W.E. UPJOHN INSTITUTE
FOR EMPLOYMENT RESEARCH

Upjohn Institute Working Papers | Upjohn Research home page

1-1-1995

Using Performance Indicators to Improve the Effectiveness of Welfare-to-Work Programs

Timothy J. Bartik
W.E. Upjohn Institute for Employment Research, bartik@upjohn.org

Upjohn Institute Working Paper No. 95-36

Citation

Bartik, Timothy J. 1995. "Using Performance Indicators to Improve the Effectiveness of Welfare-to-Work Programs." Upjohn Institute Working Paper No. 95-36. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. https://doi.org/10.17848/wp95-36

This title is brought to you by the Upjohn Institute. For more information, please contact repository@upjohn.org.

Using Performance Indicators to Improve the Effectiveness of Welfare-to-Work Programs

Authors
Timothy J. Bartik, *W.E. Upjohn Institute for Employment Research*

Upjohn Author(s) ORCID Identifier
https://orcid.org/0000-0002-6238-8181

This working paper is available at Upjohn Research: https://research.upjohn.org/up_workingpapers/36

Using Performance Indicators to Improve the Effectiveness of Welfare-to-Work Programs

Upjohn Institute Staff Working Paper 95-36

Timothy J. Bartik
Senior Economist

W. E. Upjohn Institute for Employment Research
300 South Westnedge Avenue
Kalamazoo, Michigan 49007
(616) 343-5541

I appreciate the secretarial assistance of Claire Vogelsong and Ellen Maloney with this paper. I would also like to thank Chuck Johnson, Kevin Hollenbeck and Bridget Timmeney for helpful conversations about this paper, and Craig Schreuder for help in getting information on the JTPA performance system. Preparation of this paper was funded in part by the Minnesota Department of Human Services. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author and do not necessarily reflect the views of the Minnesota Department of Human Services.

Using Performance Indicators to Improve the Effectiveness of Welfare-to-Work Programs

Timothy J. Bartik

This paper argues that it is feasible to develop good indicators of the performance of a particular welfare-to-work program, office, or contractor. Performance indicators can motivate local offices, contractors, and staff to be more effective in achieving the program's goals. Performance indicators can provide information on what program strategies lead to the greatest long-run success. To be most useful, performance indicators must be simple and timely and control for factors other than the program's effectiveness that influence whether welfare recipients "succeed."

Program managers and policy makers would like to know the "value added" resulting from a welfare-to-work program. This "value added" is the difference in welfare recipients' lives due to the program, compared to an alternate world in which the program was nonexistent. The program-induced differences of most interest will depend upon our social goals, but might include earnings gains, reduced welfare dependence, improved self-esteem, and better job skills.

The value added from a welfare-to-work program is hard to measure because it is costly and time-consuming to figure out what would have happened if the program did not exist. The change in earnings or welfare benefits for welfare recipients, from before to after the program, will be a poor measure of program effects. Many welfare recipients are suffering from temporary problems. On average, a typical group of welfare recipients will over time reduce their welfare dependence and increase their earnings, even without any special assistance. For example, in the Riverside County, California welfare-to-work experiment, the welfare recipients randomly assigned to the "control group" that did not receive any special services—more than tripled their earnings over the three-year period after the experiment's start, and less than half were still receiving welfare after three years (Riccio et al., 1994, pp. 323-324).

Social experiments—in which individuals are randomly assigned to receive services or be denied services—can measure the value added of a program. But using social experiments to monitor the performance of local welfare offices requires perpetually running social experiments in every local welfare office. Such widespread experimentation is unlikely. Among other concerns: is it ethical to deny services to some "control group" of welfare recipients on a regular ongoing basis in every local welfare office to help in program management?

Performance indicators for a welfare-to-work program are imperfect proxies for the program's value added. Performance indicators may take data on program outcomes and adjust the data so they are correlated with the program's value added. This adjustment may consider factors such as the characteristics of the welfare recipients served by the program or the health of the local economy. The advantage of performance indicators is that they do not require experimental methods. The disadvantage is that performance indicators may sometimes give a misleading impression of the value added of a particular program, welfare office, or contractor.

This disadvantage may be reduced by developing better performance indicators and by careful use of performance indicators.

Robert Behn, in his book on the Massachusetts ET programs, suggests that performance indicators for welfare-to-work programs may be used for three types of purposes: (1) justifying the overall program; (2) identifying how to improve the program; and (3) motivating better performance by managers and line workers associated with the program (Behn, 1991). Each of these purposes requires a different type of performance indicator. For justifying a welfare-to-work program, a perfectly rational policy maker would want to know the value added from the program, preferably using experimental data. In the real world, a program might be politically justified by any data showing that outcomes improved for welfare recipients, as one would expect even for an unsuccessful program. Good anecdotes might be as politically relevant as quantitative data.

To identify how to improve a welfare-to-work program, performance indicators must be positively correlated with program value added, but the value added need not be precisely measured. We need to identify which local offices, contractors, and staff members are more successful, but their value added need not be known precisely. Performance indicators for program improvement must be linked to information on the strategies of particular offices, organizations and staff members, so that we know why a particular program component is successful.

To motivate better performance by local offices and staff, we need performance indicators that are timely and understandable. These performance indicators must be easier to increase by increasing the value added of the program rather than by taking other actions that reduce value added. For example, performance indicators for motivational purposes should be easier to increase by improving earnings of welfare recipients, rather than by selecting welfare recipients for the program who will look good on the performance indicators ("creaming"). Welfare recipients with the greatest earnings prospects are not always those who will have the most value added from a welfare-to-work program. Finally, for any performance indicator to help motivate performance, program administrators must have the will to use the performance indicator to allocate some resources.

Performance indicators may also be distinguished by which part of the organization's performance is being measured, and by whom. Performance indicators may be used by federal officials to measure state performance, state officials to monitor local welfare office's performance, local offices to monitor contractors, and local offices or contractors to monitor individual staff. As one gets closer to the individual staff level of the welfare system, it becomes easier to use personal interaction and judgment to substitute for quantitative performance indicators. In addition, as one gets closer to the individual staff level, it will be more difficult to consistently interpret more complex performance indicators. Therefore, performance indicators should be simpler and less accurate as we get closer to the individual staff level, and should become more statistically sophisticated at higher levels in the welfare system.

Thus, we need a variety of performance indicators, ranging from quite simple, timely, but rough proxies for value added to more sophisticated, accurate, and complex approximations to value added. The questions addressed by this paper are twofold: (1) how can we develop better performance indicators that will be both feasible and closely correlated with value added, and (2) once such measures are developed, how should they appropriately be used for program management? These questions will be addressed based on experience with performance indicators in job training and welfare-to-work programs, and other previous research on welfare and the labor market.

PREVIOUS AND CURRENT EXPERIENCE WITH PERFORMANCE INDICATORS FOR WELFARE-TO-WORK AND RELATED PROGRAMS

There has already been some experience with performance indicators in welfare-to-work and job training programs. This experience provides some ideas for a good performance indicator system. Other ideas are provided by proposals for federal performance indicators for welfare- to-work programs and proposals for increasing private sector involvement in welfare-to-work programs.

JTPA: The Job Training Partnership Act

The most extensive performance indicator system for a social program is used by the principal federal job training program, JTPA (Job Training Partnership Act). When JTPA was enacted in 1982, the new legislation included requirements for performance standards, so that JTPA would both be more effective and seem more effective than CETA, the previous federal job training program. According to job training expert Burt Barnow, "CETA's poor image created an environment conducive to implementing a performance management system that would help make the new program more accountable. Along with the increased role of the private sector in overseeing the programs through the private industry councils, a formal performance management system was seen as an important tool for focusing local programs' attention more on outcomes" (Barnow, 1992, p. 290). Because of these concerns, the JTPA legislation required a performance management system, and even mentioned specific performance goals: placement and retention of trainees in unsubsidized employment, increases in wages and earnings, and reductions in welfare receipt.

The current JTPA performance management system operates at three levels: federal requirements for state governments, state standards for local "Service Delivery Areas" (SDAs), and SDA standards for contractors. The federal government sets performance standards, suggests procedures for adjusting those standards for different SDAs, allows states to reward well-performing SDAs, and requires states to overhaul poor-performing SDAs. Up to 5 percent of JTPA funds are set aside to be used by states to reward well-performing SDAs. States may further adjust the performance standards for different SDAs, and can choose specific rewards and sanctions for SDAs. SDAs generally contract out most actual delivery of training services to contractors. Most such contracts include performance requirements or incentives.

Current JTPA performance standards include, for adult programs, the employment rate and weekly earnings of ex-JTPA participants as of 13 weeks after exiting JTPA. Separate adult standards are set for all JTPA participants and for JTPA participants who were on welfare when entering the program. For youth, performance standards include whether the participants left the program with jobs, and whether participants acquired skills or educational credentials. The actual JTPA performance standards for 1995 are listed in Table 1.

Because JTPA performance standards for adults are follow-up measures, whether an SDA meets standards for a program year is based on outcomes for JTPA participants from the first three quarters of the current program year and the last quarter of the previous year. Follow-up data are derived from phone surveys of ex-JTPA participants. The JTPA program provides detailed standards for adjusting these phone survey results for "non-response bias" if the response rate differs by more than 5 percent between JTPA participants who leave JTPA with a job and those who leave without a job.

Table 1
JTPA 1995 Standards
for Adults (Title II-A) and Youth (Title II-C)

Title II-A Adults

- Adult follow-up employment rate, 59%
- Adult weekly earnings at follow-up, $245
- Welfare follow-up employment rate, 47%
- Welfare weekly earnings at follow-up, $223

Notes: "Follow-up" is 13 weeks after leaving JTPA program. Employment must be greater than 20 hours per week. Average weekly earnings only based on those employed at least 20 hours per week. Performance measures only include JTPA clients who received some service beyond objective assessment.

Title II-C Youth

- Youth entered employment rate, 41%
- Youth employability enhancement rate, 40%

Notes: Entered employment is whether entered employment at termination from JTPA. Employability enhancement is any one of five possible attainments, including demonstrated skill competency attainments or completing certain education or training activities.

The JTPA performance standards in Table 1 are adjusted by states for each SDA, based on the client mix and local economic conditions. The performance standard for a SDA may be below or above the standards listed in Table 1. States may either use a U.S. Department of Labor

model to adjust standards, or develop their own model. The Department of Labor model is based on a regression using historical data on all SDAs to predict their performance based on client mix and local economic conditions. Variables are only included in the DOL adjustment model if they are good predictors and have the expected effects. Using this regression, DOL sets standards so that 75 percent of all SDAs would have met the standards during the historical period.

Figures 1 and 2 (available only in hard copy) show the actual worksheets used by SDAs to figure out their own performance standards for JTPA participants who are welfare recipients. The worksheets are based on the Department of Labor's regression model for predicting SDA performance; the "weights" in the worksheets are the regression coefficients from this model. Most of the adjustment factors are client characteristics. The employment rate standards for welfare recipients are adjusted downward for SDAs with more clients who are female, school dropouts, black, offenders, lacking in significant work history, long-term unemployed, or long-term welfare recipients. These performance standards are adjusted upwards in SDAs in which a greater proportion of JTPA participants have attended some post-secondary educational institution. The employment rates tandards are adjusted downward for local areas with unemployment above the national average. As figure 1 shows, for areas with an unemployment rate 10 points above the national average, the employment rate standard is adjusted downward by 5.1 percent, from the regular standard of a 47 percent employment rate to 41.9 percent.[1] For the weekly earnings performance standard, a similar list of SDA characteristics is used to adjust the performance standards, with some reasonable modifications. As one might expect, the weekly earnings standard is adjusted more for factors affecting wages: reading skills, local earnings in retail trade, local poverty rates, and the percentage employed in manufacturing. The adjustment also reduces the weekly earnings standard for SDAs with lower population density (SDAs in rural areas) or SDAs with a high ratio of employees by place of work to residents who work (SDAs in central cities). These adjustments have the effect of relaxing the earnings standards for central cities and rural areas compared to suburbs.

As shown on the worksheets, SDAs may for planning purposes calculate their performance standards for the next program year, based on the client mix and local economic conditions they expect (these are the "Plan" type of standard). The final standards used for rewards or sanctions, however, are based on the actual client mix and economic conditions (the "Recalculated" standard). Thus, SDAs do not know precisely the standards they must meet for a particular program year until after the year is over.

SDAs often have performance-based contracts with contractors providing training to JTPA participants. These contracts may base payments to contractors on the number of participants placed in jobs, or a particular contractor may be dropped if certain standards are not met. Performance-based contracts with contractors to SDAs usually have no adjustment for the contractor's client mix.

---

[1]This calculation assumes that all other of the "weights" used to adjust the national standards for the SDA are at the national average.

A common concern is that JTPA performance standards may encourage SDAs and their contractors to engage in "creaming" (selecting JTPA participants who will do well on performance standards, even if those participants are not those most in need or those for whom JTPA has the greatest value-added). According to one researcher, "many welfare officials believe that largely because of the JTPA performance standards system, SDAs have been reluctant to enroll AFDC recipients, especially those with major labor market barriers" (Barnow, 1992, p. 305). The structure of JTPA encourages creaming. SDAs have some control over whom to serve because JTPA is a voluntary program that only serves a few of those eligible. Although the adjustment model encourages SDAs to enroll hard-to-serve clients, conversations with SDA managers suggest that many do not fully understand the adjustment model (Barnow, 1992, pp. 299-300). In addition, despite the adjustment model, SDAs have significant incentives to cream, using the many client characteristics that inevitably are excluded from the model—client motivation, for example. Some SDAs delay formal enrollment in JTPA, for example, until an individual has already received some training services, a practice that selects more motivated clients. Finally, contractors to SDAs, who often exert some influence over who is enrolled, generally do not benefit from relaxed standards if they enroll more hard-to-serve clients.

Empirical evidence suggests that JTPA does engage in some creaming. Cragg (1993) finds that in states with greater rewards for exceeding standards, SDAs enroll individuals with more work experience. This creaming reduces JTPA value added because Cragg also finds that more experienced individuals receive a lower earnings gain from JTPA. Dickinson and West (1988) find that in states that place more emphasis on exceeding standards, SDAs enrolled fewer welfare recipients or minorities. Dickinson and West also find that in states that adjusted standards for client mix, compared to states that did not do so (a former state option), SDAs served more adult welfare recipients and dropouts. Anderson et al. (1993) find that JTPA clients are on average significantly more educated than the population of JTPA-eligibles.

Although JTPA's performance standards may result in some creaming, these standards may also increase JTPA's value added. Cragg finds that SDAs in states with stronger performance incentives produced greater earnings gains for JTPA participants. This increase in SDA effectiveness more than offset the decrease in performance induced by SDAs creaming the more experienced individuals who gain less from JTPA. Dickinson and West found that states that emphasized exceeding performance standards spent more per JTPA terminee, but spending per client entering employment was no higher. Finally, the recent experimental evaluation of JTPA found it to be one of the few job training programs that have increased earnings for adult males (Bloom et al., 1994).[2] It is possible that part of JTPA's unusual success with adult males may be attributable to performance standards.

[2]JTPA also has demonstrated significant earnings gains for adult women, but so have many other training programs. JTPA has not demonstrated success for youth, but few social programs have.

Performance Standards in Two Well-Known Welfare-to-Work Programs: ET Choices and Riverside County

Two of the best known welfare-to-work programs, ET Choices and Riverside County, have aggressively used performance standards. ET Choices (Employment and Training Choices) was a "soft workfare" program in the mid-1980s in Massachusetts. ET helped welfare recipients overcome barriers to work via education, child care, transportation, and other assistance.[3] The Massachusetts Commissioner of Public Welfare in the mid-1980s, Charles Atkins, had a management style that emphasized "ten annual goals" for the department. All ten goals were supposed to be quantifiable. Atkins suggested that department staff should spend 75 percent of their time on these ten goals, 25 percent on everything else.

For the ET program, Atkins initially set a goal for placements. Over time, these standards were modified to emphasize "quality" placements, which were defined as full-time at greater than $5 per hour (equivalent to $6.76 in 1994 dollars). ET placement goals for each local office were set by dividing the state goal up among local offices according to their caseload size. No adjustment was made for the local client mix or the local economy. Annual goals for each office were then translated into monthly goals.

Atkins and the Department took several steps to increase the motivational effect of local office goals. Atkins met monthly with the directors of the local welfare offices and handed out two summary lists: local offices that had achieved their monthly ET placement goals, and local offices that had not. The Department also promised to push for upgrading the classification of line staff jobs and reducing caseloads if the state agency achieved its overall placement goals. In addition, according to the Department's Deputy Commissioner, "we said to the local offices: 'You get the error rate down, you get ET placements up, and we're going to get you first-class space.'" (Behn, 1991, p. 64). Finally, the Department fired or demoted a few local office directors for poor performance. One local office director in the Boston area was widely regarded as a poor performer, but had strong political connections. According to the Department's Associate Commissioner for Field Operations, because this director was "literally the worst director in the state, …when he was demoted and it stuck, people understood how serious we were." (Behn, 1991, p. 62)

What were the results of ET? ET was never studied by rigorous experimental methods, so there is some dispute about its value added.[4] But ET does appear to have successfully changed local welfare offices away from a focus on lowering the error rate and toward a focus on getting people into jobs. According to welfare expert Richard Nathan, one of the keys to ET's success at changing the focus of local welfare offices was performance standards: "Contrary to my expectation, I found that the exhortation to 'get people off of welfare' and the aggressive use of

[3]The source for my description of the Massachusetts ET program is the insightful book by Behn (1991).

[4]Nightingale et al. (1991) give a generally positive evaluation of ET, whereas O'Neill (1990) is generally negative.

performance goals focused on this objective brought greater pressure to bear on the bureaucracy under the ET program in Massachusetts than was the case under the seemingly more obligatory work-welfare programs of other states." (Nathan, 1993, p. 126)

Riverside County, California's GAIN program (Greater Avenues for Independence) for welfare recipients has received much attention because experimental evidence shows that Riverside is far more successful in increasing earnings than the other five California counties experimenting with the program. Riverside's GAIN program had three-year impacts on earnings gains and welfare reductions for single mothers that were over twice the average effects for all the counties (Riccio et al., 1994, p. xxviii).

An important issue is why Riverside has been so much more successful. With a sample of only six counties, it is difficult to settle this issue. One difference between Riverside and the other five counties is in goals. Ninety-five percent of surveyed Riverside case managers said that their agency viewed quick job entry for welfare recipients as more important than education and training, whereas less than 20 percent of case managers in other counties felt their agency emphasized quick job entry (Riccio et al., 1994, p. 54). An additional difference is that every local welfare office in Riverside had its own job developer who worked with local employers to place welfare recipients in jobs, whereas other counties put less emphasis on job development.

For this paper, the most relevant possibility is that Riverside's greater success is due in part to how the county chose to enforce its goals. Case managers, supervisory units, and district offices in Riverside County all have job placement goals, unlike the other five counties. Meeting the goals is an important part of the job performance evaluation of individual staff. Riverside staff report that standards for job placements are not difficult to meet, but they feel some pressure to meet and exceed goals. Goals are not adjusted for client mix of different case managers or local offices. Despite the lack of adjustment for client mix, researchers studying the Riverside program generally feel that there is no evidence of creaming by Riverside staff. Creaming may be avoided because staff members are expected to involve their entire caseload in various required activities. (Riccio et al., 1994, p. 54).

Current Federal Policy Toward Performance Standards for JOBS

The federal government has been surprisingly slow in adopting outcome-oriented performance standards for JOBS, the welfare-to-work program created by the Family Support Act of 1988. JOBS was supposed to represent a shift from the "eligibility and accuracy" culture of welfare offices, in which the focus was on reducing errors in making welfare payments, to a focus of welfare offices on increasing client self-sufficiency. The federal government might have promoted this cultural shift using performance standards, but has not done so. Under the Family Support Act, the U.S. Department of Health and Human Services was supposed to produce recommendations for JOBS performance standards by October 1993. This requirement was postponed a year, and in October 1994 HHS issued a plan for how JOBS performance standards will be developed in the future. The October 1994 document suggested that HHS will develop JOBS performance measures by October 1996, and performance standards by October 1998. With

congressional welfare reform looming, it is unclear whether any of HHS's plan for performance standards will be implemented.

In the October 1994 report, HHS argued for moving slowly on performance standards because outcome-based performance measures may be misleading indicators of program value-added. This conclusion is based on research by the Manpower Demonstration Research Corporation (MDRC), which is well-known for its research on welfare-to-work programs. Table 2 summarizes the MDRC research results cited in HHS's report. Often, differences across programs in average outcomes show no correlation with the value added of the program. Within the same state and program, variation across different counties sometimes seems to reveal some true variation in value added, but that compare a state's current performance to some baseline period. This approach assumes that a state's baseline performance helps control many unobservable features of a state's welfare system, economy, and mix of welfare recipients. HHS mentions that performance standards based on "improvements" from the baseline might be regression adjusted for client mix and local economic conditions. Possible performance measures mentioned include: percent of the AFDC caseload that receives aid for more than a specified period; the JTPA performance measures; increases in employment and earnings of program participants after leaving the JOBS program; retention of JOBS participants in unsubsidized employment.

Some observers believe that the lack of federal leadership in developing outcome-oriented performance standards for JOBS has hampered the program. According to the U.S. General Accounting Office, "[JOBS] programs are generally not well focused on recipients' employment as the ultimate goal. Our recent nationwide survey of local program administrators revealed that JOBS programs have generally not forged the strong links with local employers that may be important in helping AFDC recipients gain work experience and find jobs. Many factors hamper the development of these ties to the workplace, including the JOBS performance measurement system. This system holds states accountable for the number and type of AFDC recipients participating in JOBS activities, but not for the number who get jobs or earn their way off AFDC. Thus, programs may focus more on preparing participants for employment than on getting them jobs. In fact, the number of JOBS participants who get jobs or leave AFDC annually is unknown." (GAO, 1994).

State Initiatives for Performance Standards for JOBS

Despite the lack of federal leadership, many states have moved ahead and developed their own performance standards for their JOBS program. Table 3 summarizes a recent GAO survey asking state JOBS administrators about their use of performance measures and standards. States have developed a variety of performance indicators: whether people get jobs, the quality of jobs, the impact of these jobs on AFDC payments, whether the job or the resulting AFDC exit is really long term, and educational and training achievement under JOBS. A surprising number of states have performance standards with some consequences of whether the standard is met.

Table 2

PANEL A: Program Outcomes Across States Often Do Not Seem Highly Correlated with True Value Added Of Program

Welfare to Work Program	Percent Employed of Program Group at End of Year One	Percent Employed of Control Group at End of Year One	"Value Added" of Program (Program Group minus Control Group)
Arkansas	20.4%	16.7%	+3.7%
Baltimore	34.7%	31.2%	+3.5%
San Diego SWIM	34.7%	26.9%	+7.7%
Virginia	34.7%	31.0%	+3.8%

PANEL B: Program Outcomes Across Counties Within California's GAIN Program Seem Somewhat Correlated, But Inconsistently, with True Value Added of Program

California County	Percent Employed of Program Group at End of Year Two	Percent Eemployed of Control Group at End of Year Two	Value Added of Program (Program Group minus Control Group)
Alameda	24.5%	18.1%	+6.4%
Los Angeles	19.0%	15.7%	+3.3%
Riverside	35.2%	24.0%	+11.2%
San Diego	32.5%	26.4%	+6.0%

PANEL C: Program Outcomes Within A Particular County are Often Greater for Less-Needy Groups for Whom Program Value Added Is Less

Riverside Sub-Group	Average First Year Earnings of Program Group	Average First Year Earnings of Control Group	"Value Added" of Program (Program Group minus Control Group)
AFDC Applicants	$2832	$2038	+$794
Short-term AFDC Recipients	$2577	$1671	+$906
Long-term AFDC Recipients	$2097	$983	+$1113

Source: U.S. Department of Health and Human Services, 1994, pp. 21-23.

Table 3
Summary of Results from GAO Survey
of 50 State JOBS Administrators

Performance indicator or outcome area	Number of states collec-ting data on this as performance indicator	Number of states with state-wide performance standard for this outcome	Number of states believing this indicator should monitor JOBS at national level
Number of JOBS participants entering employment	49	27	46
Hourly wages earned at time of hire	42	16	35
Number of participants no longer receiving AFDC due to employment	33	9	44
Job retention rate after a specified length of time	26	2	40
Number of participants with reductions in AFDC grants due to employment	24	3	39
Educational/training achievement, such as graduation rates	24	14	36
Number of participants entering employment that provides health insurance coverage	16	3	29
Rate of return to AFDC, after specified time, for those who left AFDC due to employment	14	3	41
Hourly wages earned at a specified time period after initial hire	11	1	20
Expected weekly earnings at time of hire	11	2	13

Source: GAO, 1995

A recent report by the American Public Welfare Association summarizes the performance standard systems of six states believed to be the most aggressive in adopting such standards. Table 4 summarizes the performance measures used in these states, how the standards are set for each local office, and the consequences of whether standards are met. These six states also vary widely. Some states adopt standards that are narrowly focused on a few employment goals, whereas other states consider larger numbers of goals and include "intermediate goals" leading to employment, such as skills acquisition. Some states set goals based on each local office's caseload, others negotiate goals with each local office, and still others adjust goals based on the JTPA factors of client mix and local economic conditions. Some states base large chunks of local office funding on whether standards are met, whereas in other states the rewards or sanctions are informal.

There is no clear evidence on which state's approach to performance standards is most effective. States have different goals for their performance standard system. In some states, there are so many goals and so many different ways of meeting goals that it is questionable whether the standards can help redirect the state's welfare system. In states such as Oregon, where goals are negotiated and there are no formal rewards or sanctions, one wonders whether much pressure is placed on local offices to do anything differently. States such as Maryland, Pennsylvania, and Wisconsin, with a few employment-oriented goals and stronger rewards and sanctions, made stronger statements to the APWA interviewer about the standards helping to redirect the welfare system. This is a plus if one believes that these employment goals are appropriate for the welfare system, but a problem if one believes that the welfare system should be focused on skills acquisition leading to better jobs.

Recent Proposals for Privatizing the Welfare-to Work System

Some recent proposals go beyond simply setting performance goals for welfare-to-work programs, to privatizing more components of the welfare-to-work system. Examples include recent proposals by the Mayor of Indianapolis, and the chief of staff to the Mayor of Milwaukee. These proposals are partly inspired by the America Works program.

America Works is a for-profit firm providing welfare-to-work services in New York City and more recently in Indianapolis. America Works provides welfare recipients with a week-long "training" in which recipients are briefed on job search and work habits, along with some office skills training. Recipients are dropped from the program if they are late for any class. America Works uses job developers who help find jobs for the welfare recipients. America Works typically places its clients with employers for a four-month probationary period intended to lead to permanent employment. During this probationary period, an America Works staff member meets weekly with the client, and every other week with the client's supervisor, to help resolve problems. The firm is paid in part for performance. In New York City, America Works is paid $980 per client for providing the week-long training, $3855 two days after a client has been hired

Table 4
Summary of Six States' Performance Standards for JOBS

State	JOBS Performance Measures	How Standards Set and Adjusted	Consequences of Whether Standards Met
Kansas	Percentage of JOBS participants completing GED or high school; percentage of JOBS participants completing training, percentage of JOBS participants obtaining employment; average hourly beginning wage; average number of JOBS participants; AFDC case closures due to employment.	All percent goals translated into absolute numbers at local level. Allocation to local office based on share of AFDC population.	No explicit rewards or punishments
Maryland	For SDAs that run JOBS: Number of JOBS participants; number of JOBS participants who leave due to increased earnings; percentage of JOBS participants in unsubsidized employment who are employed in 26th week following termination; number of JOBS participants who achieve education/training objectives in employability development plan.	Standards for number of participants and number who leave JOBS due to earnings are based on local share of JOBS funding; retention rate goal 50 percent for all.	If satisfy participation goal, JOBS earnings exit goal, and retention goal, become eligible for incentive funds.
North Carolina	18 outcome and process goals, including: numbers of various groups to return to school, get GED, get post-secondary degree or skills certification, number who enter full-time employment; percentage who stay employed at least 9 of 12 months; percentage of recidivists to AFDC after leaving.	Percentage goals uniform across counties; county's share of state JOBS funds used to set other goals	Counties divided into exemplary, commendable, and needs improvement. Commendable must meet federal JOBS requirements and 3 of 5 process goals and 5 of 10 outcome goals. Exemplary and commendable receive extra JOBS funds. "Needs improvement" required to make changes.

Table 4
(Continued)

State	JOBS Performance Measures	How Standards Set and Adjusted	Consequences of Whether Standards Met
Oregon	18 performance goals, including average time on welfare, number getting jobs, percent returning to welfare within 18 mos., percent receiving child support, cumulative avg. wage of JOBS participants, percent who increase educational competency, cumulative avg costs per job placement.	Each local office sets own goals. Goals are supposed to be higher than agency's baseline, but achievable.	No formal rewards or sanctions, but some state or regional intervention to make changes will occur if office's performance is below baseline.
Pennsylvania	Standards only for clients referred to SDAs. Standards are adult welfare-follow-up employment rate, and adult welfare weekly earnings at followup (followup is 13 weeks). Also have six goals: entered employment rate, average cost per participant, cost per placement, wage at placement, and rate of placement in jobs with medical benefits.	The "standards" are adjusted for local conditions using JTPA regression model; goals are uniform statewide	SDAs exceeding one or both standards are eligible to receive share of incentive grant fund. If don't meet goals, can be required to develop correction plan.
Wisconsin	Statewide, "standards" only for AFDC-UP. For AFDC-UP except for Milwaukee, only 45 percent of budget is fixed; remainder of budget based on fixed dollar pay-ments for number of work-active participants, and num-ber getting full-time or part-time jobs. For Milwaukee, all JOBS participants randomly divided between two teams of organizations. Future alloca-tions of budget and clients between teams based on points received for various work activities, with greatest points for full-time job.	Essentially standards are uniform. Within Milwaukee, two teams should have identical client mix.	Rewards and penalties extremely strong: essentially all of Milwaukee's JOBS budget depends on performance, and 55 percent statewide of AFDC-UP budget.

Source: APWA, 1994.

for a permanent job (after the four-month probation), and the remaining $650 after that permanent job has lasted three months.[5]

Although America Works claims high placement and retention rates, these high rates may occur because America Works is creaming from the welfare population, not because the program has a high value added. Kicking tardy clients out of the program helps select the more motivated and organized welfare recipients.

Stephen Goldsmith, the Mayor of Indianapolis, has suggested that welfare-to-work could be reorganized as a competitive, performance-based system with many different providers. Different private organizations could be "chartered" to provide welfare-related services. Payments to each competing office could be based on job placement or on educational achievement. (Goldsmith, 1994)

David Riemer, the chief of staff to the Mayor of Milwaukee, has proposed turning over the entire welfare system to competing private organizations, in what he calls an "Employment Maintenance Model" (Riemer, 1994-95). Welfare recipients would be randomly assigned to different "employment maintenance organizations." Each EMO would have to guarantee that their clients would be brought above the poverty line through transfer payments, community service jobs, or private sector jobs. EMOs would be reimbursed by the government for 75 percent of their transfer payment costs, and 100 percent of their costs for community service jobs. EMOs would receive an incentive payment of 5 percent of earnings on unsubsidized jobs by all EMO clients for five or ten years after clients entered the EMO. Under this payment system, EMOs make money by reducing transfer payments and increasing placements in long-term unsubsidized jobs. EMOs would have great flexibility in taking any action they think would increase long-term unsubsidized employment for their clients. EMOs could deny welfare payments to clients who "unreasonably" refused a job offer, with arbitrators ruling on the fairness of any such benefit denial.

Riemer's plan has the advantage of clearly providing large incentives to accomplish the goal of unsubsidized long-term employment for welfare recipients. His proposal also raises concerns in three areas: (1) feasibility; (2) affordability; (3) fairness. First, is it really feasible for private organizations to make any money under his scheme? Second, would this scheme be affordable even if it worked? Finally, is it legitimate to randomly assign individuals to what appear to be almost different laws and regulations?

[5]This description of America Works is based on Rosenberg and Stern (1993) and Fein (1994).

SOME SUGGESTED GUIDELINES FOR PERFORMANCE INDICATORS FOR WELFARE-TO-WORK PROGRAMS

1. *We need to agree on a few easy-to-understand performance standards that are frequently measured in order to really focus the attention of staff people in the welfare system on the social goals we consider to be important.*

Robert Behn's book-length analysis of the Massachusetts ET program concludes that the program's success in reorienting the focus of the welfare system was crucially dependant on having a simple goal:"Despite the numerous problems with using any single goal to manage for performance, the selection of a single, simple, specific goal is critical. It is essential, of course, that the goal mesh with the overall mission. But the other characteristics of the goal are not as important. Indeed, there is a distinct advantage of using a very simple goal (e.g., the number of job placements); for once the manager starts making adjustments (e.g., to reflect the characteristics of the person being placed or the characteristics of the region), people will begin quibbling over how large that adjustment should be. And if the manager places primary emphasis on the mission and is constantly alert to the distortions created by the goals, the exact nature of the goals chosen is not as important as is the use of them." (Behn, 1991, p. 79)

Additional evidence on the importance of simple goals is provided by the JOBS program in Maryland. Maryland's JOBS program focuses on three simple goals: the number of JOBS participants, the number of JOBS participants who leave JOBS due to increased earnings, and the job retention rate for JOBS participants who obtain employment. The American Public Welfare Association (APWA) interview with Maryland JOBS staff suggests that these simple goals are important to the program's success: "According to Maryland JOBS staff, the use of absolute numbers presents local programs with a precise and undisputable target to work towards. In addition, JOBS staff suggested that raw numbers are more difficult to manipulate and easier to adjust than rates, …JOBS staff indicate that their use of incentive funding has had a positive impact on program performance. First and foremost, the goals send a clear and simple message to all state and local staff that the program is designed to help clients find work and remain employed." (APWA, 1994, p. 48).

Finally, as discussed previously, the simple Riverside job placement goals for all employees help make Riverside's program a success.

The problem with simple goals is that we may lack social consensus on goals for welfare-to-work programs. The goals used by state JOBS programs fall into at least five categories: (1) job placement; (2) quality of the job placement (wage, insurance benefits, etc.); (3) reducing welfare payments or getting off welfare; (4) some measure of the length of time for which the job or the welfare reduction lasts; (5) overcoming employment barriers (e.g., lack of job skills). All these goal categories may help achieve long-run economic success. But these goals may often conflict. Pushing a welfare recipient into low-wage employment immediately may come at the expense of acquiring skills that could help that person get a better job later. On the other hand,

putting a welfare recipient in a training or educational program may come at the expense of not receiving valuable on-the-job experience.

We can create performance standards that reflect all these goals, but the result is a more complex performance management system, which probably has less immediate effect on motivating a local welfare office and staff to move in any particular direction. The more we can agree on a few, simpler goals, the more likely we are to move quickly in some direction. Of course, moving aggressively in the wrong direction may not be a good thing.

The problem is that choosing social goals for a welfare-to-work program is as much an issue of values as of facts, and even the facts are in dispute. We do not know whether a typical welfare recipient benefits more in the long-term from taking any job now, or from getting more education now to get a better job later. Beyond that factual issue, there is disagreement over society's obligations to welfare recipients and welfare recipients' obligations to society. Does society have a right to insist that every welfare recipient take any available employment opportunity? Do welfare recipients' have a claim on social support for skills acquisition, and if so, how extensive is that claim?

2. *Total fairness is not necessary for a performance indicator system to be a useful motivator for staff and local welfare offices.*

No performance management system will be perfectly fair, because it is impossible to measure perfectly and quickly the value added of each welfare office, contractor, and employee. Any system will sometimes reward an individual or organization whose true performance is not as good as others, and will sometimes sanction an individual or organization whose true performance is no worse than others that are not sanctioned.

But what we need is not a perfect performance measure, but one that the individual or organization finds it easiest to increase by improving program value added. It should be hard to increase the performance indicator by manipulating client mix, program operations, or program data in ways that do not increase value added. Whether a performance indicator is useful for motivation depends not only on its accuracy as a measure of value added, but on aspects of program design: What control do local offices or staff members have over client mix or details of program operations? Can state managers monitor ways in which local offices or individuals may subvert the purposes of the program to artificially boost performance?

An analogy to performance indicators in the private sector is useful. Private sector organizations often base performance standards and incentives on output measures that only correspond imperfectly to the value added that an employee has to profits. For example, a salesman may have sales quotas. The volume of sales associated with a salesman does not necessarily have a strong relationship with the salesman's impact on the firm's profits. For example, if a company exerted no control over price charged by its salesmen, and just blindly paid bonuses based on units sold, the company would quickly go bankrupt. But if the company exerts some control on prices charged by salesmen, then salesmen who are trying to beat their

sales quota are also likely to increase profits. This is true even if some salesmen find it easier to meet their quotas due to their assigned customer base.

If performance indicators are perceived as grossly unfair, this may create poor morale in local welfare offices. Performance indicators need to meet some minimum fairness standard. Welfare office staff should perceive that those offices and staff who fail to meet standards probably aren't performing on average quite as well as those who are. The perceived fairness of a performance management system also depends on the size of its rewards and sanctions and how hard the standards are to meet. If the system is designed so that rewards of meeting standards are only modest, and the standards are not unreasonably hard to "pass," there is likely to be less resentment over the system. If the system is designed so that next year's entire budget depends on meeting one simple standard, we may be putting too much stress on an imperfect indicator. Finally, the perceived fairness of a performance management system may depend on whether the system allows for special factors that are not captured in any quantitative adjustments to the standards. For example, under JTPA, Governors are allowed to adjust standards based on identifiable special factors that explain why an SDA may find it difficult to meet a particular standard.

3. *Creaming, of the right type, is necessary and good.*

Creaming—or selecting particular welfare recipients who may be better off for more intensive services—may **sometimes** make sense. One important advantage is that **some types** of creaming may increase the effectiveness of job development. Job development has been important in many successful work and training programs: JTPA, the Riverside County GAIN program, and the programs run by the Center for Employment Training (CET) in San Jose California.[6] Job development is finding employers who might hire clients of the program, and aggressively marketing clients of the program to these employers. Successful job development requires some creaming, because it is difficult to market social service clients to employers unless there is some reasonable likelihood that these clients will succeed on the job. The National JTPA study provides some direct evidence that job development can be successful. According to the National JTPA study, which relies on an experimental methodology, the on-the-job training (OJT) component of JTPA, which includes a heavy emphasis on job development in finding subsidized training slots for clients, is one of the most successful components of JTPA (Bloom et al., 1994).

Job development seems like such a small intervention in the labor market that it is surprising that it is so successful. One reason for job development's success is that employer information on new hires is poor, resulting in costly turnover and employer use of imperfect proxies for productivity in making hiring decisions. According to John Bishop's survey of smaller businesses, "Managers of small and medium firms were very often unpleasantly surprised by the performance of new hires. After six months on the job, more than one-quarter of new hires

[6]For more on CET, see Zambroski and Gordon, 1993, and Caves et al,1993. U.S. Department of Labor (1995) provides a useful summary of evaluations of training programs.

were producing less than 75 percent of what was anticipated when they were hired."(Bishop, 1993, p. 336) In reaction to these common problems, employers do a great deal of hiring though employees or friends of the employer. This increases the average productivity level of those hired, but makes it more difficult for persons without job contacts to be hired.

A highly undesirable employer response to imperfect information is to discriminate against black or lower-class jobseekers, based on preconceptions about their likely productivity. Kirschenman and Neckerman's interviews with 185 Chicago employers give some insight into employer discrimination. According to one Chicago manufacturer interviewed, "I would in all honesty probably say there is some [discrimination against blacks] among most employers. I think one of the reasons, in all honesty, is because we've had bad experience in that sector, and believe me, I've tried. And as I say, if I find—whether he's black or white, if he's good and, you know, we'll hire him. We are not shutting out any black specifically. But I will say that our experience factor has been bad. We've had more bad black employees over the years than we had good" (Kirschenman and Neckerman, 1991, p. 212). Besides race, employers sometimes also discriminate based on social class, neighborhood, and gender. According to another Chicago employer, "We have some black women here, but they're not inner city. They're from suburbs and…I think they're a little bit more willing to give it a shot, you know, I mean they're a little bit more willing [than black men] to give a day's work for a day's pay."(p. 217)

Because employers have poor information about job seekers and prejudices against many disadvantaged groups, job development can improve job prospects of disadvantaged persons by screening to identify the more productive group members, and marketing these more productive individuals to employers. In this way, job development can be a positive approach that may help reduce employer discrimination against disadvantaged groups. Of course, vigorous enforcement of laws against racial discrimination will also be necessary. It is also important to monitor job developers to make sure that their practices are not catering to employers' racial prejudices.

The productivity and information advantages of screening also help explain the growth in temporary help agencies. Many firms now only hire workers permanently after they have already worked at the firm for a while through a temp agency. Unlike job developers, however, temporary help agencies have no mission to serve the disadvantaged. Screening through temp agencies may sometimes reinforce rather than alleviate the job hiring practices that lead to labor market problems for the disadvantaged.

Creaming for job development is facilitated by the common practice in employment and training programs of dividing clients into "job-ready" and "not job-ready," with the not job-ready diverted to training. Job developers have a better "product" to market if they are only marketing persons who are job-ready or who have completed training.

Beyond its usefulness for job development, creaming may also be of more general benefit in training programs for disadvantaged persons, because some disadvantaged persons may have too many problems to be helped with the resources at hand. Friedlander's analysis of who benefits from welfare-to-work training programs suggests a "triage" pattern in the earnings benefits from

these programs (Friedlander, 1988). Friedlander used data from various welfare-to-work experiments, and divided the welfare recipients in the experiments into subgroups based on the extent of prior welfare history and prior earnings. Among his findings:

> *The groups that were most job-ready and least welfare-dependent, as defined by previous work and welfare experience, had below-average program impacts [on earnings] that were generally not statistically significant…*
>
> *Earnings impacts were found most consistently for individuals in the mid-dependency tier…*
>
> *Earnings impacts were not found consistently for subgroups in the most dependent tier…*
>
> *Statistically significant welfare savings were found for several of the more dependent subgroups, but this was not consistent across samples. Welfare savings were also found for some of the mid-dependency groups, although generally these were not statistically significant…* (Friedlander, 1988, pp. xvii-xviii)

Therefore, if our social goal is to increase the earnings of welfare recipients, these findings suggest a "triage" strategy: focus services on welfare recipients whose problems seem of "average" difficulty. Those with the fewest problems will probably get off welfare and significantly increase earnings without any help, and those with the most major problems are less likely to be affected by the typical welfare-to-work program, which will have modest, limited resources. Welfare-to-work programs should cream among the welfare recipients who most need the services. We do not necessarily want to target welfare-to-work services on the most dependent group.

4. *The wrong type of creaming can be discouraged.*

The wrong type of creaming is selecting welfare recipients for services who will score better on some performance indicator but will receive smaller than average or no value added from the welfare-to-work program. This type of creaming is discouraged or eliminated if the welfare offices or staff persons whose performance is being judged do not control selection of the group whose success is being measured. This occurs quite naturally for welfare-to-work programs if we judge performance based on what happens to the entire welfare caseload, or based on some subset of welfare population that has certain objective characteristics (kids over certain age, etc.). In either case the local office or staff cannot pick or choose the welfare recipients whose success will determine their score on some performance indicator. Because the office's or staff's performance will be judged based on the success of some externally defined group, the most straightforward way for the office or staff to increase the performance indicator is to increase the average value added of the entire group.

This point was made in the October 1994 HHS Report on welfare-to-work performance indicators:

> *…from the Department's perspective, of primary importance in specifying "who counts" in a mandatory welfare employment program is the need to remove the ability of staff to select those who "count"—and those whose performance for which they will be held accountable. It is a precondition for eliminating the problem of "creaming." To address this concern, it is important that the base for measuring performance standards be the entire caseload or an entire segment of the caseload—similar to the basis used for calculating participation rates—thus providing a strong mechanism to prevent creaming*. (HHS, 1994, p. 29)

Eliminating the ability of local welfare offices to select those whose success will determine their performance measure will not eliminate the selective provision of services, nor should it. The local welfare office will probably do some sort of "triage creaming." The available services will be most focused on those welfare recipients in a position to most benefit from them, whereas those who do not need the available services, or who have problems too far-reaching for these services to make a difference, will receive less attention.

5. *Data on "outcomes" for welfare recipients in a welfare-to-work program or training program can be regression adjusted so that the adjusted numbers are positively correlated with the value added from the program.*

Both Friedlander of MDRC and Zornitsky and his colleagues from Abt Associates have examined the strength of the positive correlation between "adjusted" outcome data from a welfare-to-work program, and the "value added" from the program. Table 5 summarizes their findings. This research suggests that individual outcomes, adjusted for the characteristics of the individual and the local economy, are often significantly correlated with the value added of the program for that individual.

It is important to adjust outcome data from welfare-to-work programs for client mix and local economic conditions for at least three reasons: (1) such adjustment is a further incentive to avoid creaming; (2) adjustment for local conditions increases the perceived fairness of standards for judging local office performance, contractor performance, and individual performance; (3) state managers need to have adjusted outcome measures that are at least reasonable proxies for value added if they are to reliably identify successful programs that might be imitated.

Friedlander's remarkable result in Table 5 is that even very crude adjustments to outcomes data can result in significant positive correlations of the adjusted outcomes data with value added. For example, if we simply give extra "points" for job entries for clients with low prior earnings, the resulting "weighted job entry" points for each individual has a significant positive correlation with the individual's earnings gain from the program in four out of five sites in Friedlander's study.

Table 5
Summary of Correlations Between Several Performance
Measures and Several Value-Added Measures
in Two Research Studies

PANEL A: Zornitsky results for correlation between various performance measures and value added measures (earnings gains in $, welfare reductions in $) for AFDC Homemaker-Home Health Aide Demonstration

Performance Measures	**Correlation with earnings gain:**		**Correlation with welfare reductions :**	
	Unadjusted measure	Adjusted measure	Unadjusted measure	Adjusted measure
Placed at termination	0.053 (0.036)	0.051 (0.053)	0.027 (0.295)	-0.006 (0.813)
Avg Placement Wage	0.086 (0.001)	0.083 (0.002)	0.041 (0.109)	-0.020 (0.438)
Employment Status				
3 months	0.068 (0.007)	0.079 (0.003)	0.007 (0.783)	-0.033 (0.202)
6 months	0.119 (0.000)	0.126 (0.000)	-0.003 (0.920)	-0.042 (0.108)
Avg Weekly Earnings				
3 months	0.113 (0.000)	0.120 (0.000)	0.023 (0.363)	-0.032 (0.222)
6 months	0.142 (0.000)	0.151 (0.000)	0.011 (0.661)	-0.046 (0.080)
Welfare Status				
3 months	-0.017 (0.506)	-0.005 (0.839)	-0.010 (0.687)	0.063 (0.017)
6 months	-0.024 (0.344)	-0.017 (0.517)	0.017 (0.511)	0.089 (0.001)
Welfare Benefits Level				
3 months	-0.046 (0.072)	-0.043 (0.098)	0.046 (0.073)	0.090 (0.001)
6 months	-0.048 (0.059)	-0.047 (0.074)	0.050 (0.048)	0.101 (0.000)

Notes: Levels of significance reported in parentheses. Adjustment is based upon regression that includes following variables: age, education, race, marital status, number of children, pre-program months of unemployment, pre-program average monthly AFDC benefits, site unemployment rate, site average wage. Reported correlation is between performance measure and estimated value added for individual. Welfare dependency reductions criterion is dollar reduction in welfare benefits, calculated so that reductions are negative numbers. Source is Zornitsky and Rubin, 1988, p. 76.

Table 5
(Continued)

PANEL B: Friedlander results based on analysis of welfare to work experiments at five sites.

Performance Measures	Number of sites with statistically significant correlation of "right" sign between measure and earnings gain		Number of sites with statistically significant correlation of "right" sign between measure and welfare reductions	
	Unadjusted measure	Adjusted measure	Unadjusted measure	Adjusted measure
Job entry 2 to 3 quarters after random assignment	3	4	0	2
Job entry 4 or more quarters after random assignment	3	4	0	2
Off welfare 3 quarters after random assignment	1	2	0	3
Off welfare 6 quarters after random assignment	1	2	1	3

Notes: Adjustment is done giving more "points" to job entries or welfare exits for persons with weaker prior earnings history. Calculated correlation is between number of points for individual and estimated value added for that individual for each of five sites examined in the study. Source: Friedlander, 1988, p. 102, p. 108.

The study by Zornitsky and his colleagues at Abt uses a more complete regression model to control for a variety of client background characteristics, including prior unemployment, preprogram welfare dependency, age, education, race, and marital status. Their regression model also controls for local economic characteristics such as the state unemployment rate. Using this adjustment, they find that a variety of adjusted performance measures for each individual are significantly positively correlated with the estimated earnings gains for that individual, based on experimental data. Among the good adjusted performance measures for earnings gains, in the sense of being significantly positively correlated, are the entered employment rate, the placement wage, and employment status and average weekly earnings at three months and six months. The estimated true reductions in average welfare payments to each individual, based on experimental evidence, are highly positively correlated with the regression-adjusted performance measures for welfare status and welfare benefits at three months and six months after termination from training.

One issue is whether such regression models are so complicated that they eliminate the desired simplicity of performance measures. But regression-adjusted performance measures do not have to be stated to staff or local welfare offices in a complicated manner. Before the program year begins, a state could calculate goals for that local welfare office for the upcoming year after

adjusting for predicted client mix and area economic conditions. Goals can be expressed as the actual number of job placements, expected average wages, etc. The line staff does not need to know the details of procedures used to calculate this goal. The preliminary goals could be relaxed at the end of the program year if actual local client mix and economic conditions turned out to be more difficult than expected.

6. *Performance measures work better the more directly related they are to goals*.

In both the Zornitsky and Friedlander studies, earnings- and employment-related performance measures are more strongly correlated with earning value added than are welfare-related performance measures. Also, welfare-related performance measures are more strongly correlated with the program's effect on reducing welfare than are earnings- and employment-related performance measures.

7. *Performance measures are probably easier to develop for earnings value added than for welfare savings value added*.

Based on Friedlander and Zornitsky, a good performance indicator for the effect of a program on welfare, compared to a good indicator for earnings effects, requires more adjustment to the indicator and this adjustment is more difficult. Unadjusted performance measures frequently have significant correlations of the "right" sign with earnings value added, whereas unadjusted indicators frequently have correlations with the program's welfare savings that are insignificant or the "wrong" sign. Even after adjustment, performance indicators have more problems in measuring welfare savings than in measuring earnings gains. In Zornitsky's study, the maximum correlation between a performance indicator and welfare savings is considerably less than the maximum correlation between a performance indicator and earnings gain. In Friedlander's study, fewer sites (three out of five) show a correctly-signed and significant correlation with welfare savings than with the earnings gain (four out of five), even after adjusting the performance indicators.

The problems with performance indicators for welfare savings occur because welfare savings from social programs are greater for individuals who would otherwise be welfare dependent. Even with a program that is successful, many of these women will still be heavily welfare dependent. A program that targeted such women for services might look bad even if it was quite successful. In contrast, the earnings gains from welfare-to-work programs are not so concentrated among individuals who otherwise would have low earnings.

8. *It is probably more important to adjust for client characteristics than to extend the length of the follow-up period for collecting data for performance indicators, although both help*.

As shown in Zornitsky's study, if performance indicators are unadjusted, the correlation between the performance indicator and value added may have the wrong sign. This would imply that we would be rating some offices or programs as inferior that were superior. Avoiding this

problem by adjusting the performance indicators for client mix and other local area characteristics is important.

If we extend the follow-up period over which performance data are collected, the correlations between the performance indicator and true value added does get stronger. But even shorter follow-up periods with the adjusted performance measures are sufficient to get indicators that have the right sign and are statistically significant. Extending the follow-up period helps but is not essential. Against the advantages of a longer follow-up period must be weighed the extra costs of collecting data for a longer follow-up period.

9. *In adjusting performance indicators for the unique characteristics of each welfare office, among the important variables to include are: (a) for the individuals served by the office prior earnings history, prior welfare history, race, marital status or gender of household head, education and age; and (b) various measures that will reflect both site job availability (unemployment rate or job growth) and site wages*.

These are the types of controls used both in JTPA (see Figures 1 and 2) and Zornitsky and Rubin (1988) (see notes to Table 5). The control for prior earnings history seems particularly important. Friedlander gets a good correlation between adjusted performance measures and true value added with just a simple adjustment for prior earnings. The regression used by Zornitsky shows that preprogram months of unemployment has one of the strongest negative associations with the individual's earnings six months after termination from the program, but a strong positive correlation with the earnings gains from the program (Zornitsky and Rubin, 1988, p. 71, p. 79). Prior welfare benefits also is an important control. Zornitsky finds that average monthly public benefit payments in the 12 months before entering the program are strongly positively associated with welfare benefits received six months after termination from the program, but strongly negatively associated with the welfare reductions from the program (i.e., women with high welfare benefits prior to entering the program tended to have their welfare payments reduced more due to the earnings and other effects of the program). Thus, prior earnings and welfare experiences are important controls, because they can potentially cause the outcomes of the program to be unrepresentative of the program's true performance.

10. *Randomization is often possible as a permanent device for improving program management*.

We don't have to adopt Riemer's proposal of contracting out the entire welfare system to get program management benefits from doing random assignment. If we contract out some portion of the welfare-to-work system to different service providers and use random assignment to determine which client is assigned to which service provider, we can reliably determine the relative performance of the different providers even if each provider only works with 100 welfare recipients. As mentioned before in this paper, Milwaukee is randomly assigning its entire JOBS caseload to two different teams of service providers. In Kalamazoo, a county of less than a quarter million people, the Upjohn Institute is using random assignment among three different providers for the initial job search and job development phase of Michigan's Workfirst program

for welfare recipients. The relative performance of these three different contractors will determine future assignments of welfare recipients to different providers, and some contractors may even be dropped from the program if their performance is significantly poorer than other contractors.

Random assignment cannot be done in all welfare offices for all welfare recipients. Some welfare offices in rural counties have such small caseloads that it would be uneconomical to have several competing service providers, each with their own overhead costs. Even within large urban areas, there may be some individuals located in subareas in which there are relatively few other welfare recipients, and in which it may not make sense to have more than one provider. But competition among different service providers may be more viable than some would assume. For example, a request for proposal could specify that several contractors would be selected to be randomly assigned part of the caseload and require each contractor to use vans to pick up their welfare recipient clients throughout a defined geographic area.

11. *How to best improve the long-term value added of welfare-to-work programs is an important unresolved issue.*

Welfare-to-work programs seem to differ widely in long-run effectiveness in ways that are hard to predict. We would like to develop performance indicators that could be used in the short run as indicators of long-run success and use these short-run indicators for program management purposes. Directly using long-run indicators of program success to reward or sanction individual offices or staff members is likely to be so delayed that it would be ineffective as a management tool.

One might suppose that the long-run success of a welfare-to-work program would be related to greater educational content, placement in better jobs, or targeting folks more likely to be long-run chronic welfare recipients or the permanent poor. But these hypotheses are not universally supported by program data.

Consider the issue of greater educational content in welfare-to-work programs. Friedlander and Burtless's recent book (1994) indicates that Baltimore's welfare-to-work program has greater long-run (equals five years in their book) success than four other welfare-to-work programs they consider, and Baltimore appears to have somewhat more educational content than these other welfare-to-work programs. As another example of the potential success from greater education content, the Minority Female Single Parent job training program of San Jose's Center for Employment Training achieves great long-run success and does so with a job skills-focused approach (Zambrowski and Gordon, 1993) . On the other hand, some welfare-to-work programs with greater educational content appear to have failed. Most of the other demonstration sites for the Minority Female Single Parent job training program showed no significant effects on earnings. CET was distinguished from these other programs by a greater focus on skills demanded in the local labor market, and by teaching basic skills as part of a course focused on locally demanded job skills, not as a separate course. The New Chance program succeeded in getting many teenage welfare mothers to successfully complete their GED, but the program appears to have had no significant effect on success in the labor market (U.S. DOL, 1995).

Finally, the National JTPA Study suggests that for women, JTPA classroom training does not have significant earnings effects, but on-the-job training (OJT) does have significant earnings effects (Bloom et al., 1994). Education and job training for disadvantaged women are more successful if the job skills taught are linked as closely as possible to specific jobs in demand in the local labor market. But it is not clear how this insight helps us develop a short-run performance measure. Courses at a local community college do not come equipped with labels showing their relevance to the local job market.

Consider the issue of the relationship of the quality of initial job placements for welfare recipients and long-run success. Being placed in better jobs should increase long-run success, because job retention should be greater at higher wage jobs, and higher wage jobs on average have more on-the-job training that will promote subsequent upward mobility. Unfortunately, there is little direct empirical evidence in favor of this plausible hypothesis. In examining the data, it is difficult to know whether persons who are placed in better jobs do better in later years because of that placement, or whether persons with better skills as a result get better jobs later and better jobs now. In other words, is the positive correlation between a person's job quality over time a true causal relationship or merely reflective of the reality that some persons have more valued skills? In the Baltimore welfare-to-work program, which showed greater long-run success than the other programs analyzed in the Friedlander and Burtless book, there was no sign that welfare recipients obtained higher paying jobs than they did in the other programs. There is a great need for more research here. We need to know what types of jobs are most conducive to long-run success. Perhaps there are characteristics of jobs other than wages that are really the most crucial in determining whether the job leads to long-run success.

Consider the issue of targeting "hard-core" welfare recipients. It might seem plausible that programs that do this will have the greatest long-run success. The evidence suggests, however, that most of the programs with the greatest success in increasing long-run earnings, such as CET or the Baltimore welfare-to-work program, have little effect on long-run welfare receipt. What these programs seem to do is increase the long-run earnings potential of persons who would eventually leave welfare anyway. We do not know how to reliably increase the long-run earnings of persons who would otherwise have been long-run welfare recipients. This is a policy area in which we need more research and experimentation.

12. *Costs for data collection for performance measures can be held down by relying on administrative data rather than special surveys*.

Surveys are very expensive compared to using administrative data. Much of the data needed to create performance measures can be generated from internal welfare department data, assuming the appropriate questions are asked at intake, and the records are computerized. For example, intake interviews can collect data on education, race, age, family composition, and prior earnings and welfare receipt, all of which are important controls to use in adjusting outcome measures to create good performance indicators. The matching of these individual records with other welfare department records and with earnings records from unemployment insurance (UI)

files (the "ES-202" data) will increase the accuracy of the prior welfare receipt and prior earnings data.

In addition, some performance measures will use longer-run follow-up data. Longer-run follow-up data allows the creation of performance measures that will be more strongly correlated with the true impact of the program. In addition, we want to try to unravel the puzzle of what policies and practices will contribute to long-run success, and this requires longer-run follow-up data.

These longer-run follow-up data can be collected through surveys, or through administrative earnings records from the unemployment insurance system. The available research shows that collecting earnings data from the UI system is far less expensive than doing surveys—probably less than one-fifth the cost (National Commission for Employment Policy, 1992). It is hard to track down welfare recipients once they leave the welfare system. There are also questions about the accuracy of reported earnings from surveys of welfare recipients.

The disadvantages of using UI system data are several. First, UI data does not include information on hours worked or hourly wage rates. Second, data from a given state will miss persons who have moved to another state. However, these persons will also be hard to track down in a survey, and a survey will miss many other individuals as well. Third, there will be some slight delay from using unemployment insurance data. ES-202 data will probably not be available for performance indicator purposes except with a one-quarter lag. In addition, ES-202 data only reports information for the entire quarter. Therefore, if one wants to know someone's earnings in the three months after that person obtains a job, one would need to wait for data from the quarter after job placement, which might not be available for up to six months after the initial job placement. A survey can collect data exactly 13 weeks after the job placement, and the survey data would be available to the program almost immediately. Therefore, any system of rewarding or sanctioning local offices or staff using performance indicators will be based more on lagged performance under a system using UI data than under a system using survey data. This disadvantage is outweighed by the lower cost and greater accuracy of UI data.

One advantage of surveys of welfare recipients is that they can be designed to collect unique information. For example, interviews with welfare recipients could ask direct questions of how welfare recipients rated the quality of the welfare-to-work services they received, whether and how they felt these services were helpful, and how they felt they were treated by program staff. Collecting such sensitive information is likely to require that the surveys be done by some group independent of the local welfare office to assure confidentiality of the survey respondents and reduce any possible manipulation of the results. Whether the value of such additional information outweighs the additional costs of such surveys is a tough issue.

POSSIBLE MODELS FOR STATE PERFORMANCE MONITORING

This concluding section summarizes the lessons learned from previous experience and the suggested guidelines by outlining a model for state government performance indicators for welfare-to-work programs. I consider state monitoring of local office performance, state analysis of what program approaches are most effective, local office monitoring of contractors, and local and contractor monitoring of staff.

State Monitoring of Local Office Performance in Welfare-to-Work Programs

State monitoring of local office performance in welfare-to-work programs should focus on only a few key performance standards. Keeping standards to a small number helps give a stronger direction to local offices, and avoid overwhelming local offices with too many conflicting goals. To ensure that standards are perceived as fair, the standards should be adjusted for local economic conditions and the local client mix, and in particular for the prior earnings and welfare history of clients. Performance standards should be based on the success of the entire welfare caseload, or on some segment of the welfare caseload that is objectively identifiable, so that local offices cannot select which welfare recipients will count in meeting the standards. Even though the standards may be calculated by a sophisticated model that projects client mix and local economic conditions, standards should be stated in a simple form to local offices and staff. For example, standards could be stated as the number of welfare recipients in the local office expected to meet particular goals (job placement, off welfare, etc.). Standards should be made clear in advance, based on predicted local office conditions, so that offices and staff know in advance what goals to meet that year. Standards may be later relaxed if local conditions significantly change, for example if the local economy deteriorates. Wherever possible, data needed for calculating the performance measures should be obtained from administrative sources such as welfare department records or unemployment insurance earnings files. Use of administrative data will hold down data collection costs and increase data accuracy.

Whether a local office meets performance standards should have some consequences or real incentives for the local office budget, to ensure that the standards are taken seriously. These consequences should be modest rather than deciding a large portion of the budget to avoid overstressing standards that inevitably will be imperfect. Technical assistance should be provided to local offices that fail to meet the performance standards.

State Analysis of What Program Approaches Have the Greatest Long-Term Effectiveness

We need a great deal of additional research to address an important unanswered question: what program approaches will have the greatest long-run effectiveness in helping welfare recipients to increase earnings and reduce welfare dependency. State welfare agencies need to be part of the effort to address this issue, because they possess much of the data needed. State welfare agencies have a strong incentive to address this issue, because long-term welfare dependence is costly and its solutions may vary from state to state.

To find out what contributes to long-run success for welfare recipients, the state welfare agency should use UI earnings records to follow welfare recipients for five years after receiving welfare-to-work services. Data files should be created that link information on an individual's long-run earnings history to information on services received from welfare and training agencies, and on initial job placements. The data should be analyzed to see what local offices, particular services, or particular job placements are linked to long-run success for particular types of clients. If a particular office, program or job placement seems to help, follow-up interviews should be conducted to confirm this statistical finding, and get more insight into why this intervention is successful.

Local Office Monitoring of Contractors

In cities of mid-sized population or above, local welfare offices can readily identify several contractors who can compete in providing specified welfare-to-work services. Welfare recipients can be randomly assigned among these contractors. The relative performance of the contractors may be used to determine the amount of payment to contractors, how many future welfare recipients will be assigned to a contractor, and whether the contract will be renewed.

Local and Contractor Monitoring of Staff

Local welfare offices and contractors may wish to assign individual case managers or job developers with responsibility for meeting part of the organization's overall performance goal. Goals are helpful in focusing staff attention even when they are easy to meet. Whether individual staff members meet their goals should be a part of their performance evaluation.

A FINAL WORD

With an effective performance standard system at the state, local office, and contractor level, the overall effectiveness of the welfare-to-work system should improve each year. Continued research by state welfare offices should identify why particular offices or programs are more successful, and help other offices or programs adopt these practices. With continuous improvement of the welfare-to-work system each year, a state should be able to significantly increase earnings and reduce welfare dependence over a ten-year period.

References

American Public Welfare Association, *JOBS: Measuring Client Success*, The Institute for Family Self-Sufficiency, November 1994.

Anderson, Kathryn H., Richard V. Burkhauser, and Jennie E. Raymond, "The Effect of Creaming on Placement Rates Under the Job Training Partnership Act," *Industrial and Labor Relations Review*, Vol. 46, No. 4 (July 1993).

Barnow, Burt S., "The Effects of Performance Standards on State and Local Programs," in Charles F. Manski and Irwin Garfinkel (eds.) *Evaluating Welfare and Training Programs*, Cambridge, MA: Harvard University Press, 1992.

Behn, Robert D., *Leadership Counts: Lessons for Public Managers from the Massachusetts Welfare, Training and Employment Program*. Cambridge, MA: Harvard University Press, 1991.

Bishop, John H. "Improving Job Matches in the U.S. Labor Market," *Brookings Papers: Microeconomics*, 1993.

Bloom, Howard S., Larry L. Orr, George Cave, Stephen H. Bell, Fred Doolittle, and Winston Lin, *The National JTPA Study Overview: Impacts, Benefits, and Costs of Title II-A*, Abt Associates, Inc., January 1994.

Cave, George, et. al., *JOBSTART: Final Report on a Program For High School Dropouts*, Manpower Demonstration Research Corporation, October, 1993.

Cragg, Michael Ian, *The Impact of Performance Incentives on Providing Job Training to the Poor: The Job Training Partnership Act (JTPA)*, Stanford University Doctoral Dissertation, June 1993.

Dickinson, Katherine P. and Richard W. West, *Evaluation of the Effects of JTPA Performance Standards on Clients, Services, and Costs*, National Commission for Employment Policy, 1988.

Fein, Ester B., "For Job-Finding Concern, a Troubled Past," *The New York Times*, March 1, 1994.

Friedlander, Daniel, *Subgroup Impacts and Performance Indicators for Selected Welfare Employment Programs*, Manpower Demonstration Research Corporation, August 1988.

Friedlander, Daniel and Gary Burtless, *Five Years After: The Long-Term Effects of Welfare-to-Work Programs*, New York: Russell Sage Foundation, 1994.

Goldsmith, Stephen, "End the Welfare Delivery Monopoly," *The Wall Street Journal*, August 23, 1994.

Kirschenman, Joleen and Kathryn M. Neckerman, "'We'd Love to Hire Them. But…': The Meaning of Race for Employers" in Christopher Jencks and Paul E. Peterson, eds. *The Urban Underclass*, Washington, DC: The Brookings Institution, 1991.

Nathan, Richard, *Turning Promise into Performance: The Management Challenge of Implementing Workfare*, New York: Columbia University Press, 1993.

National Commission for Employment Policy, *Using Unemployment Insurance Wage-Record Data for JTPA Performance Management*, NCEP Research Report No. 91-07, June 1992.

Nightingale, Demetra Smith, et. al., *Evaluation of the Massachusetts Employment and Training (ET) Program*, Washington, DC: The Urban Institute, 1991.

O'Neill, June, *Work and Welfare in Massachusetts: An Evaluation of the ET Program*, Boston, MA: Pioneer Institute for Public Policy Research, 1990.

Riccio, James, Daniel Friedlander and Stephen Freedman, *GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program*, Manpower Demonstration Research Corporation, September 1994.

Riemer, David R., "Replacing Welfare With Work: The Case for an Employment Maintenance Model," *Focus*, Vol. 16, No. 2 (Winter 1994-1995).

Rosenberg, Jan and Sol Stern, "America Works: A Venture to End Dependency," *City Journal*, Summer 1993.

U.S. Department of Health and Human Services, *Report to Congress: Recommendations on Performance Standards for the JOBS Program*, Administration for Children and Families, Office of the Assistant Secretary for Planning and Evaluation, and Manpower Demonstration Research Corporation, 1995.

U.S. Department of Labor, Office of the Chief Economist, *What's Working (and what's not): A Summary of Research on the Economic Impacts of Employment and Training Programs*, U.S. Department of Labor, January 1995.

U.S. General Accounting Office, *Welfare to Work: Measuring Outcomes for JOBS Participants*, Report to the Ranking Minority Member, Committee on Finance, U.S. Senate, April 1995.

U.S. General Accounting Office, *Welfare to Work: Current AFDC Program Not Sufficiently Focused on Employment*, December 1994.

Zambrowski, Amy and Anne Gordon, *Evaluation of the Minority Female Single Parent Demonstration: Fifth-Year Impacts at CET*, Mathematica Policy Research, Inc., December 1993.

Zornitsky, Jeffrey and Mary Rubin, *Establishing a Performance Management System for Targeted Welfare Programs*, National Commission for Employment Policy Research Report No. 88-14, August 1988.