wholey, et al. San Francisco: Jossey-Bass, 1994.
- Wholey, J. S. 1979. *Evaluation: Promise and Performance*. Washington, D.C.: The Urban Institute.
- Wholey, J. S., H. P. Hatry, and K. E. Newcomer. 1994. *Handbook of Practical Program Evaluation*. San Francisco: Jossey-Bass.