Session 5: Emerging and other strategies for productivity and performance

Title: Testing Effectiveness of the Adjusted Performance Measure Strategy: Evidences from a U.S. Federal Job Training Program

Do Han Kim, University at Albany, SUNY, USA
Pascal Courty, European Institute, Italy
Gerald Marschke, University at Albany, SUNY, USA*

Paper Presented at the International Conference of "A Performing Public Sector: The 2nd Transatlantic Dialogue", Leuven (Belgium), June1-3, 2006

**This paper is a preliminary analysis. Please do not cite without authors' permission.*

* Do Han Kim is contacted at: University at Albany, SUNY, 135 Western Ave., Albany, NY12222, USA; Tel)518-432-6118; email)dk7535@albany.edu: Pascal Courty is contacted at: European University Institute, Villa San Paolo, Via della Piazzuola 43 Firenze, 50133 Italy; Tel)+39 055 4685925; email) Pascal.Courty@IUE.it: Gerald Marschke is contacted at: University at Albany, SUNY, BA112A, 1400 Washington Ave., Albany NY12222, USA; Tel)518-442-5274; email)marschke@albany.edu

Introduction

Although there has been much interest in introducing performance measurement in the public sector, many have expressed reservations that performance management systems may generate negative responses. Especially, outcome-based performance measures[1], which has been one of the popular methods to evaluate social programs, has been criticized as misleading agency's efforts to increase measurable program outcomes rather than to improving original program goals such as value-added or program impact. For example, if job training agencies', schools', or hospitals' performance are measured by their program outcomes such as job placement, test scores, or mortality rate, they can maximize their measured performance by serving only job-ready training applicants, the smartest students, or the healthiest patients rather than trying to serve the hard-to-serve clients[2]. This behaviour is problematic because the most in need clients for social programs are unfairly excluded from the public services due to the performance incentive systems.

The adjusted performance measures (APM) has been developed to reduce the limitations of outcome-based performance measures by adjusting each organization's performance standard in the consideration of the percent of hard-to-serve clients and uncontrollable environments. If organizations serve a larger number of hard-to-serve clients than average, their performance standards will be set lower than average organizations' performance standards. In addition, local economic conditions outside

[1]In this paper, outcome-measure is defined as measures of outcomes of the program at some designated date following Barnow (1992)'s definition of gross outcome measures. For more details of typologies of performance measures, please look at Barnow (1992), Hatry (1999), and Propper and Wilson (2003).

[2] In this paper, we define "hard-to-serve" clients as "individuals with labor market deficiencies or barriers to employment" following Barnow and Constantine (1988)'s terminology in performance management in job training program literature. According to them, most common deficiencies are thought to be a lack of basic skills, particularly reading skills, and a lack of work experiences. Also, some of demographic characteristics which training agencies have had generally difficulty in finding jobs after training like older workers, members of minority groups, or female, were also included as "hard-to-serve" clients. For more details about a definition of hard-to-serve, please look at Barnow and Constantine (1988).

of organizational control but correlated with performance outcomes are also reflected in the adjustment model to level each organization's environmental conditions.

This study introduces APM strategy as alternative to a simple outcome-based performance measures and investigates the effectiveness of APM using the case of the US federal job training programs. Between 1982 and 1999, the U.S. Department of Labor (DOL) used outcome–based performance management system and a regression-based adjustment method as part of the Job Training Partnership Act (JTPA) programs, to serve the economically disadvantaged. We use exogenous changes in target group weights to measure the influence of the adjustment method on enrollment of hard-to-serve clients and other sub-groups. The quality of the enrollee change among the same client group will be examined as proxy measures for training agency's client selection rule changes.

This paper is composed as follow. The next section introduces the basic mechanism of the APM, and strengths and weakness of this approach. The third section describes historical development of the adjustment model used by US DOL in the JTPA. The empirical analysis will follow and, finally, the policy implications will be suggested.

Adjusted Performance Measures (APM)

Why Do We Need Adjusted Performance Measure?

The rationale for the use of the APM strategy started from realization of limitations of the outcome-based performance management system. First and well-recognized, outcome-based performance incentive systems cause unintended perverse behaviour like, most infamously, cream skimming which is agency's client selection discrimination to maximize incentive awards. If organizations' performance were

measured by their gross program outcomes, agencies are likely to be tempted to increase their performance outcomes by selecting clients who are likely to increase measured performance easily rather than by improving quality of services to hard-to-serve clients. Cream skimming is problematic, especially in public programs, because citizens' opportunities to receive basic social services are restricted based on unfair and arbitrary criteria like demographic characteristics or educational level.

The second limitation of outcome measure is that it cannot single out program outcomes which are generated by organizations' efforts from outcomes which are generated by organizations' environments. Program outcomes in social programs are usually influenced by uncontrollable factors like local economic conditions or unexpected shock from outside the organizations. For example, program outcomes of public job training are largely influenced by their local economic situations such as job market situations or unemployment rates. Proponents of the APM argue that program outcomes which are determined by the environment should be controlled to level the organizational playing field.

The APM attempts to fix these limitations of outcome-based performance measures using a statistical method. In the APM approach, each organization's performance outcomes are adjusted by organizations' uncontrollable environments and by the proportion of hard-to-serve clients among service recipients.

How Does Adjusted Performance Measure Work?

Stiefel et al. (1999) explain the procedure of adjusting performance measures as follow. First, a group of experts in the field select potential adjustment factors based upon the previous experiences and theoretical backgrounds. Potential adjustment factors include controllable factors but uncorrelated with actual organization's

performance like initial demographic characteristics of clients and uncontrollable factors but correlated with performance outcomes like local economic conditions. Second, model developers chose types of data they use among available data like datasets of previous program years or datasets of the organization in similar environments. Third, they estimate the influences of potential factors on performance outcomes by running a regression with performance outcomes as dependent variable and with potential adjustment factors as independent variables. Fourth, statistically significant regression parameters are selected as adjustment factors and the regression coefficients become adjustment weights. Fifth, using selected adjustment factors and weights, they calculate the expected performance outcome for each agency reflecting its target population mix and environment. Finally, the difference between actually measured performance outcome and predicted (expected) performance outcome will be adjusted performance outcomes for each organization[3]. As a result, under the APM system, each agency confronts different performance standard (=expected performance outcome), and agency's performance will be evaluated by adjusted performance outcomes (= expected performance outcomes– measured performance outcomes) instead of by simple measured performance.

What does this mean? If any agencies accept only very productive participants like the smartest students or job-ready training applicants, their expected performance will be set higher than other agencies and they should achieve much higher measured performance outcomes to receive performance incentives. Also, if training agencies enroll a higher percent of hard-to-serve clients than other agencies, their expected performance will be reduced than others and they have lower performance standards to meet. Therefore, there are no more incentives to enroll only high productive clients

[3] For detailed description of the construction of APM, please refer to Stiefel et al. (1999).

or to avoid hard-to-serve clients, and ideally incentives to cream skimming can disappear. In addition, by including agency's local environments as adjustment factors, the APM is also capable of handling the second limitation of outcome-based performance management system. By assigning high expected performance to good neighbourhood agencies and give lower expected performance to bad neighbourhood agencies, the APM reduces the bias of the simple outcome measures and it will provide relatively accurate information about agencies' performance.

The APM, however, is not perfect. First of all, the adjustment model uses only some measurable and collectable information about clients' characteristics.[4] Thus, if agencies cream skim using information on clients which are not included in the model but influential on performance outcomes, they can increase performance outcomes even after the adjustment process (Barnow, 1992; Barnow & Constantine, 1988; Courty & Marschke, 2004; Cragg, 1997). In addition, if relevant factors are omitted from the model, the estimated weights of the other factors included in the model can be biased (Barnow & Constantine, 1988)[5]. Using Monte Carlo simulation, Brooks (2000) showed that, if there exists a correlation between observed demographic characteristics and unobserved merit, or, if the APM model does not explain a very large part of the variance in the outcome measure, the APM will not be reliable.

US Department of Labor's Adjusted Performance Model in the JTPA

Performance Management System in the JTPA

[4] Barnow and Constantine (1988) showed a number of reasons why not all relevant factors are included in the regression adjustment models. First, there is not a consensus as to what all the relevant factors are. Second, practically, adding factors would add to the data collection burden for agencies and make the required computations more complex. Third, statistical problems might result when factors are added because if several factors are highly correlated, then the individual weight may be less precise and change significantly from year to year. Finally, many factors are difficult to measure.

As one of the largest federal job training programs, the JTPA used the outcome-based performance incentive system and adjustment model between 1982-1999. Referring to the JTPA, the program goal was to increase earning abilities of economically disadvantaged and their performances were supposed to be measured by increased employment and earnings of the participants and reduction of welfare dependencies. In absence of value-added measure, DOL developed outcome measures focusing on labor market outcome like job placement and weekly earnings of training terminees. During 1980s, DOL created two different performance standards, national average performance standards and adjusted performance standards, and let governors chose one of them and apply it to states. However, as of 1992, DOL stopped using national average performance standards and required all states to use adjusted performance standard approach. Still, the Governor had discretion over using the adjustment model developed by DOL or have the state develop their own adjustment models[6]. The incentive grants were distributed up to 6% of JTPA funds for job training providers who exceeded some or all of the performance standards, and if any job training providers fail to meet performance standards persistently for two years, they were supposed to be penalized by the Governor's reorganization plan.

DOL's Adjustment Model

DOL estimated adjustment factors and weights using information collected in previous program years every year during the 1980s, and every other year during the 1990s. Adjustment weights were calculated for each performance measure: follow-up

[5] For examples of omitted factor bias, please see (*Training Hispanics: Implications for JTPA System*, 1990)

[6] Out of 33 states who answered our survey, 29 states used Department of Labor's adjustment method, and only 4 states developed their own models.

adult entered-employment rate, follow-up weekly earnings, and welfare entered-employment rate. Equation(1) shows the simplest form of regression model where P is performance outcomes, and X is potential factors for participants' characteristics (more specifically, percentage of each hard-to-serve clients), and Z is potential factors for local economic conditions. After running the regression including all potential factors on performance outcomes, if there are any insignificant factors, those factors are excluded from the model. Then, they ran the restricted model again and excluded insignificants factors repeatedly until all the factors would be statistically significant. Statistically significant factors become adjustment factors and their regression coefficients become adjustment weights (Barnow, 1992; *Guide to Performance Standards for the Job Training Partnership Act for Program Years 1998 and 1999*, 1998; *Guide to Setting JTPA Title II-A Performance Standards for PY 87*, 1987)

$$P_{ijn} = \beta_0 + \sum \beta_i X_{ijn} + \sum \tau_{jn} Z_j + \upsilon_{ijn} \quad \text{.....Eq. (1)}$$

In the 1980s, the adjustment model used a limited number of adjustment factors like training applicants' demographic information such as age, gender, and race, educational level, welfare receipt, and local economic conditions such as unemployment rate and annual earnings. However, as time went on, there were emphasis on the important roles of clients' labor market information, and, in the 1990s, many labor market histories like work histories, labor force participation, and period of unemployed were newly introduced. Table1 shows the changes of adjustment factors and weights during the 1990s. One interesting point of the table is that most adjustment factors seem to have patterns for their lives in the model. When adjustment factors were introduced to the model at first, most of them had a relatively

large amount of negative values. Then, as time went on, the amount of adjustment weights decreased gradually, and old adjustment factors eventually disappeared from the adjustment worksheets and new factors were introduced. The lifecycles of adjustment weights exactly supports Barnow and Constantine's prediction which was published during the early JTPA periods. In their 1988's report, they said that "…one possible reason that the estimated weights in the regression models have become smaller and smaller in successive years is that training agencies have become more adept at selecting participants on the basis of factors that are omitted from the models"[7]. (*Guide to Performance Standards for the Job Training Partnership Act for Program Years 1998 and 1999*, 1998; *Guide to Setting JTPA Title II-A Performance Standards for PY 87*, 1987)

<Table 1 Adjustment Factors and Weights>

Once DOL finalizes adjustment factors and weights, they create adjustment worksheets and distribute them to each training agency. As you can see in table2, the adjustment worksheet includes adjustment factors, adjustment weights, national average, and national departure points. Each training agency is supposed to fill the column "B: SDA factor values (=percent enrollment for the client group)" and its adjusted performance standard is decided by a sum of deviations of percent enrolment of hard-to-serve clients and the national average enrollment of that groups multiplied by adjustment weights. In our example, the adjustment weight for females was -0.052 and national average of the percent of female was 67.7. If a training agency enrolls females more than the national average, for example 10% higher than the average,

[7] The phenomena will be discussed more in the policy implication section.

like 77.7%, then the effect of local factors on performance is -0.52 (=(77.7-67.7)*(-0.052)). Thus, the adjusted performance level will decrease from 59.0 to 58.48 and the training agency confronts a lower performance standard than the national average since they enroll a higher percent of hard-to-serve clients (females). Therefore, training agencies enrolling a higher number of hard-to-serve clients with negative adjustment weights will be rewarded by reduced performance standards, and incentives to cream skimming will decrease under the adjusted performance model. (Barnow, 1992; *Guide to Performance Standards for the Job Training Partnership Act for Program Years 1998 and 1999*, 1998; *Guide to Setting JTPA Title II-A Performance Standards for PY 87*, 1987).

Table2 Example of Adjustment Work Sheet

A. Local Factors	B. SDA Factor Values	C. National Averages	D. Difference (B--C)	E. Weights	F. Effect of Local Factors on Performance (D*E)
Female	*77.7*	67.7	*10.0*	-0.052	*-0.52*
		G. Total (∑ F)			
		H. National Departure Point			59.0
		I. Model--Adjusted Performance Level (G+H)			*58.48*
		J. Governor's Adjustment			
		K. SDA Performance Standard			

Research Hypothesis

In this section, we draw a prediction on the most rational client selection rule of training agencies' under the APM and outcome-based performance incentive system. The next section will examine it using a statistical model.

First of all, we assume that training agencies purport to obtain incentive awards and avoid sanctions under the performance management system. Performance incentives (or sanctions) are distributed based upon whether training agencies' measured performance outcomes exceed their performance standards. Hence, in order to increase possibilities of

incentive awards, training agencies should maximize *a difference between measured performances and adjusted performance standards.* The measured performance outcomes in our analysis are gross program outcomes measured by job placement rate and weekly earnings several months after the training graduation. Adjusted performance standards are performance standards for each training agency adjusted by the percent of hard-to-serve clients and their local environments. Which client select rule does enable training agencies to increase measured performance and reduce performance standards at the same time?

First, training agencies can easily increase measured performance by enrolling only high productive applicants and reject hard-to-serve applicants whose characteristics are in the adjustment model. They are cream skimming and, without the adjustment model, they can improve the probability of incentive awards by cream skimming. Yet, under the adjustment model, the increased measured performance effect by cream skimming will be cancelled out by increased performance standards because their percent enrolment of hard-to-serve clients will be lower than the national average, and their performance standards will be adjusted upward. Theoretically, the amount of increased measured performance made by cream skimming is very similar to the amount of increased performance standard. Therefore, as we described in the previous section, the traditional cream skimming strategy is ineffective under the APM.

Second, contrasted to the first option, training agencies can enroll a relatively high number of hard-to-serve clients to lower their performance standards. However, as you can imagine easily, this strategy is also ineffective because their measured performance is supposed to move downward as they enroll hard-to-serve clients one by one. Thus, the decreased performance standards will be cancelled out by the decreased measured

performance. It implies that APM can reduce cream skimming but it cannot encourage training agencies to enroll hard-to-serve clients.

The third possible strategy for training agencies is selecting the most productive clients among hard-to-serve clients using information which is correlated with clients' performance but omitted from the model. In this case, since they are cream skimming among the hard-to-serve clients, they still have a lowered performance standards and, furthermore, they can enroll more productive client groups. For example, even though the DOL adjustment model gives the same weights to female welfare recipients, there are several different types of female welfare recipients which the adjustment model cannot distinguish. Some of female welfare recipients are possibly well-motivated than others, or some of female welfare recipients are more serious about their jobs than others. During application interview, training agents can collect this kind of unmeasurable information and select only well-motivated and sincere applicants among female welfare recipients. By doing this, they can increase performance outcomes and decrease performance standards at the same time.

In this study, we hypothesize that training agencies are smart enough to take the third strategy and enrolled high productive clients among the hard-to-serve clients to maximize their possibilities of incentive awards. Particularly, we suppose that training agencies would select more productive hard-to-serve clients if adjustment weights increase for those groups and would select relatively less productive hard-to-serve applicants if adjustment weights for those groups decrease. The reason is simple. If adjustment weights increase from program year t to program year t+1, training agencies feel pressure to enroll more productive applicants than before because their performance standards does decrease, but not as much as in the first t year iteration. To the contrary, if adjustment weights decrease from program year t to program year t+1, training agencies

feel free to enroll just relatively less productive applicants than before because their performance standards does decrease more than before as they enroll those types of clients. Consequently, our research hypothesis can be described as follow.

Under the APM and outcome-based performance incentive system, if the adjustment weight for the hard-to-serve group 'i' increases, training agencies will select more productive applicants among hard-to-serve clients, and, if the adjustment weight for the hard-to-serve group 'i' decreases, training agencies will select relatively less productive applicants among hard-to-serve clients. Thus, there exists a positive relationship between adjustment weight change and performance outcome change for the same hard-to-serve clients.

Data

We use Standardized Program Information Report (SPIR) data collected by the DOL and distributed by the WE Upjohn Institute of Employment Study. The data contains demographic information on all individuals served, services provided, and outcomes attained at the individual level under JTPA during program years 1993-98. We evaluate the major adult program, Title II-A.

We contacted employment offices of 50 states and surveyed whether they used the DOL adjustment model or developed their own models. Out of 33 states who responded our survey, 29 states answered that they used DOL model and 4 states answered that they developed their own models. From DOL, we obtained copies of the "Guide to Performance Standards for the Job Training Partnership Act for Program Years 93, 94-5, 96-7, and 98-9" and the reports have adjustment worksheets for each of adult performance measures during our research period, follow-up entered employment rate (FER) and follow-up weekly earnings (FWE). The adjustment worksheets have information about adjustment factors, national average enrolment, and the adjustment weights. We also tried to collect information on adjustment models used in 4 states who

did not use DOL model, but states didn't have data about it. Thus, we decided to use training agencies in 29 DOL model states as a treatment group and 4 non-DOL model states as a control group for our analysis.

Referring to adjustment worksheets, 33 factors were included in the adjustment model on and off between py93-98. We selected 13 large adjustment factors whose number of enrollees are at least 10 percent of total number of enrolees out of 33 factors[8] and created 2^{13} non-overlapping and exclusive "index" groups. We use the index as a unit of analysis and performance outcomes are averaged at the index level and adjustment weights are added up at the index level. For example, "index female" represents female training terminees whom none of other adjustment factors apply to except for "female" factor and the index female's adjustment weight will be the same as the adjustment weight for female. As an index level, the data has an unbalanced panel format with 1837-2203 units and 6 time periods.

Analysis Model

Equation (2) displays the base model tested in the study. Because our data structure is a panel (cross-sectional time series) and also we are interested in the relationship between the changes of adjustment weights and changes of the performance outcomes, we will run a first-differenced fixed-effect model. In the left side, we include the change of average performance outcomes of the client group i at program year t (Pit), and, in the right side, we include the change of adjustment weight for group i at time t (Wit) as an explanatory variable and change of economic conditions like unemployment rate

[8] Selected 13 adjustment factors are female, age 30-54, black, minority male, high school dropouts, post high school attendees, handicapped (changed into "individuals with disability" in 1994), Unemployment Compensation claimants or exhaustee, long-term Aid to Families with Dependent Children recipients, cash welfare recipients, lacking significant work histories, unemployed 15 weeks or more, not in labor force, and veteran. Basic skill deficient and reading skills below 7th grade also

(unempit) and average earnings (earnit) are included as control variables. If training agencies enroll more productive clients under increasing adjustment weights, the sign of beta1 will be positive. We will run this regression model for two separate groups of states: states with DOL adjustment model and states without DOL adjustment model. With two separate runs, we expect a pseudo-experiment situation with DOL model states as treatment group and with non DOL model states as a control group. If training agencies actually respond to the DOL model, only the first group's regression coefficients should be significant.

$$\Delta P_{it} = \beta_0 + \beta_1 \Delta W_{it} + \alpha_1 \Delta unemp_{it} + \alpha_2 \Delta earn_{it} + \Delta u_{it} \quad \text{.....eq. (2)}$$

Equation (3) shows a regression model for an analysis using pooled sample states. This time we include all 33 states as a pooled sample, and the regression equation has a dummy variable for DOL adjustment model and an interaction term of DOL adjustment model and adjustment weight. This model is an extension of the first model, and it enables us to conduct a statistical test about if there exist any difference between DOL adjustment model states and non DOL model states in respect to the relationship between adjustment weights and performance outcomes. We predict a statistically significant regression coefficients beta 2 and beta 3.

$$\Delta P_{it} = \beta_0 + \beta_1 \Delta W_{it} + \beta_2 dolD + \beta_3 dolD * \Delta W_{it} + \alpha_1 \Delta unemp_{it} + \alpha_2 \Delta earn_{it} + \Delta u_{it}$$

.....eq.(3)

Results

Table 3 shows results of 8 different regression analyses. The first two regressions are results of base the model with DOL adjustment model states. The first

---

covered a large proportion of population, but they are excluded from the analysis because they have too much missing information.

regression uses a FEE as a dependent variable and the second regression use FWE as a dependent variable. As we expected, regression coefficients for adjustment weight changes in both analyses are positive and statistically significant at 0.05 level. The first analysis can be interpreted as, if adjustment weight of entered employment rate increased by one unit, the training terminees' job placement rate improved by 9.5 percent. The second analysis also shows that, as adjustment weight of weekly earnings increased by one unit, training terminees' weekly earnings improved by 32 dollars.

The third and fourth regressions also run the base model using 4 non DOL adjustment model states, and, as we expected, neither of the regression coefficients for adjustment weight changes were significant.

The 5^{th}-8^{th} columns are results of the second analysis model with the pooled sample. The fifth and sixth model conduct a significance test for average performance outcome differences between DOL adjustment model group and non DOL model group, and the final two analyses look at the slope differences between two groups by including interaction terms. The fifth model shows that training terminees in states with the DOL model employed average 0.06 percent higher than training terminees in states without the DOL model, and DOL model states' average weekly earnings are higher than non DOL model states by 7 dollars on average. Both differences were statistically significant. The final model shows a significantly different slope coefficient for weekly earnings between states with the DOL model and without the DOL model. However, the slope difference for entered-employment rate was not significant.

<Table 3 Results of Regression Analysis>

Policy Implications

The first implication of this study can be found from the theoretical arguments. If the APM works well, the limitations of the outcome-based performance management system can be alleviated. There have been large efforts to find evidence for cream skimming in the US job training programs, but results of empirical studies are mixed[9]. We claim that the weak evidence for the cream skimming in the job training programs can partially be ascribed to the APM of the JTPA. Especially, most research attempted to find evidence of cream skimming by looking at enrollees' composition of demographic characteristics and labor market conditions which are already included in the DOL's adjustment model.

Second, the data analysis shows that training agencies actually changed their client selection rule in response to the adjusted performance measure approaches. The statistically significant regression coefficients to the DOL model samples indicate that the select clients' characteristics are significantly changed by the adjustment weight changes. Especially, they selected more productive clients as the adjustment weight for the groups increases because, under the increasing adjustment weights, training agencies feel pressure to enrol the clients due to possibly increasing performance standards and they will become more picky. In contrast, if adjustment weight decreases for some groups of clients, training agencies became more generous in selecting clients because they would be compensated by reduced performance standards as they enrol those characteristics of clients.

Third, training agencies also attempted to cream skim under the APM using omitted factors from the adjustment model. The positive signs of regression coefficients show that training agencies attempted to cream skim among the hard-to-

[9] For a summary of empirical studies on job training programs, please look at Heckman, Heinrich, and Smith (2003) or Barnow and Smith (2004).

serve client groups. As Barnow (1992) pointed out, even though the DOL adjustment model includes potential factors affecting performance outcomes like demographic, human capital, and labor market variables, there are still many factors omitted in the model that influence performance outcome significantly. If training agencies have information on omitted factors, they are capable of engaging in creaming regardless of the adjustment models. Marschke (2002) also shows that training agencies increase their performance outcome by manipulating the types of training which are omitted in the DOL's adjustment model instead of changing client mix.

Finally and most importantly, policy makers should remember that government programs should pay attention to bureaucrats' responses to the programs and modify the programs during the implementation considering bureaucrats' behaviours. Our research presents that bureaucrats in training centers continually change their client selection rules after the adjustment model introduced. As indicated in explaining table 1, many adjustment factors were introduced and disappeared because the relationship between adjustment factors and performance outcomes faded away. Also, half of the adjustment weights moved toward zero value. It looks like that the adjustment factors experience an evolutionary life. When model developers find a significant relationship between the factor and performance outcome, a new adjustment factor is introduced in the adjustment model. Then, bureaucrats in training agencies detect newly entered adjustment factors and change their behaviour (in our case, client selection rule) to improve their performance outcomes and obtain the incentive. Because of the bureaucrats' gaming responses to the new factor, the effects of the adjustment factor diminished as time goes on, and finally the factors become insignificant in the regression model and disappear from the adjustment model.

In the same context, Courty and Marshchke (2003) argue that good performance measures can be determined only after observing bureaucrats' gaming responses that the model generated because performance measures and policy goals are endogenous.

Conclusion

Workforce Investment Act of 1998 (WIA) drops off the model-based performance management system, and gives state or local site discretion in negotiating performance standards and allows petition to the DOL if circumstances change significantly. As Barnow and Smith (2004) point out, without the model-based adjustment model, the cream skimming incentive will be higher under the WIA than under the JTPA. This study shows how APM can complement to traditional outcome-based performance management system. With the APM, the outcome measure provides unbiased and accurate measurements about an organization's performance and also reduces unintended perverse behaviours. With the APM, outcome based performance management system can function as an alternative to value-added measures.

Still, empirical studies in this paper indicate that the adjustment model should be evolved in response to bureaucrats' behavioural changes. As we already showed, the power of adjustment factors and weights seemed to be reduced right after bureaucrats understand the mechanism and attempt to game with it. The only way we can keep the effectiveness of the APM is keep trying to find omitted factors and include them in the model and also improve the model considering bureaucratic responses to the model.

In many cases, politicians and public officers spend most of their time on designing the programs and neglect to revise or modify after the program

implemented. Our study shows that, to be effective, government programs should be revised and improved even after the program implemented.

Reference

Barnow, B. S. (1992). The Effects of Performance Standards on State and Local Programs: Lessons for the Job Opportunities and Basic Skills Programs. In C. Manski & I. Garfinkel (Eds.), *Evaluating Welfare and Training Programs* (pp. 277-309). Cambridge, MA: Harvard University Press.

Barnow, B. S., & Constantine, J. (1988). *Using Performance Management to Encouraging Services to Hard-to-Serve Individuals in JTPA*. Washington, D.C.: National Commission for Employment Policy.

Barnow, B. S., & Smith, J. (2004). Performance Management of U.S. Job Training Programs: Lessons from the Job Training Partnership Act. In C. J. O'Leary, R. A. Straits & S. Wandner (Eds.), *Job Training Policy in the United States*. Kalamazoo, Michigan: W.E.Upjohn Institute for Employment Research.

Brooks, A. C. (2000). The Use and Misuse of Adjusted Performance Measures. *Journal of Policy Analysis and Management, 19*(2), 323-329.

Courty, P., & Marschke, G. (2003). Dynamics of Performance Measurement Systems. *Oxford Review of Economic Policy, 19*(2), 268-284.

Courty, P., & Marschke, G. (2004). Benchmarking Performance. *Public Finance and Management, 4*(3), 288-316.

Cragg, M. (1997). Performance Incentives in the Public Sector: Evidence from the Job Training Partnership Act. *Journal of Law, Economics, and Organizations, 13*(1), 147-168.

Guide to Performance Standards for the Job Training Partnership Act for Program Years 1994 and 1995. (1994). Menlo Park, CA: Social Policy Research Associates.

Guide to Performance Standards for the Job Training Partnership Act for Program Years 1996 and 1997. (1996). Menlo Park, CA: Social Policy Research Associates.

Guide to Performance Standards for the Job Training Partnership Act for Program Years 1998 and 1999. (1998). Menlo Park, CA: Social Policy Research Associates.

Guide to Setting JTPA Title II-A Performance Standards for PY 87. (1987). Washington, D.C.: Employment and Training Administration, U.S. Department of Labor.

Hatry, H. P. (1999). Performance Measurement Principles and Techniques: An Overview for Local Government. In R. C. Kearney & E. M. Berman (Eds.), *Public Sector Performance: Management, Motivation, and Measurement* (pp. 304-328). Boulder,CO: Westview Press.

Marschke, G. (2002). Performance Incentives and Organizational Behavior: Evidence from a Federal Bureaucracy. *Journal of Human Resources, Revising for invited resubmission*.

Propper, C., & Deborah, W. (2003). *The Use and Usefulness of Performance Measures in the Public Sector* (No. 03/073): CMPO.

Stiefel, L., Rubenstain, R., & Schwartz, A. E. (1999). Using Adjusted Performance Measures for Evaluating Resources Use. *Public Budgeting and Finance, 19*(3), 67-87.

Training Hispanics: Implications for JTPA System. (1990). National Commission for Employment Policy.

Table 1adjustment weight change during PY92-PY99

Adjustment factors	PY92-3		PY94-5		PY96-7		PY98-9	
	FER**	FWE***	FER	FWE	FER	FWE	FER	FWE
Female*	-0.072	-0.425	-0.056	-0.443	-0.052	-0.602	-0.05	-0.683
55 years old & over		-1.126	-0.118	-0.774	-0.105	-0.484	-0.13	-0.61
age 30 to 54*			-0.098					
Black*	-0.064	-0.27	-0.086	-0.325	-0.035	-0.226	-0.027	-0.177
Other minority				-0.1		-0.042		-0.065
Minority male*						-0.279	-0.026	-0.306
High school dropout*	-0.184	-0.271	-0.084	-0.276	-0.073	-0.24	-0.066	-0.145
Post high school attendees*		0.415	0.066	0.659	0.032	0.235	0.008	0.334
High school dropout under 30					-0.02		-0.015	-0.088
Handicapped (individuals with disability PY94-)*	-0.083	-0.367	-0.09	-0.558	-0.075	-0.28	-0.096	-0.315
UI or UC claimant (UC claimants or exhaustees)*		1.062		0.361	0.037		0.022	0.081
Long-term AFDC recipient*	-0.151		-0.234		-0.025		-0.018	-0.086
Cash welfare recipient*					-0.054	-0.093	-0.031	-0.072
SSI recipient					-0.091	-0.027	-0.133	-0.265
Offender			-0.057					
Limited English speaking						-0.259		-0.251
Basic skills deficient					-0.034	-0.193	-0.037	-0.286
Reading skills below 7th grade		-0.148	-0.032	-0.344				
Lacking significant work history*	-0.074	-0.292	-0.059	-0.144	-0.05	-0.15	-0.055	-0.098
Unemployed 15 weeks or more*	-0.111		-0.103	-0.242	-0.086	-0.091	-0.073	-0.076
Not in the labor force*	-0.122		-0.113		-0.103		-0.108	-0.044
GA/RCA recipient	-0.137		-0.05					
Veteran (Vietnam era veteran)*	-0.16		-0.135		-0.03		-0.081	
Homeless		-0.595		-0.602			-0.043	-0.136

*: 14 adjustment factors included in our analysis **:Follow-up entered employment rate ***: Follow-up weekly earnings
This table is developed by combining adjustment weight worksheets from (*Guide to Performance Standards for the Job Training Partnership Act for Program Years 1994 and 1995*, 1994; *Guide to Performance Standards for the Job Training Partnership Act for Program Years 1996 and 1997*, 1996; *Guide to Performance Standards for the Job Training Partnership Act for Program Years 1998 and 1999*, 1998)

Table2 Results of Regressions

Samples	DOL model states		Not DOL model states		Pooled sample (all states)			
Performance Outcomes	FER (1)	FWE (2)	FER (3)	FWE (4)	FER (5)	FWE (6)	FER (7)	FWE (8)
Adjustment Weight	**0.095****	**31.821****	**0.064**	**19.269***	**0.085****	**29.698****	**0.087****	**27.886****
(SE)	*(0.040)*	*(5.213)*	*(0.099)*	*(10.458)*	*(0.037)*	*(4.693)*	*(0.037)*	*(4.697)*
DOL dummy					**0.016****	**7.108****	**0.013**	**16.498****
(SE)					*(0.007)*	*(2.914)*	*(0.008)*	*(3.206)*
DOL*AW interaction							**-0.015**	**19.566****
(SE)							*(0.022)*	*(2.792)*
Constant	*-0.025*	*-13.663*	*-0.020*	*-10.599*	*-0.025*	*-13.489*	*-0.025*	*-11.277*
(SE)	*0.009*	*3.478*	*0.023*	*7.524*	*0.009*	*3.166*	*0.009*	*3.180*
Unemp	0.02	4.34	0.04	12.69	0.02	5.05	0.02	5.03
(SE)	0.002	0.915	0.008	2.723	0.002	0.856	0.002	0.855
Earning	0.03	15.06	0.02	12.04	0.03	14.69	0.03	14.76
(SE)	0.001	0.393	0.002	0.866	0.001	0.353	0.001	0.353
Time dummy	y	y	y	y	y	y	y	y
Model Summary								
Number of Obs.	49169	33215	8832	5903	58001	39118	58001	39118
Prob.>F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Adjusted R2	0.1044	0.1711	0.1172	0.2248	0.1061	0.1778	0.1061	0.1788

**: statistically significant at 0.05 level

1. All variables are weighted by the size of index to correct heteroskedasticity problems.
2. Time dummy variables are included to reflect time trend in each program year.