

IMPROVING STATE TANF PERFORMANCE MEASURES

HEATHER HAHN
PAMELA LOPREST

Acknowledgements

This report was prepared as part of a study funded by the U.S. Department of Health and Human Services (DHHS), Office of the Assistant Secretary for Planning and Evaluation (ASPE) under Task Order Number HHSP23337013T. We would like to thank all the individuals who participated in interviews and discussions for this project, particularly state and local TANF staff who gave their time so that others could learn from their experience. We also appreciate the comments we received from our project officer Kristen Joyce, other ASPE and DHHS staff, our internal project advisors Swati Desai, Olivia Golden, and Sheila Zedlewski, and insights from audience members at the Welfare Research and Evaluation conference and National Association of Welfare Research and Statistics conference where these results were presented. We would like to acknowledge the excellent research assistance of David Kassabian and Alex Stanczyk on this project. All errors and omissions remain those of the authors.

Table of Contents

<u>Executive Summary</u>	i
<u>I. Introduction</u>	6
<u>Background and Context</u>	6
<u>Study Goals and Research Questions</u>	10
<u>Methodology</u>	11
<u>II. State Summaries</u>	14
<u>Maryland</u>	14
<u>Minnesota</u>	14
<u>New York City</u>	15
<u>Utah</u>	16
<u>Wisconsin</u>	17
<u>Additional Research States</u>	18
<u>III. Findings</u>	21
<u>Performance Measure Definitions and Goals</u>	22
<u>Outcome Measures</u>	22
<u>Other Measures</u>	26
<u>Other Issues</u>	27
<u>Performance Measure Implementation</u>	29
<u>History and Development of TANF Performance Measurement Systems</u>	29
<u>System of Using Performance Measures</u>	30
<u>Assessing the Effectiveness of Measures</u>	35
<u>Implementation Challenges</u>	36
<u>Role of Data</u>	37
<u>Data Collection</u>	37
<u>Data Analysis</u>	38
<u>IV. Lessons Learned</u>	40
<u>V. State Recommendations for the Federal TANF Performance Measurement System</u>	42
<u>Improvements to the Federal TANF Performance Measurement System</u>	42
<u>Implications of Hypothetical Changes</u>	42
<u>VI. Conclusions</u>	45
 <u>Appendix A. Expanded State Summaries</u>	 46
<u>Maryland</u>	46
<u>Minnesota</u>	49
<u>New York City</u>	53
<u>Utah</u>	56
<u>Wisconsin</u>	59

Executive Summary

Performance measurement is a tool government can use to improve program performance and address accountability. The Temporary Assistance for Needy Families (TANF) program, like many federal government programs, requires measurement of program performance to help ensure federal funds are being used to reach stated program goals. The primary performance measure used by the federal TANF program toward the goal of promoting work among recipients is the work participation rate (WPR). Some states have gone beyond the WPR and added additional performance measures for their state TANF programs, making them useful laboratories for understanding the possibilities and challenges of broader and varied performance measurement in TANF. This study exploits this opportunity by gathering and synthesizing information from a set of states with more innovative performance measurement systems. These states' experiences provide models and information for other states wishing to broaden their TANF performance measurement systems and inform discussion and review at the national level on federal TANF performance measures.

Our study methodology included visits to five sites that have innovative TANF performance measurement systems (Maryland, Minnesota, New York City, Utah, and Wisconsin), phone conversations with five additional states with specific experiences relevant to the study (Arkansas, California, Florida, Texas, and Washington), consultations with a number of knowledgeable experts, and a review of related materials. The site visits included discussions with state and local officials, data analysts, advocates, and researchers, and took place between March and May 2011. The phone conversations were conducted during the same time period.

Findings

Each study site had a different performance measurement system. We summarize here the different types of performance measures, the implementation of these measures (including setting targets, reviewing performance, and using incentives and penalties), and how states sought to assess the effectiveness of the measures. We then summarize the overarching lessons learned from the study. Details of each system are provided in the report and the site-specific appendices to the report.

Performance measures. Performance measures in the study sites include a variety of outcome measures and additional measures. Although the states measure similar outcomes and other program indicators, each state tends to have its own specific definitions for the measures. Employment outcomes were the most common, but they vary in type and definition. Employment outcomes include new placements, ongoing employment, wages, wage gains, and employment retention. Minnesota uses the unique Self-Support Index that measures adults three years after a base quarter who are working a minimum number of hours or who are off cash

assistance (including those receiving disability benefits but not those off due to a time limit or sanction). Nonwork outcomes are also measured by a number of states. These include educational attainment and case closures for positive reasons, such as for employment or increases in child support income. Most states also include other measures such as engagement in activities and timeliness or efficiency of processing cases. Several, but not all, of our sites included the federal WPR in their performance measurement system.

While the measures used were varied, none of the study states has implemented performance measures directly related to child outcomes, family stability, access (take-up rates), or poverty. Arkansas has an outcome goal of increasing the number of former recipients who move out of poverty, but has not been able to develop indicators to measure this outcome. In addition, we explored whether states had integrated measures for TANF and workforce programs, particularly the measures required for the Workforce Investment Act (WIA) programs. While a number of study sites had integrated service delivery, only Texas had integrated measures and targets for TANF and other WIA clients. Several study sites did note the influence of the WIA measures on their measures of employment outcomes.

Performance measure implementation. The study sites varied in how long they have had their performance measurement systems. Wisconsin and New York City had systems essentially since the institution of TANF. Utah started developing its new system in 2010. Some systems were developed in response to being dissatisfied with the federal WPR, others were developed to help meet the WPR requirements. All states felt performance measurement systems were aligned with and served to help them meet their TANF program goals.

In most locations, performance targets are set in advance of the period of performance, with some performance measures weighted more than others. In some cases, the specific measures or targets are adjusted for demographics or other local conditions. New York City sets a target range for each indicator and measures performance against this range. Minnesota uses a regression model to calculate an expected range of values for a county's success rate on the Self-Support Index, making adjustments for variation across counties in economic and demographic characteristics. Unlike the other states, Minnesota evaluates county performance based on outcomes three years after the period of performance.

States take different approaches to how they review performance using the measures and targets set. In most of the states, performance reviews occur at least every few months, although in some cases monthly reports or quarterly reviews are essentially status checks, with counties or offices held accountable for performance over a longer period of time. Some places, such as New York City and Wisconsin, have formalized meetings for review and report-card type reports that serve as the basis for reviewing performance and developing strategies for improvement. In both of these sites, comparing or publishing performance across areas serves as an additional

incentive to perform. Utah's system ties into the state's existing computerized "dashboard" that tracks performance for each caseworker and allows managers at each level to review performance.

Finally, the states we visited use different incentives and penalties in conjunction with their performance measures to move toward their goals. To meet performance goals, some states such as Minnesota (and previously, Maryland) link funding or monetary incentives to performance; some states use competition among offices or vendors¹ (including New York City and Wisconsin); some, such as Utah, use neither. It is important to note that many of the discussion participants emphasized that staff on the state and local levels respond to measures and that "what gets measured gets done"—expressing the feeling that the mere measurement and review of the measures caused changes in behavior.

Performance measure assessment. There are a number of ways states seek to assess the effectiveness of their performance measurement systems. None of the states we visited had formally evaluated whether their measurement systems improved performance. Nonetheless, during our site visits, states shared their observations of how their performance measurement systems had contributed to changes in their TANF programs, improved outcomes for families, and generated further refinements of the performance measurement systems. Overall, discussion participants emphasized that the performance measures focus attention and action on whatever is measured. Each of the site visit states reported that their performance measurement systems had allowed them to more rapidly identify problems and provide technical assistance or work toward other solutions. Several of the site visit states also reported that their performance measurement systems had led local offices to focus more intently on engaging with clients to meet client needs. States generally assess the effectiveness of the performance measurement systems by assessing how well the states meet their program goals. As evidence of the success of the focus provided by the performance measures, officials in Utah, Wisconsin, and New York City each pointed to the improved employment outcomes for their clients. Maryland pointed to their work participation rate increasing substantially in response to implementing performance measures, among other things (performance-based funding and universal engagement requirements). Minnesota's performance measurement system led to a very specific change in their program. The state began targeting specialized services to African American and Native American TANF populations after analyses of the demographic control variables for the Self-Support Index highlighted persistent racial and ethnic disparities.

¹ Throughout this report, we generally use the term *vendor* to refer to a contracted employment services or case management services provider.

Lessons Learned

Through this study, we have identified several factors that study sites believed contributed to the success of their performance measurement systems.

Few and clear performance measures. Officials in New York City and Wisconsin emphasized the importance of limiting the number of measures and keeping them simple. Having a limited number of simple measures helps focus the attention of caseworkers and administrators, and makes it easier to identify high and low performance. People we spoke with in Minnesota made these same points by remarking that the complexity of their measures presents a challenge. The two measures for which Minnesota counties are held accountable—the work participation rate and the range of expected performance on the Self-Support Index—have a statistically complex methodology, making them difficult to understand and limiting their usefulness for day-to-day management, despite the basic concept being fairly straightforward.

Active management of performance. New York City and Wisconsin each highlighted the importance of having active management and steady attention for the performance measures. Both New York City and Wisconsin have regular meetings with top-level agency staff to review the JobStat, VendorStat, or KidStat reports. Not only does this keep attention focused on the measures, but New York City officials noted it also fosters working relationships and collaborative efforts to improve performance. Similarly, officials in Wisconsin cited the KidStat meetings for helping the state better identify struggling agencies in need of technical assistance, as well as identifying agencies in a position to share best practices.

Commitment to using data. The commitment and capacity for using data to inform policy were key factors in the success of the performance measurement systems in each of our site visit states. An upfront investment in data analysis systems gives states the ability to track, analyze, and interpret performance data. More important than the sophistication of the computer systems is the sophistication of the analytical staff. Minnesota representatives noted that their commitment to collecting and analyzing data allowed them to use data to inform policy despite the challenge of having a legacy data system. In addition, Utah has drawn heavily from longitudinal research of their TANF population and program to inform the development of their performance measurement system. Maryland's work performance specialists (state staff detailed to local offices to help with data and technical assistance) were viewed at the state and local levels as important keys to improving the work participation rate.

Priority of top managers. While New York City and Wisconsin noted the value of including top managers in their review meetings, several Utah discussion participants identified the leadership and support of their agency's executive director as a key to the success of their revised performance measurement system. Her clear and vocal support of the revised system has allowed staff at all levels to have the confidence to risk changing their ways of doing business.

Commitment to evolution. The importance of having a system that can evolve was discussed explicitly in New York City, Minnesota, and Wisconsin. Goals, measures, and relative weights of the measures are revisited and revised with each subsequent contract in Wisconsin (every two to three years) and annually in New York City. This evolution allows for better alignment with broader program goals as well as more responsiveness to program and data changes.

In conclusion, this study shows that a number of states have implemented performance measurement systems in state TANF programs well beyond the federal WPR measure. The states we visited unanimously believe their measurement systems are helping or have helped them actively manage performance to better reach their goals. The lessons learned from these states provide important information for other states considering expanding performance measurement in their TANF programs and for the federal government.

I. Introduction

Performance measurement is a tool government can use to improve program performance and address accountability. The Temporary Assistance for Needy Families (TANF) program, like many federal government programs, requires measurement of program performance to help ensure federal funds are being used to reach stated program goals. The primary performance measure used by the federal TANF program toward the goal of promoting work among recipients is the work participation rate (WPR). The WPR is the percentage of families with work-eligible individuals in which a family member participates in specified work activities including work, job search, and associated activities for a specific number of hours. Some states have gone beyond the WPR and added performance measures for their state TANF programs, making them useful laboratories for understanding the possibilities and challenges of broader and varied performance measurement in TANF. This study exploits this opportunity by gathering and synthesizing information for a set of states with more innovative performance measurement systems. These states' experiences provide models and information for other states wishing to broaden their TANF performance measurement systems, as well as inform discussion and review at the national level on federal TANF performance measures.

The term *performance measurement* can have several different meanings. It can refer to using data indicators to manage program performance down to individual caseworkers' performance, including accountability measures in performance-based contracts, using outcome indicators to evaluate program performance, or tracking and monitoring program data for goal-setting and policy. The states reviewed in this study use a combination of these types of performance measurement in their performance measurement systems. We use the term *performance measurement system* to refer to a state's overall approach, including the goals of performance measurement and the specific measures used, as well as the way the measures are used, including setting performance targets, reviewing performance, and creating performance incentives or penalties.

Background and Context

Researchers have found that using performance measures in public programs can improve outcomes in a wide range of public settings and for a broad array of city and state services. These measures often motivate, even in the absence of specific consequences, particularly when they clearly signal what leaders and the public think is important. They also

can be an important source of learning, highlighting differences in performance over time or among local offices or units. These differences can be mined for lessons about success or failure and used to pinpoint areas for improvement.² Program performance measures can focus on process or outcomes. Process measures address administrative or operational activities of a program—the means to getting to an end result rather than the goal itself (e.g., the percentage of applications processed promptly). Outcome measures focus on program goals such as job placement rates, employment retention, or wage rates in welfare-to-work programs. Of course, there is a continuum between the two types of measures since interim outcome measures such as achievement of a skill certificate indicate progress toward the ultimate outcome of employment.

Some of the main challenges to implementing performance measurement systems include defining measures that are linked to program goals, choosing levels of expected performance, putting the measures into practice, and developing consequences for not reaching targets.

To be effective, measures must be connected to program goals. However, defining measures that match specific goals can be challenging. Another challenge in implementing performance measures is the ability to collect timely and accurate data at the geographic level needed—potentially, at the worker, supervisory team, local office, or county level for a state measurement process, or at the state level for federal use. Typically, the solution is to start with what is available and then enhance over time, but information systems investments may be crucial to eventual success. A further challenge is distinguishing the true effect of program performance from the effects of outside influences, such as changes in economic conditions, demographics, and policy changes.

Implementing measures in an intergovernmental system raises additional challenges, including concerns about fairness and about “gaming the system.”³ States operate programs in different environments with different human capital and economic resources. Some start out way ahead of others, so it is important to define equitable performance measures. In addition, the specific definition of measures or systems of measurement (for example, rankings) may create unintended incentives or be more open to “gaming”—leading entities to take actions to meet a target performance measure without achieving the underlying goal. A related issue that can occur in programs aimed at vulnerable groups is an incentive to “cream,” that is, to direct resources to those most likely to succeed rather than those most in need. Promising approaches for corrective action could include directly targeting performance for various subgroups and frequently reviewing measures and their consequences in order to fix any incorrect incentives that emerge.

² Olivia Golden, *Reforming Child Welfare* (Washington, DC: Urban Institute Press, 2009).

³ Shelley Metzenbaum, with Allison Watkins and Adenike Adeyeye. “A Memo on Measurement for Environmental Managers: Recommendation and Reference Manual,” 1st ed. (College Park: Environmental Compliance Consortium, University of Maryland School of Public Policy, 2007).

TANF primarily holds states accountable for program performance through its WPR requirements.⁴ States must meet a 50 percent WPR for all families and a 90 percent rate for two-parent families or risk financial penalties.⁵ These rates apply to all work-eligible individuals on TANF. Federal statute and regulation defines the activities that can be counted as work participation, the number of hours that can be counted in each activity, and how states must calculate the WPR. The current rules were established through the 2005 Deficit Reduction Act (DRA) that reauthorized the original TANF program established in 1996. The WPR is primarily a process measure, focusing on engagement in work-related activities as well as work, rather than a pure outcome measure of recipients' success in the labor market.

Some have argued that TANF should shift toward outcome-based accountability with a focus on reducing poverty and hardship, improving child well-being, and helping parents work and gain skills for better jobs.⁶ Given the multiple goals of TANF beyond work (including assisting needy families so that children may be cared for in their own homes or in the homes of relatives; ending dependence of needy parents on government benefits by promoting job preparation, work, and marriage; preventing and reducing the incidence of out-of-wedlock pregnancies; and encouraging the formation and maintenance of two-parent families⁷) additional measures could be considered. Others have argued that TANF's program outcomes should be integrated with other workforce programs, such as the "common measures" used in the Workforce Investment Act (WIA) and other U.S. Department of Labor programs.⁸

In addition, some have criticized the WPR measures as requiring states to put a large amount of time into tracking and meeting specific activities requirements that do not always reflect the goal of improving employment outcomes.⁹

⁴ TANF also includes penalties for failure to maintain required state spending levels, failure to comply with the five-year time limit on use of federal funds, and failure to satisfy reporting requirements, among other penalties. For a complete list, see section 409 of the Social Security Act.

⁵ The TANF statute and regulations (particularly 45 CFR 261.2(n)) contain detailed rules concerning how the numerator and denominator are calculated for participation rate purposes.

⁶ Elizabeth Lower-Basch, Center for Law and Social Policy, Testimony to the Subcommittee on Income Security and Family Support, Hearing on TANF's Role in Providing Assistance to Struggling Families, March 11, 2010.

⁷ Paraphrased from Sec. 401. Purpose of the Block Grants to States for Temporary Assistance for Needy Families, Public Law 104-193, August 22, 1996, 104th Congress.

⁸ Mark Greenberg, Emil Parker, and Abbey Frank, "Integrating TANF and WIA into a Single Workforce System: An Analysis of Legal Issues" (Washington, DC: Center for Law and Social Policy, 2004).

⁹ Sheila Zedlewski and Olivia Golden, "Next Steps for TANF Reauthorization," Low-Income Working Families Policy Brief (Washington, DC: The Urban Institute, 2010).

The WPR performance measures have led some states to make policy changes that increase their WPR but do not actually increase work activity among individuals. For example, some states use small, monthly “earnings supplements” for TANF recipients whose earnings are high enough that they otherwise would be ineligible for TANF. Having these working recipients on the caseload helps the state meet its WPR. As another example, at least 27 states have implemented or authorized solely state-funded programs in response to the DRA requirements.¹⁰ States identify groups not likely to help their work participation rate, such as two-parent families or clients completing their educations, and serve them through solely state-funded programs, which are not counted toward the maintenance-of-effort requirement and not subject to any of the federal TANF requirements. The use of solely state-funded programs may be for gaming, for providing appropriate services to individuals when their activities would not be countable toward federal participation rates, or both.

Numerous other program outcomes also could be used to track program success and state accountability. A number of potential alternative outcome measures and related challenges to implementation were reported by the U.S. Department of Health and Human Services (HHS) in 2000.¹¹ The “High Performance Bonus” in the original TANF program, for example, included a broader measure of employment success and the reduction in out-of-wedlock pregnancies. However, the High Performance Bonus system was largely viewed as ineffective in identifying high performance, measuring real program improvement, or affecting state behavior, and Congress repealed funding for the High Performance Bonus structure in the DRA. Since the repeal of the High Performance Bonus, the Administration for Children and Families (ACF) has continued to calculate state job entry, job retention, and earnings gains rates based on matching monthly listings of adult TANF recipients against the quarterly wage files on the National Directory of New Hires. ACF uses this data source to measure and report employment among TANF recipients, though these rates are affected by economic and demographic factors and state eligibility rules as well as state performance.¹² The federal government also tracks a family formation and stability measure (the share of all children under 18 residing in married-family couple groups) and the percentage of families with children and earnings that participate in the Supplemental Nutrition Assistance Program.

¹⁰ Liz Schott and Sharon Parrott, “Designing Solely State-Funded Programs” (Washington, DC: Center on Budget and Policy Priorities, 2009).

¹¹ U.S. Department of Health and Human Services, “Report on Alternative Outcome Measures: TANF Block Grant,” 2000. <http://aspe.hhs.gov/hsp/alt-outcomes00/index.htm>.

¹² U.S. DHHS, *Temporary Assistance for Needy Families (TANF) Eighth Annual Report to Congress*, chapter V and appendix table 5:17 (Washington, DC: Administration for Children and Families, 2000). Measures use the National Directory of New Hires.

Despite the challenges facing the federal WPR measure, there is little systematically known about what states have done to go beyond the WPR to measure performance in their TANF programs. This study documents the state TANF performance measurement systems of several states that have gone beyond the federal measure, the choices they have made, and how they have addressed the challenges described above.

Study Goals and Research Questions

The study focused on three areas of research: performance measure definitions and goals, performance measure implementation, and effective use of performance measures. Specific research questions addressed in each area include:

Performance Measure Definitions and Goals

- How do states' performance measures relate to their TANF goals? Do the performance goals of counties, contractors and internal agency staff differ? Do states consider multiple measures toward reaching one goal?
- What process did states use to develop measures? Were measures vetted with constituent groups? Agency staff? What roadblocks did they face and what are the emerging promising practices?
- What measures are states using and how are the measures defined? To what extent are states measuring outcomes versus process? What measures are states using to expand on employment outcomes (beyond the work participation rate)? Are any states using population measures (i.e., measuring outcomes for a larger population than just TANF recipients)? Have states articulated goals and corresponding performance measures for nonwork aspects of their programs such as child-only cases and diversion expenses? Are states measuring child outcomes or family stability?
- How and to what extent have states integrated TANF measures with other social service or workforce program performance measures, in particular, those required by the WIA?
- What data are used to implement measures? What are the limitations of available data? How are the data collected, what is the impact of data lags, and how are the data validated? What is the relationship between sophistication of data systems and use or success of performance measures?

Performance Measure Implementation

- How long have measures been in place and have they stood the test of time? Has there been controversy about performance measurement issues among political stakeholders or operationally, within the agency?
- Was specific staff training required to implement measures?
- Who is responsible for implementing the measures—local offices, counties, states? How do these levels interact?

- How have states dealt with or tried to reduce strategic or gaming behavior (e.g., only serving those who can more easily meet set targets)?
- How are performance measures used in contracts with vendors? What are the limitations and successes of using contractors to reach performance goals?
- Is there a regular process for review/revision of performance measures and required targets? Have the performance measures changed over time and have they been influenced by changing government circumstances such as the recent economic downturn?

Effective Use of Performance Measures

- How do states set targets for performance measures? Are targets adjusted across counties or local offices?
- What are the consequences of meeting or failing to meet performance targets (e.g., use of financial incentives or penalties)? Is staff held accountable, or paid based on performance? Is publicity of relative scores (across counties, offices, or individual staff) part of the consequence?
- How do states determine/evaluate the effectiveness of performance measures? What changes and modifications have states made to their programs in response to the results of their performance measurement efforts?
- How and to what extent have states seen improvement in outcomes for families receiving TANF as a result of measuring performance?

Methodology

The main source of information for this study is from five site visits to Maryland, Minnesota, New York City, Utah, and Wisconsin. This information was supplemented with telephone calls to an additional five states (Arkansas, California, Florida, Texas, and Washington) and review of materials.

The sites for visits were chosen through a multistage selection process. First, an extensive canvass of existing materials and consultations with 26 relevant experts was conducted to identify promising state candidates and important aspects of state TANF performance measurement systems. This process identified 10 possible state TANF programs. Using the information gathered and guided by the study research questions, we developed the following set of dimensions of state TANF performance systems and arrayed the state candidates along them:

- performance measures beyond the federal work participation rate: work-related measures, child outcome measures, process measures, or other measures;
- clearly defined process of developing measures to align with goals for TANF program;
- process-based performance management systems, including dashboards;
- use of performance-based contracts beyond the federal work participation rate;

- highly integrated workforce/TANF systems that use WIA-type measures for TANF populations (or have the potential for this);
- performance systems with set targets and consequences for failing to reach or reaching targets;
- performance systems that adjust targets across entities for caseload characteristics, economic, or other factors; and
- extensive data/evaluation systems in use for TANF program.

In conjunction with HHS staff, we chose five sites that had the most extensive performance measurement systems and many of these dimensions.¹³ All the sites chosen were states with the exception of New York City, which has an extensive performance measurement system that is separate from and in addition to that of the state.¹⁴ Our site visits involved discussions with state human services agency/TANF program manager officials, data systems managers, local TANF program directors and contractors familiar with the implementation of performance systems, and interested policy advocates or experts outside of the government. In all, we held five to six meetings in each site, typically with multiple respondents in each meeting. A general discussion guide was developed for all sites and modified for specific respondents and to reflect information we had collected prior to the site visit. Site visits took place between March and May of 2011.

In addition to these site visits, we conducted hour-long telephone conversations with the remaining five states identified in our initial canvassing effort. A number of these states stood out on one specific dimension of the list above, so the phone call focused on gathering information related to that dimension. In California we focused on their statewide system of measures with incentive payments, in Texas and Florida we focused on the integration of measures with WIA programs, in Washington we focused on the measures they have developed in addition to work, and in Arkansas we focused on their measures of moving families out of poverty. A fuller description of these calls is included in the next section of this report on state summaries. We should note that all the topics discussed with site visit states could not be covered in this brief call. Throughout this report we include these states as examples where relevant.

Finally, at the time we were conducting this study, the American Public Human Services Association (APHSA) conducted a survey of its members in state human services agencies, asking about their use of TANF performance measures. APHSA shared the results with us and

¹³ Initially, Washington State was also chosen for a site visit but was unable to participate given limitations on staff time.

¹⁴ Throughout the report we use the terms *sites* and *states* interchangeably to refer to all five of the site visits, including New York City.

we used these as context for our report and validation of our site selection. Of the 14 responses to their survey, none identified additional states that stood out as having innovative TANF performance measurement systems on par with those already included our study.

II. State Summaries

Maryland

Maryland's TANF program (Temporary Cash Assistance, or TCA) is supervised and administered by the state through county-level social services departments that may use vendor contracting. Maryland includes outcome- and process-based measures through a series of reports originally developed when the state shifted to a universal engagement model for its caseload in 2003 in anticipation of the Deficit Reduction Act. To support this model, the state adopted a performance management process based on New York City's JobStat model, with monthly, county-level reports that score performance on a range of indicators and give a combined score. State officials regularly met with the governor's staff and local TCA managers as part of the JobStat system. The local allocation of work program funds was 25 percent performance based.

The formal JobStat system has been suspended in Maryland. Regular face-to-face meetings have been discontinued since 2008, and the state has not issued a JobStat template since May 2009. Use of performance-based funding has also been discontinued. Several reasons were mentioned for suspending the JobStat performance system including changes in state-level priorities, reductions in resources, and the fact that the WPR had increased substantially and seemed stable.

Currently, the measures originally reported in the JobStat templates still are tracked. Family Investment Program Indicators reports are made available to local managers with many of the same indicators but not in a JobStat report-card type format with weighted/combined scores, and without the regular meeting structure. As part of the original performance-based system, the state put in place work participation specialists, state staff that were detailed to local offices to help with the performance measurement system, including data issues, verification, and technical assistance to improve performance. These specialists remain in the local offices (some smaller local offices share one roving specialist) and continue to support local performance efforts.

Minnesota

Minnesota's TANF program (Minnesota Family Investment Program, or MFIP) is supervised by the state and administered by the counties and tribes.¹⁵ Since 2003, Minnesota has tied a portion of county TANF funding to two performance measures. One is the county's

¹⁵ For simplicity, we use the term *county* to refer not only to counties but also to the tribes that administer the TANF program in Minnesota.

performance on the Self-Support Index, a measure based on the percentage of adults three years from a base quarter who meet a positive outcome, defined as either (1) working 30 hours or more per week (and still receiving cash assistance) or (2) no longer receiving cash assistance (including those receiving Supplemental Security Income (SSI)). The county's actual performance is compared to expected performance on the Self-Support Index. The other performance measure is the county's performance on a measure paralleling the federal work participation rate. Counties automatically receive 95 percent of their TANF grant allocation from the state but only receive the remainder of the funds if they meet the expected performance on the two measures (2.5 percent per measure), or if they submit corrective action plans for addressing performance shortcomings.

The state conducts complicated data analyses to determine expected and actual county performance on the Self-Support Index and the work participation rate. The state also conducts extensive analyses of disparities by race, ethnicity, and other demographic variables. This measure is computed for all federally and state-funded MFIP participants, unlike the TANF work participation rate that only includes certain federally funded participants.

In addition to the two measures tied to county funding, quarterly reports to counties also include several management indicators not tied to funding, such as the number of and reasons for case terminations, median placement starting wage, the distribution of participants' countable months remaining on TANF, the number of participants returning to assistance, the percentage of recipients remaining off TANF after 12 months, and the number of "unaccounted-for cases" (neither exempt nor engaged in activities).

New York City

New York City's Human Resources Administration (HRA) implemented performance measurement systems for TANF shortly after welfare reform. We focus in this report on NYC's innovative JobStat and VendorStat systems. JobStat is a program that provides local offices (Job Centers) with performance data through the JobStat report and facilitates regular meetings (each Center is reviewed roughly every quarter) with Job Center leaders to review performance and develop problem-solving strategies. VendorStat is a program for vendors that provides performance data through a VendorStat report and holds regular meetings (every week with each vendor's performance reviewed every seventh week) to review performance and address issues. These vendor reports are used in conjunction with performance-based contracts and focus on cases in the Back to Work program (NYC's employment program). To get a full picture of the engagement of recipients, additional data tracking reports are generated including an engagement report, an employment report, and reports for each relevant HRA program area. TANF cases that are permanently or temporarily not employable are served through other NYC TANF programs,

such as WeCARE (Wellness, Comprehensive Assessment, Rehabilitation and Employment), for which separate data reports are kept.¹⁶

Data for performance reports comes from operational databases (information coming from front-line staff using the systems) which are uploaded weekly into a data warehouse. HRA's Office of Data Reporting and Analysis uses data from this warehouse (and other management information systems) to generate JobStat and VendorStat reports. HRA has a data intensive culture that is made possible by its investment in operational databases and the data warehouse as well as in human resources, including computer programmers and data analysts.

Initially, HRA tied performance on the JobStat report to management bonuses, but has since switched to using gift certificates and noncash awards to reward high performance. Publicly published reports and competition among Job Centers also have important roles in motivating high performance. HRA payments to vendors are tied to performance on several key indicators on the VendorStat report.

The January 2011 JobStat report (version 11.0) contains 29 measures including job placements, participation rate, employed cases, measures of engagement (employment plan, child care application, timely conciliation process), and application process/timeliness (for cash assistance and food stamps). The 2011 VendorStat report contains 19 measures related to client assessment, compliance, placement, and retention in employment. Some specific measures include the paid and unpaid employment plan rate, the failure to comply rate, the 30-day placement rate, the mean and median wage of job placement, and the 90- and 180-day retention rate.

HRA reviews and can adjust the measures on the JobStat and VendorStat reports annually.

Utah

Utah's TANF program (Family Employment Program, or FEP) is administered by the state through regional service areas. Utah began redesigning its TANF program in April 2010 with a Work First model that included an overhaul of its TANF performance measurement system. Before the redesign, performance measures for caseworkers focused on the federal work participation rate measure. While the countable activities of the federal work participation rate are still included among Utah's performance measures, the state has broadened its focus to

¹⁶ In addition to performance-based employment contracts, HRA also has contracts for clients with physical and mental health barriers to employment, in the WeCARE program. In addition to job placements and retention, the WeCARE program also tracks and measures SSI/SSDI receipt.

include additional performance measures that officials say better help achieve the goal of moving customers into employment. The new performance measurement system, still in transition, focuses on TANF customers' outcomes, including not only employment and earnings but also educational attainment and case closures for positive reasons. The new performance system ties in to Utah's existing computerized dashboard that tracks performance for each caseworker. Full implementation of the new system will depend in part on how quickly the underlying computer programming can be completed.

When the new system is completed, performance of individual caseworkers as well as aggregate administrative units will be tracked on the following outcomes: increased earnings (from entering employment, job retention, or increased wages), increased income (such as SSI or child support), high school/GED completion, successful completion of job club, and "positive" enrollment closures. The following process measures also will be tracked: compliance (i.e., are caseworkers following procedures?), quality (i.e., are caseworkers taking appropriate actions to serve their customers?), 50 percent customer participation in verified activities (i.e., meeting the federal work participation rate), and customer hours in countable services.

Measures, targets, and consequences apply at each level of Utah's TANF administration, such as economic service-area (regional) managers, local office managers, and individual caseworkers. Reviews of performance will identify areas for improvement or additional training.

The TANF dashboard system uses the same computer system as WIA, although there is no explicit coordination between the TANF and WIA measures and the dashboard is designed specifically for TANF, not WIA.

Utah's goals for TANF are to help customers enter unsubsidized employment, retain jobs, and increase wages; the state's new performance measures were designed to promote these goals. Progress, measured in improvement over time, is what matters most to state officials, according to administrators. Additionally, while the federal work participation rate still is tracked and measured, state administrators said staff have been instructed and trained to work toward specific positive outcomes for customers while not worrying about potential negative effects to the work participation rate.

Wisconsin

Wisconsin's TANF program (Wisconsin Works, or W-2) is administered by counties and private agencies through contracts with the state. Wisconsin includes outcome-based performance measures in its contracts with counties and other TANF providers. Performance standards include the following primary outcomes: job entry, earnings stabilization, customer satisfaction surveys, educational attainment, success of job skills training (obtaining employment after completing job skills training), and SSI/Social Security Disability Insurance (SSDI) receipt.

The performance standards deliberately do not include recipients' participation in federally countable work activities, as used to calculate the federal work participation rate.

Until recently, provider agencies either passed or failed each measure and those that were successful had the right to renew their contracts without competition. A new system currently being implemented scores agencies' performance on each measure as Exceeds, Satisfactory, Needs Improvement, or Fail, with points assigned to each level of performance. Agencies will be able to use their accumulated points toward renewing their contracts as well as to compete for new contracts elsewhere in the state.

In addition, Wisconsin has set aside a \$1 million incentive earnings pool. Any agencies that meet or exceed their performance standards will share the pool, according to a formula that accounts for agency size and other metrics. The agencies can use the funds to reinvest or innovate in any ways they choose.

Although the TANF performance measures in Wisconsin do not include measures of child well-being, child well-being is evaluated through the state's quality service reviews (QSRs). Based on the child welfare model, the TANF QSRs involve intensive case reviews of a small number of cases per agency and look at child outcomes as well as many other measures. The state does not plan to use the QSR for contract decisions, in an effort to keep the process collaborative and nonadversarial.

Additional Research States

Arkansas

Arkansas is an example of a state with a focus on outcome measures, not just process measures. In 2001, the state legislature passed a requirement for tracking 13 outcome measures related to the Transitional Employment Assistance (TEA) program, Arkansas' TANF program, moving the program away from an initial focus on caseload reduction. Subsequently, the list was reduced to five simpler outcome measures including moving recipients into employment, retention of employment, increasing earnings over time, and moving recipients over the poverty line. Given it is one of the few states to have a measure related to poverty, we wanted to explore this. It turns out that the outcome has never been implemented because the state was unable to develop related indicators; staff said, developing a corresponding set of measures proved conceptually difficult. A current effort is underway to contract with an outside group to develop possible indicators for this outcome as well as update all of Arkansas's TEA indicators.

California

California has in statute outcome-based performance measures that would give high-performing counties financial rewards. These measures, known as Pay for Performance, have

never been funded or implemented, although data were published for a few years on the state's web site. The three Pay for Performance measures are (1) the rate of TANF cases with earnings; (2) the federal WPR, excluding exempt cases and including sanctioned cases other than those with mental health, substance abuse, or domestic violence; and (3) cases with earned income three months after leaving assistance.

Florida

Florida's TANF performance measures mirror those in WIA. Two of the measures on the state's WIA monthly report (i.e., the short-term report) are outcome measures for TANF, an entered employment rate, and a wage earnings rate. The long-term report has three measures for the TANF program, separately, that emulate the WIA long-term outcomes or common measures: (1) an entered employment rate, (2) an earnings rate, and (3) a retention rate (individuals that have at least three quarters of wage data). These measures mirror the state's measures for WIA, although they differ somewhat from the federal-level measures. Unlike with WIA, for which the state sets targets and negotiates performance standards, Florida does not set targets for TANF performance. However, the workforce boards can earn additional TANF incentive dollars based on performance in comparison to other boards. In addition, TANF performance information for Florida's regions and individual providers is publicly available on the state's web site.

Texas

Texas has integrated its TANF and WIA performance measurement systems. In Texas, a single performance measurement system focuses on outcomes for all customers at its one-stop centers, regardless of which program funds services for each customer. The system does account for the fact that some customers have more difficulty obtaining and retaining employment by having, for example, both a systemwide entered employment rate and a subset of that called At-Risk Entered Employment. All TANF recipients are included in this subcategory, as are ex-offenders and others.

Washington

Washington State has a fully developed, data-driven government wide system for measuring and improving state agencies' performance, called Government Management Accountability and Performance (GMAP).¹⁷ At the WorkFirst program level (Washington's TANF program for adults), the six partner agencies that administer the program develop objectives and goals, create measures connected to those goals, and set targets for measures. Data are used to report on each measure at the state and local levels. The measures developed are

¹⁷ A full description of the state GMAP process can be found at <http://www.accountability.wa.gov>.

extensive and go beyond the WPR. In 2010, the governor directed the WorkFirst system to conduct a total reexamination of the programs, which could potentially impact elements of performance measurement going forward.

III. Findings

This section describes the major findings from our study, drawing mainly from our five site visit states, supplemented with examples from the five additional research states and other supporting materials. As an overview of the dimensions of state TANF performance measurement systems, Table 1 summarizes the types of performance measures used and the consequences associated with the measures. This section also discusses these dimensions as well as other aspects of the state systems, including details of their implementation and use of data.

Table 1. Dimensions of State TANF Performance Measurement Systems

Dimensions	MD	MN	NYC	UT	WI	AR	CA	FL	TX	WA
Types of Performance Measures										
Employment outcomes ^a	X	X	X	X	X	X	^b	X	X	X
Education outcomes				X	X	X			X	
Positive case closures			X	X		^c				X
Timely and efficient processing	X	X	X	X	X	X	X	X	n.a.	X
Consequences in Performance System^d										
Financial penalties/incentives		X			X					
Publication of results	X	X	X		X					
Targets adjusted across areas		X	X	X	X					

a. See table 2 for more details on employment outcome measures.

b. California has in statute performance measures related to employment outcomes. However, these provisions have never been funded or implemented.

c. Arkansas has a performance measure for increasing the percentage of former recipients who move out of poverty, although this measure has not been implemented.

d. The phone calls with states did not gather consistent information on the consequences in their performance measurement systems, as the focus of the phone calls was narrower.

Performance Measure Definitions and Goals

Performance measures in the states we visited include a variety of outcome measures and other measures. Although the states measure similar outcomes and other program indicators, each state tends to have its own specific definitions for the measures.

Outcome Measures

Employment outcomes. Arkansas, California, Florida, Maryland, Minnesota, New York City, Texas, Utah, Washington, and Wisconsin each have performance measures related to employment outcomes for TANF recipients, such as measures related to obtaining and retaining unsubsidized employment. Each state defines and measures these outcomes differently. For example, while each state has a performance measure of whether TANF recipients are employed, Minnesota measures whether TANF recipients are employed three years from a base period (not whether they have been employed throughout that time). Florida, Maryland, Texas, and Utah measure customers entering employment. Arkansas also measures new job placements, but counts only placements where the client is working 30 hours a week at \$5.15 or higher. Wisconsin measures TANF recipients who secure jobs expected to last at least 30 days. New York City encourages retention by having payment milestones for its vendor contracts for job placements after the 30th, 90th, and 180th day of the placement, while California, Florida, and Texas have retention measures that are separate from the job entry measures. Arkansas also measures retention but for the whole caseload not by individuals. They measure retention as the number of clients employed divided by all clients who have been employed at least once. Washington measures employment by program, including job search, community jobs, and selected education and training programs.

In addition to their measures of job entry and retention, the study states each measure employment wages and some have additional work-related measures that capture a wider range of employment outcomes. Wisconsin measures the number of TANF recipients whose earnings are at least as high as the benchmark wages for their particular area of the state. Wisconsin also has measures to capture improved employment outcomes for TANF recipients who obtain jobs after successfully completing a job skills training program, who switch to higher paying jobs, or who obtain additional jobs to increase their total wages. Maryland measures clients who have experienced earnings gains and separately measures placements in full-time jobs with wages at \$10 an hour or more as well as placements into paid internships or apprenticeships. New York City measures mean and median wages. Arkansas measures closings due to employment. Utah measures increased earnings as well as case closures that are due to increased earnings. To further promote positive employment outcomes for customers, Utah gives caseworkers credit for increasing customers' work hours, even if the hours do not reach the threshold for the federal work participation rate. Minnesota tracks both the number of participants who leave TANF for

employment and their placement wages, although neither is a performance measure with consequences.

Table 2 provides more detailed information about each state's work-related performance measures and data sources. Additional information on the states' use of data is provided later in this report.

Table 2. Site Visit States' Work-Related Performance Measures and Data Sources

State	Name of measures	Description	Data source
Maryland	Earnings gain rate	Percentage of increased earnings over time for employed individuals	Client Automated Resources and Eligibility System (CARES), Unemployment Insurance—reported wages
	Job placements	Total number of job placements	
	Job retention rate	Percentage of individuals who obtained employment in one calendar quarter and remain employed in the following quarter	
	Paid internship/apprentice placements	Total number of participants placed into paid internship/apprenticeship positions	
	Full-time \$10-per-hour job placements	Total number of job placements with 30 or more hours per week at \$10 or more hourly wage	
Minnesota	Self-Support Index	Participant is off cash assistance under certain situations or employed and on cash assistance after three years	Participant-submitted paystubs, TANF eligibility system, Department of Employment and Economic Development—supplied data
	Placement wages	Wages of participant at time of placement	
	Employment exit rate	Percentage of participants who leave for employment	
New York City	Mean and median wage of job placement	Mean and median wages of recipients at placement; included in the VendorStat reports	Welfare Management System (the state's eligibility and case management system); and NYC Work, Accountability and You (NYCWAY) system (employment- and activity information-related system) and other MIS systems; HRA Payment and Claiming System (PaCS) for the paid vendor milestones; and employed are required to submit paystubs or documents from an employer.
	30-day placement rate	Placement in employment within 30 days; included in the VendorStat reports	
	90-day retention rate	Rate of participants' retention in employment after 90 days; included in VendorStat reports	
	180-day retention rate	Rate of participants' retention in employment after 180 days; included in VendorStat reports	
	Qualified reported placements	Reflects cases that were closed due to earnings and cases that had their benefit rebudgeted due to increased earnings; Included in JobStat reports	
	Federal Participation Rate	Federal participation rate; Included in JobStat reports	
	Other measures of employed cases with proper documentation	Measures to encourage accurate and timely rebudgeting of cases with earnings and to keep information on placements; Included in JobStat reports	

Utah	Increased earnings	Customers with increased earnings from entering employment, job retention, or increased wages	TANF case record management system, participant-submitted paystubs, Unemployment Insurance—reported wages
	Positive enrollment closures	Recipients who transition off assistance for employment-related reasons	
	Partial work participation rate credit	Credit that caseworkers receive for increasing customers' hours	
Wisconsin	Job entry	Percentage of unduplicated W-2 individuals who secure full- and part-time jobs expected to last 30 days or more, who obtain a job with higher gross wages, or who obtain an additional job to increase total wages	Customer satisfaction survey results, Unemployment Insurance—reported wages, eligibility system, case management system
	Success of job skills training	Percentage of individuals who obtain jobs within 90 days of successfully completing job skills training	
	Earnings stabilization	Number of individuals who have entered employment and have quarterly UI reported earnings equal to or greater than the benchmark for that area	

Nonwork outcomes. Positive recipient outcomes other than employment are measured in Utah, Texas, Washington, and Wisconsin. Texas, Utah, and Wisconsin have performance standards related to educational attainment, recognizing that educational attainment can improve employment prospects. Wisconsin measures the percentage of W-2 individuals who achieve skills increases (i.e., moving up at least one grade level in basic education, literacy skills, or English as a second language), pass a GED or high school diploma equivalency test, or successfully complete a job skills training course. In Utah, caseworkers are given credit when recipients attain a high school diploma or a GED, although caseworkers are not otherwise held accountable for recipients' educational attainment. Texas measures the percentage of people enrolled in degree or certificate programs who have achieved the certification within three quarters after exit.

Utah and Washington also promote positive outcomes for recipients through measures of case closures. Utah has a measure of positive enrollment closures that includes not only recipients who leave TANF for employment, but also those who transition off assistance for reasons such as ineligibility from child support income and marriage or reunification. Washington has a measure of self-sufficiency exits, which includes the percentage of exits that last three months or more and were for self-sufficiency reasons including child support income, excess net income, exceeding the earned income limit, or the unit requesting closure. They also track (but do not have targets for) long-term exits—those who remain off TANF for 12 months.

Maryland, Minnesota, New York City and Washington also track diversion cases. Maryland measures the number of payments made as part of its diversion program, called Welfare Avoidance Grants (WAG). There are no defined performance expectations based on the WAG reports, but the data are sent to each department. This indicator was formerly one factor in their performance-based funding. Similarly, New York City tracks these cases (called active single issue cases). Minnesota has a four-month diversionary work program and tracks whether those who were on the program are off assistance at 5 and 12 months. Similarly, Washington measures diversion recipients who do not enter TANF within 12 months. Maryland also measures the percentage of cases reaching their 60-month limit in an effort to limit families that reach this limit.

Washington tracks and reports monthly on an extensive set of additional measures developed from program goals.¹⁸ Not all of them have formal targets or are reported at the local or regional WorkFirst program level. Examples of indicators tracked and reported include length of stay; number of cases at 61 or more months of receipt; overlap between TANF and child care receipt in total and by age; number of comprehensive evaluations being done for new, returning, and ongoing recipients; education and work readiness results for recipients; and barriers for new and returning recipients. In addition, Washington has measures focusing on child support orders as a way to increase the income of TANF and post-TANF families. One measure is the percentage of TANF cases with child support orders and the other is percentage of noncustodial parents making child support payments.

TANF subgroup outcomes. A number of states have performance measures related specifically to subgroups of TANF recipients, typically those applying for Supplemental Security Income (SSI). Both Minnesota and Utah consider enrollment in SSI to be a positive case closure, akin to leaving TANF for employment. Wisconsin has a performance standard for the percentage of individuals with pending SSI/SSDI applications who are then awarded SSI or SSDI. Specialized agencies in Milwaukee, Wisconsin, whose main responsibility is assisting families with applications for SSI and SSDI, have additional performance standards related to this subgroup of TANF recipients. These agencies are accountable not only for the percentage of individuals who are awarded SSI or SSDI, but also for completing transition plans for families moving from TANF to SSI or SSDI, and for the percentage of these families with completed applications for the state's Caretaker Supplement Program (a child-only TANF program for children whose parents are receiving SSI). NYC also tracks SSI applications, appeals and receipt in its engagement reports.

Utah has a performance standard related to another subgroup of TANF recipients —those mandated to participate in the state's new job club program. Job club counselors are accountable

¹⁸ These are posted online and can be accessed at <http://www.workfirst.wa.gov/performance/measures.asp>.

for the percentage of job club participants who successfully complete the job club within two to four weeks.

Further, Minnesota adjusts the “range of expected performance” for each county on the Self-Support Index to take into account demographic differences among subpopulations of TANF recipients as well as economic conditions across counties.

Child, family stability, access, or poverty outcomes. None of the study states has implemented performance measures directly related to child outcomes, family stability, access (take-up rates), or poverty. Some state thoughts on the potential for including these measures are discussed later. One example of a state that has considered these measures is Arkansas. Arkansas has established in law since 2003 program outcomes for their TEA program, including “increase the percentage of former transitional employment cash assistance recipients who move out of poverty, including the value of food stamps and the Federal Earned Income Tax Credit and child support.” The Department of Human Services, and later the Department of Workforce Services, was required to develop and maintain indicators for this and other outcomes. However, according to state staff and outside consultants, the state has never been able to develop indicators to measure this outcome with the data and resources available. The state has recently let a request for proposals to review and update their indicators for all outcomes, including work toward developing indicators for this fifth outcome and potentially helping the state think about new outcomes.

Other Measures

Engagement in other activities. Included in Minnesota’s management indicators report is the number of MFIP unaccounted-for cases, defined as the percentage of the nonexempt caseload with no reported activities during the quarter. This measure is used to engage clients with the ultimate goal of moving them off assistance. Maryland includes universal engagement as one of its measures. All clients are required to engage in an activity starting at the time of application.

New York City has a weekly caseload engagement status report, which tracks the number of TANF recipients engaged in employment and other activities, including treatment programs, SSI pending, work experience programs, education/training, and job search. This report is separate from the JobStat and VendorStat reports. New York City’s JobStat also measures success in connecting individuals requiring childcare with a vendor responsible for providing childcare or referral services for child care, which is seen as an important step in preparing for work.

Timely and efficient processing. The states in our study stand out for their use of outcome measures; however, most of these states also have performance measures related to the timeliness, efficiency, or effectiveness of their processes. New York City’s JobStat report contains several measures of the application’s efficiency, including the cash assistance interview

wait time and the cash assistance application timeliness rate. In Milwaukee, Wisconsin, specialized agencies responsible only for W-2 eligibility and assessment have performance standards related to timeliness of interviews, W-2 placements, and completion of screenings and assessment. These Milwaukee agencies, as well as all other TANF agencies in Wisconsin, have a performance standard based on the results of a customer satisfaction survey. The survey includes questions about the responsiveness of the caseworkers and the clarity of the caseworkers' explanations. Maryland has separate reports on the timeliness of application processing for TCA and other programs.

Federal work participation rate. Minnesota, New York City, Utah, and Wisconsin each have urged caseworkers to use measures other than the WPR. While Utah and Minnesota still include the work participation rate among a broader set of measures, Wisconsin has eliminated the work participation rate as a required performance standard within the state. New York City does not include the work participation rate as a measure in its vendor contracts but it is part of its JobStat report for local offices. Maryland tracks the work participation rate both currently and as part of the former JobStat system. Arkansas and Washington also continue to include the WPR in their set of performance measures.

Other Issues

Integration with WIA. The delivery of TANF employment and training services is integrated with WIA programs in some of our study states. However, the integration of service delivery is different from integration of performance measures. Of the states we visited, Utah has the most integrated TANF and WIA service delivery system. The Utah Department of Workforce Services administers both WIA and TANF through its one-stop centers and the two programs share a computer system. Nonetheless, according to Utah officials, the performance measurement system for Utah's TANF program is entirely separate from the WIA performance measurement system.

To learn more about the potential for integrating TANF and WIA performance measurement systems, we had phone discussions with two states, Florida and Texas, which have integrated these measures to a greater degree.

Florida has performance measures for TANF that mirror the WIA performance measures: entered employment, average wages, and retention. According to Florida officials, the Workforce Investment Board decided to use the same measures for all of the workforce programs so that TANF clients would receive the same attention and level of services as clients of other programs. However, unlike with WIA, the state does not set specific performance targets for working with the TANF population. Each region's performance on these measures for the TANF population is compared to the state average and publicly reported.

Texas has the most fully integrated performance measures for TANF and WIA, among the states included in our study. According to Texas Workforce Commission (TWC) officials, the state uses the WIA common measures and other performance measures for all participants in the state's workforce programs, regardless of which state or federal program funds their participation. Although for informational purposes they track outcomes separately for each type of customer, they have integrated measures and targets for serving all customers, rather than setting separate targets for each type of customer. Nonetheless, the TWC has both a systemwide entered employment rate and a subset of that, called At-Risk Entered Employment, for those customers at risk for not having successful outcomes, such as TANF recipients and ex-offenders. The contracts with the Texas workforce boards include several measures in addition to the measures of employment outcomes. For example, Texas also has an educational achievement measure based on the measure applied to WIA youth. The TWC monitors the performance of the workforce boards on the contracted measures and also provides annual cash awards, some of which have considered the workforce boards' performance with TANF customers.

Performance measures versus monitoring. In addition to what each study state described as their performance measures, many states shared information about other indicators and data that they regularly track. Their performance measures were typically what was tracked, associated with targets, connected to consequences, and actively managed. Other indicators took on various forms and purposes in the systems. For example, as mentioned above, several states tracked operational measures of processes including timeliness of application, assessment completion, and due process case resolution. In some places, such as New York City, these measures are included in the formal performance measurement system with targets and consequences. In other places, such as Minnesota, these are tracked for operational management purposes at the local office level, but are not considered by state staff as part of the formal performance management system. Other types of indicators include measures of outcomes or activities that the state thinks are important to monitor and track over time, but for which they do not have developed targets or consequences in a performance measurement system. For example, Washington has an extensive set of indicators that are reported monthly (such as the number of job search placements, recipients' time to employment, and the number of comprehensive evaluations completed), but not all have targets and not all are produced for the local TANF planning areas. State staff described these as being descriptive and informative. In another example, New York City puts out a quarterly workforce development performance dashboard that reports on the work of more than 10 city agencies and departments on a set of common indicators, although some are measured differently. This information is meant to increase cross-agency communication and strengthen the workforce development system. This is an example of using data indicators for broader government wide policy or goal setting.

Performance Measure Implementation

History and Development of TANF Performance Measurement Systems

The timeframe for the development of TANF performance measures in our site visit states ranged from the mid-1990s in Wisconsin to April 2010 in Utah. For both Wisconsin and New York City, the performance measurement systems were natural aspects of their competitive processes for selecting TANF service providers and in New York City, among their local Job Centers as well. The performance measurement systems have been part of their TANF programs since the program's initial implementation, although the systems have continued to evolve. In Wisconsin, for example, this competitive process was mandated by the program's originating statute; however, the measures themselves have gradually shifted over time to become more outcome-focused.

In contrast, Minnesota and Utah each developed performance measurement systems out of a sense that the federal work participation measure was not aligned with the state's goals for TANF. Minnesota first implemented six outcome-focused performance measures for its counties in 1999, and began tying funding to performance on the Self-Support Index in 2003. Although Minnesota does hold its counties accountable for the federal work participation measure, state officials said the goal of adding a measure of county performance on the Self-Support Index was to give counties credit for moving recipients into employment and off the TANF rolls, which are the state's goals for the TANF program. Utah also holds its counties accountable for the federal work participation measure, but state officials said that measure "made us crazy and it didn't make us effective." Utah's performance measurement system is intended to champion the belief that getting jobs is more important than counting hours. In fact, Utah developed its new system through a formal process improvement methodology that involved defining the program's goals, looking for process constraints and meaningful levers, and then ridding the system of any aspects that did not meaningfully further the program's goals. Utah is still implementing its outcome-focused performance measurement system, which does not include financial penalties for counties.

Maryland developed its system in response to concerns by the governor and legislature about low work participation rates, and the anticipation that federal changes through the DRA would be more stringent. In 2003, Maryland began its universal engagement policy to ensure "everyone does something" immediately rather than wait until the end of the initial 24-month period, and implemented JobStat in the beginning of 2004 to measure performance. The state implemented toward the end of 2004 the allocation of some work funds to counties based on performance.

System of Using Performance Measures

The performance measurement systems in the site visit states include an array of strategies for establishing performance targets or benchmarks, reviewing performance, and creating incentives or penalties related to performance. In addition, the states use the measures as part of a broader active management strategy to move toward goals.

Setting performance targets or benchmarks. In most locations, performance targets are set in advance of the period of performance, with some performance measures weighted more than others. In some cases, specific measures or targets are adjusted for demographics or other local conditions.

New York City's JobStat system, which applies to the city's Job Centers, sets a target range for each indicator (low threshold to excellent), and the goal is for Job Centers to meet or exceed the Excellent threshold, but the Low threshold is the minimum performance level. Job Centers get points for their performance relative to these targets ("index" scores), receiving 0 points for falling below minimum, 100 percent of available points for Excellent or higher, or some points related to the distance from the Excellent target. Each indicator has a different weight. The Job Center's "index" scores for each indicator are combined to create an overall score. The overall score and scores on each indicator are ranked among all Job Centers and shown in comparison to the regional and citywide averages.¹⁹ The target range for indicators is generally the same across Job Centers, although special circumstances (e.g., a Job Center largely serving refugees) can be taken into account. The specific range and weights on each indicator are set by the city agency based on a number of factors, including agency priorities, federal regulations, lawsuits, and others. New York City sets targets for vendors through its contracts. Targets for placement rates are the same across all vendors, although staff remarked they were set at an "achievable" rate of 25 percent, which the city found was a challenging but realistic target for the population being served.

Maryland's JobStat target system was similar to New York City's JobStat system, although less complex. Each indicator had a specific goal and different weight. A county's performance relative to goals on each indicator was combined for a total score. Target goals were set relative to state priorities and federal regulations and did not vary across counties.

Wisconsin's contracts with county and private providers include targets for each of the measures, which may be adjusted for local economic conditions or agency type, such as for the specialized local agencies in Milwaukee. Until recently, provider agencies either passed or failed

¹⁹ The actual computation is somewhat more complex. It is described in detail in the publication "JobStat: Program Accountability and Participation Reference Guide," Human Resources Administration, Department of Social Services, New York, New York, <http://www.nyc.gov/html/hra/downloads/pdf/jobstatguide2.0.pdf>.

each measure depending on whether they met the specific targets. A new system currently being implemented scores an agency's performance on each measure as Exceeds, Satisfactory, Needs Improvement, or Fail, depending on where their performance falls relative to the targets. Points are assigned to each level of performance, and the measures are weighted with different numbers of total possible points.

The details of Utah's performance measurement system are still being developed. Baseline data collected during the current transition year will be used to establish benchmarks for the fiscal year that begins in July 2011.

Minnesota sets the county target for its work participation rate at 50 percent (less the reduction in the caseload since the federal base year) or a five percentage point increase from the county's rate the previous year, adjusted for county caseload declines. However, Minnesota is unique among our site visit states in that it does not set prospective targets for county performance on its Self-Support Index. The Self-Support Index is the percentage of adults three years from a base quarter who are "successful," defined as either (1) working 30 hours or more per week (and still receiving cash assistance) or (2) no longer receiving cash assistance (including SSI) but not because of a time limit or sanction. Minnesota establishes a range of expected performance retrospectively, taking into account actual conditions and participant characteristics in the county both currently and at a point three years prior. State administrators compare a county's actual performance (that is, the percentage of adults who meet the success criteria) to a confidence interval (range of expected performance) developed using a logistic regression model and additional statistical techniques. The model calculates an expected range of values for a county's success rate on the Self-Support Index, controlling for about 30 independent variables relating to economic and demographic characteristics. These independent variables include, for example, marital status, race/ethnicity, education level, county population, and county unemployment rate.

Arkansas does not have different targets for TEA program regions in the state, although state staff noted there are very different employment contexts across the state. They expressed interest in moving toward a regression-based adjustment system, but lacked the statistical expertise on staff to implement such a system. They also expressed the difficulties that might be created by the relatively few TEA regions (five in the state) relative to some smaller unit for which adjustments might make more sense. This interest in regression adjustment was at least in part due to staff familiarity with similar systems in DOL programs over time.

Reviewing performance. In most of the states, performance reviews occur at least every few months, although in some cases monthly reports or quarterly reviews essentially are status checks, with counties or offices held accountable for performance over a longer period of time.

New York City publishes the JobStat report once a month on the internet and holds a JobStat review meeting with each Job Center at least once a quarter. At these meetings, top HRA managers meet with Job Center management to review performance. Job Center staff are expected to come into the meeting ready to discuss any areas of low performance and plans to correct them. Positive results and how centers achieved them are also discussed. Within these meetings, performance is compared to past performance as well as compared across Job Centers. City officials described these meetings as a chance to hold Job Centers accountable for performance and to identify problem areas, as well as to provide technical assistance and work out solutions. They are integral to the performance system—one official noted they were as important as the JobStat report. The meetings provide a venue for the city to actively manage and work with Job Centers. Staff from the Data Reporting and Analysis division are also present at the meeting to assist with training or technical assistance.

New York City also reviews performance of its vendors through weekly formal VendorStat meetings between city staff and vendors. Each of the seven vendors is the focus of a meeting every seven weeks. These meetings also serve as a venue to hold vendors accountable, identify problem areas, and discuss solutions.

Maryland also held regular meetings to review JobStat performance with county offices. Since the use of JobStat has been suspended, regular meetings no longer occur for all counties. Ad hoc meetings continue to be held with a few counties that struggle with meeting work participation rate and universal engagement goals.

Wisconsin's system of review is similar to New York City's JobStat model. Wisconsin provides local offices a monthly W-2 summary report, dubbed a green-flag report, which visually displays according to a color scheme whether each office's performance on each measure is failing (red), needs improvement (yellow), is satisfactory (blue), or exceeds expectations (green). The scores for each required performance standard all add into a master scale that totals 100 points if all standards are exceeded. The report compares performance across all of the offices and compares performance from one month to the next. These reports form the basis for monthly KidStat review meetings with top agency officials. Considerable attention is given to the KidStat meetings, which are held in a designated conference room painted with the KidStat logo.

Utah's new performance measurement system ties in to the state's existing computerized dashboard that tracks ongoing performance for each caseworker. The dashboard system gives staff at each level (i.e., executive officials, service-area managers, local managers, and team leaders) the ability to see their own performance and the performance of the administrative units below them. However, peers at each level may not see each other's performance.

Minnesota's Self-Support Index retrospective methodology does not allow for ongoing performance review. However, interim data on the Self-Support Index is provided to the counties after one and two years for informational purposes, even though county performance is officially measured retrospectively after three years. In addition, Minnesota provides counties with data on its management indicators quarterly.

Creating incentives and penalties. The states we visited use different incentives and penalties in conjunction with their performance measures to move toward their goals. To meet performance goals, some states link funding or monetary incentives to performance, some states use competition among offices or vendors, and some use neither. Many of the discussion participants emphasized that people respond to measures and that “what gets measured gets done”—expressing the feeling that the mere measurement and review of the measures caused changes in behavior.

Financial penalties and incentives in the states we visited take a variety of forms. In Minnesota, county TANF funding is contingent on meeting performance standards for both the Self-Support Index and the work participation rate. If a county's actual success rate on the Self-Support Index falls below the range of expected performance, 2.5 percent of TANF funding is withheld until the county submits a corrective action plan. Likewise, Minnesota counties that either do not achieve work participation rates of at least 50 percent participation (less the reduction in the caseload from the base year) or do not improve their rates by at least five percentage points from the previous quarter will have 2.5 percent of their funding withheld until they submit corrective action plans. Maryland, in an effort to focus attention on work, decided in 2004 to reallocate work program funds to local departments of social services, with 25 percent of the formula based on a combination of work-related performance measures (work participation rate, job placements, job retention, caseload decline, universal engagement, and applications denied due to employment). This system is no longer used, in part due to changing priorities, reductions in resources, and the fact that the work participation rate had increased substantially and seemed stable, officials said. For New York City work vendors, payment is contingent on milestone achievements for clients, not per head or per activity.

Financial incentives were used most actively in Wisconsin, where local agencies that meet performance standards have a competitive advantage in renewing their contracts. Specifically, agencies will be able to use their accumulated performance points toward renewing their contracts as well as to compete for new contracts elsewhere in the state. Wisconsin also is the only one of our site visit states to have an active incentive fund for agencies meeting or exceeding performance standards. These agencies share a \$1 million pool according to a formula that accounts for agency size and other metrics. California has in statute performance measures that would give high-performing counties financial rewards, although this pay-for-performance system has never been funded or implemented. The three measures in place are (1) the rate of TANF cases with earnings by county; (2) the federal work participation rate, excluding exempt

cases and including sanctioned cases other than those with mental health, substance abuse, or domestic violence; and (3) cases with earned income three months after leaving assistance. In addition, New York City sometimes has provided successful Job Centers with gift certificates or noncash awards as incentives or rewards.

To create further incentives for positive performance, performance information is shared publicly or among peers (e.g., local offices and caseworkers) in Maryland, Minnesota, New York City, and Wisconsin. New York City in particular reported that competition among Job Centers and among vendors plays an important role in encouraging performance. In Utah, on the other hand, performance of peers is explicitly not available. Office managers can see performance information only for the people they supervise, not for other offices. Utah uses its performance measures only to identify areas for improvement and potential corrective action and does not have any further consequence or incentive scheme.

Beyond the financial penalties, incentives, and competition in the states' performance measurement systems, the attention paid to performance is important. New York City officials noted that a key to their system is making the measures part of everyday activity for the offices, not just at the JobStat meetings. They felt measures and data have to be used in the "regular course of business" to be effective. In addition, New York City's performance-based vendor contracts are actively managed, meaning accountability is beyond just getting paid per milestone. Vendors are expected to be responsive to suggestions and work closely with agency staff throughout the course of the contract, as evidenced by the frequent VendorStat meetings. The vendor we visited appreciated the meetings as an outlet for two-way discussion and input. The vendor also reported using the measures to allocate resources across activities. Officials in our site visit states noted some challenges related to their performance incentives. Minnesota's use of financial penalties has been a controversial aspect of its performance measurement system since its inception. Minnesota officials said there had been consensus about the value of the performance measures until funds were attached to performance. A Minnesota advocate who opposed attaching funding to performance explained that because the work participation rate measure was harmful to client outcomes it should not be incentivized with funding. Moreover, it was difficult for counties to know whether their performance would be within the expected range on the self-support index, so their loss of funding would always be a surprise. She argued that corrective action plans were a more productive approach to poor performance than were penalties because corrective action plans focus on how the county will address the problem. Indeed, a Minnesota county official reported that her county had submitted a corrective action plan to avoid a penalty and that the process had made them pay greater attention to ways their performance had not met the standard. On the other hand, Minnesota officials noted that a few counties have chosen to take the financial penalties for not meeting standards, rather than creating corrective action plans.

In several places, it has been noted that performance-based contracts that tie payments to placements and retention can lead to less emphasis on clients who are harder-to-place, sometimes called “creaming.” New York City notes that their “active management” performance-based system is designed to avoid this through regular meetings with top-level HRA staff on vendor caseloads and performance. In addition, placement rate goals are set at a level that is realistic and challenging. Also, those clients who are either permanently or temporarily unable to work or in need of intensive support services are not included in the Back-to-Work vendor caseloads, but in separate program activities.

Performance measure penalties and incentives also can create the potential for gaming or other unintended consequences. Officials in our site visit states said they consider the possibility of gaming at several points in their performance measurement systems. First, they try to design their measures to minimize gaming. Wisconsin officials said, for example, that they have created more inclusive standards to minimize the number of clients who can be excluded from the performance measures, and they base measures on when data are entered to minimize the potential for backdating. Second, when states are reviewing performance data, they look for unusual patterns that might indicate gaming.

Assessing the Effectiveness of Measures

There are a number of ways that states seek to assess the effectiveness of their performance measurement systems. None of the states we visited had formally evaluated whether their measurement systems improved performance. Nonetheless, during our site visits, states shared their observations of how their performance measurement systems had contributed to changes in their TANF programs, improved outcomes for families, and further refined the performance measurement systems. Overall, discussion participants emphasized that the performance measures serve to focus attention and action on whatever is measured. Each of the site visit states reported that their performance measurement systems had allowed them to more rapidly identify problems and provide technical assistance or work toward other solutions. Several of the site visit states also reported that their performance measurement systems had led local offices to focus more intently on engaging with clients to meet client needs.

States generally assess the effectiveness of the performance measurement systems by assessing how well the states meet their program goals. As evidence of the success of the focus provided by the performance measures, officials in Utah, Wisconsin, and New York City each pointed to the improved employment outcomes for their clients. Maryland pointed to their work participation rate increasing substantially in response to implementing performance measures, among other things (performance-based funding and universal engagement requirements). Minnesota’s performance measurement system led to a very specific change in their program. The state began targeting specialized services to African American and Native American TANF

populations after analyses of subgroup performance on the Self-Support Index highlighted persistent racial and ethnic disparities.

Information collected through state performance measurement systems can also be used to refine the systems themselves, either by revising the measures or by moving the targets. Wisconsin not only revises measures included in each subsequent contract, but may raise benchmarks depending on prior performance. Both Wisconsin and Minnesota officials spoke specifically about how they incorporate feedback from their counties or other providers into refinements of their systems. Minnesota officials gave the example of a county that believed its performance was unfairly hurt by clients moving into or out of the county. The state added a “mover” variable to their regression analysis that, indeed, was an empirically important variable in evaluating county performance. New York City regularly receives input from all involved parties (Job Centers and vendors) on potential changes and has a formal review and revision of the JobStat report every year. Not only are new measures included or deleted, but the order of measures on the report and the wording of labels (e.g., moving to positive versus negative language) have been changed. Agency staff felt strongly that this review was important to keeping the performance system on track. On the other hand, one of the reasons Maryland suspended its JobStat system is that it did not continue to change and had begun to feel “stale” and be less useful.

Implementation Challenges

States identified several issues that could be challenges to implementation, although most thought they were dealing with these issues effectively. Some discussion participants noted the challenge of shifting caseworkers’ mindsets from focusing on processes to focusing on outcomes. They also noted the challenge of shifting the way managers use data to manage, which may be a change for some in the social services field. Other challenges included keeping the measures simple enough to be understandable and usable, balancing the needs of the state’s performance measurement system and the federal reporting requirements, and implementing a performance system at a time of staff reductions and caseload increases.

The interactions between states and counties can also pose challenges. Several state and county officials mentioned minor tensions between counties and states over the performance measures, with counties sometimes believing that the measures did not fully recognize their performance. Some discussion participants acknowledged that controversies over performance measures are inevitable, given the natural resistance people feel toward being evaluated. However, both county and state officials highlighted the value of county involvement and flexibility in implementation decisions, in some cases even in setting performance benchmarks, given the counties’ understanding of conditions and circumstances at the local level.

In addition, several site visit states identified data-related challenges, which are discussed in the next section of this report.

Role of Data

Methods of data collection and data analysis are central to performance measurement systems. The site visit states discussed their strategies and the challenges of identifying, collecting, and using data that are accurate, fair, timely, and useful, without requiring burdensome paperwork. Important aspects of the systems in our study sites were the computer systems for collecting and analyzing the data, the sources of data, and the personnel involved with the data. All of the study sites had in-house staff devoted to data analysis.

Data Collection

The typical starting point for the data in the sites we visited was caseworker entry into a mainframe eligibility system implemented in the 1990s for the AFDC program. One of the exceptions is Maryland, where eligibility information is entered into a 1990s-era system but information on work activities is entered into a web-based system implemented in the mid-2000s. The other exception is Utah, which has a more sophisticated, modern system that automatically incorporates information from clients' online applications as well as information entered by caseworkers. Most states also use data warehouses that pull information from the eligibility and workforce systems and, in at least some places, from other social service programs.

Regardless of the age of the computer systems, each state faces choices about how to collect data to inform its performance measurement systems. For example, Utah uses data from its case management system rather than its eligibility system because the data in the case management system are timely, well documented, and more reliable. New York City officials have chosen to rely almost entirely on data that are already entered as part of the regular course of business. Doing so reduces the burden on local offices and ties the measures to the reality on the ground. On the other hand, New York City officials noted that this approach limits the data available and incurs the risk that frontline workers do not always use codes consistently. Similarly, to improve data reliability and reduce the data-collection burden for TANF staff, Wisconsin and Maryland rely on employment wage data collected through the unemployment insurance system. Although this reduces the data-collection burden for TANF staff and has the benefit of coming from a trusted third-party source, the data lag by at least a few months, creating a challenge for measuring customers' increased earnings. Maryland noted the general difficulty in using performance data that are released with lags. Minnesota and Utah avoid these limitations by relying on the paystubs that TANF recipients submit directly to the TANF office, a practice Minnesota officials cited as critical to the success of their system for measuring employment outcomes. However, Minnesota then faces the challenge of collecting income data

on people who are no longer TANF recipients. Using data collected by other agencies can be useful data-collection strategy, but a number of states commented on the difficulty of establishing data-sharing agreements among state agencies. California faces a data-collection challenge across its counties. Because California's TANF program is county administered and not all counties use the same electronic system, uniformity among data systems and mapping data elements across the counties have been significant challenges. More generally, Utah officials discussed the overarching challenge of balancing the data-collection needs of the state's outcome-focused performance measurement system with the data needs of the federal TANF reporting requirements.

To ensure the accuracy of data used for performance measurement, states have a number of strategies, including training caseworkers on data entry, independently verifying information, and reviewing aggregate data for anomalies. New York City officials said that Job Center managers are motivated to train caseworkers on the proper use of data codes because they want to receive credit for their actions and successes with clients. Wisconsin cross-matches information entered by caseworkers on earnings and SSI with data obtained from the unemployment and SSA systems. Minnesota and Utah, again, verify earnings information by collecting paystubs from TANF recipients.

Data Analysis

Information from the data systems is then extracted and analyzed to provide information and create reports on performance. The extent to which this is automated varies by state and by type of information. In some places, most notably Utah, this is highly automated, allowing caseworkers and managers direct access to extensive information. In other places, data analysts must conduct the analyses periodically or ad hoc. For example, officials in Minnesota noted that the implementation of their performance measurement system completely depends on having central office staff with statistical expertise, particularly to determine the range of expected performance for each county.

In all places, the relationship between statistical and IT staff and policy staff was identified as a crucial aspect of the success of the performance measurement system. Several states mentioned that determining how to best calculate certain imprecise performance measures is a challenge best addressed by data and policy staff together. Collaboration between data and policy staff is important for both old and new computer systems. The older computer systems were designed for caseworkers to provide services, not for collecting performance data, so programmers and data analysts must make adjustments to the systems to meet the current policy needs. Data staff in Utah highlighted the importance of the close relationship among their data and policy staff in working together to design a new computer system that meets the day-to-day needs of caseworkers as well as the needs of their performance measurement system and federal reporting requirements. If systems are not carefully designed to serve these multiple uses, details

such as data field names and the relationships among data elements can cause difficulties both for caseworkers and for data analysis. New York City, Maryland, and Minnesota highlighted the value of the sophisticated data analysts on their staff. New York City has staff familiar with the measures and the data who provide information and assistance in the JobStat meetings. These staff work with Job Center staff to train and educate them on the meaning of different indicators and the nuances of the specific measures. New York City officials emphasized that the role of data in their system is “not a passive technology solution, but an active management solution.” Minnesota officials said, similarly, that a minimum requirement for a successful use of data is a “spirit of collaboration.”

IV. Lessons Learned

The site visit states identified several factors that they believed contributed to the success of their performance measurement systems.

Few and Clear Performance Measures

Officials in New York City and Wisconsin emphasized the importance of limiting the number of measures and keeping them simple. Having a limited number of simple measures helps focus the attention of caseworkers and administrators, and makes it easier to identify high and low performance. People we spoke with in Minnesota made these same points by remarking that the complexity of their measures presents a challenge to understand the underlying methodology. The two measures for which Minnesota counties are held accountable—the work participation rate and the range of expected performance on the Self-Support Index—have a statistically complex methodology, providing clear goals but limiting their usefulness for day to day management and making them difficult to understand, especially for those outside the system.

Active Management of Performance

New York City and Wisconsin each highlighted the importance of having active management and steady attention for the performance measures. Both New York City and Wisconsin have regular meetings with top-level agency staff to review the JobStat, VendorStat, and KidStat reports. Not only does this keep attention on the measures, but New York City officials noted it also fosters working relationships and collaborative efforts to improve performance. Similarly, officials in Wisconsin cited the KidStat meetings for helping the state better identify struggling agencies in need of technical assistance, as well as identifying agencies in a position to share best practices.

Commitment to Using Data

The commitment and capacity for using data to inform policy were key factors in the success of the performance measurement systems in each of our site visit states. An upfront investment in data analysis systems gives states the ability to track, analyze, and interpret performance data. More important than the sophistication of the computer systems is the sophistication of the analytical staff. Minnesota representatives noted that their commitment to collecting and analyzing data allowed them to use data to inform policy despite the challenge of having a legacy data system. In addition, Utah has drawn heavily from longitudinal research of their TANF population and program to inform the development of their performance measurement system. Maryland's work performance specialists (state staff detailed to local

offices to help with data and technical assistance) were seen by the state and local level as keys to improving the work performance rate.

Priority of Top Managers

While New York City and Wisconsin noted the value of including top managers in their review meetings, several Utah discussion participants identified the leadership and support of their agency's executive director as a key to the success of their revised performance measurement system. Her clear and vocal support of the revised system has allowed staff at all levels to have the confidence to risk changing their ways of doing business.

Commitment to Evolution

The importance of having a system that can evolve was discussed explicitly in New York City, Minnesota, and Wisconsin. Goals, measures, and relative weights of the measures are revisited and revised with each subsequent contract in Wisconsin (every two to three years) and annually in New York City. This evolution allows for better alignment with broader program goals as well as more responsiveness to program and data changes.

V. State Recommendations for the Federal TANF Performance Measurement System

Site visit participants shared their own suggestions for improvements to the federal TANF performance measurement system and their thoughts on challenges or limitations to potential program changes. Site visit participants expressed frustration with the federal work participation rate measure. They wished instead to be held accountable through measures that focus on customer outcomes, especially employment, and that allow states greater flexibility to meet the varied individual needs of their customers.

Improvements to the Federal TANF Performance Measurement System

Many discussion participants suggested specific changes to the work participation rate measure that they believe would improve outcomes for their clients. The most common suggestions included (1) expanding the list of countable activities, (2) increasing the flexibility around the number of hours of engagement, such as allowing for setting required hours of engagement according to clients' individual capacities, (3) counting educational activities toward a GED or high school diploma as core activities, and (4) reducing the administrative burden of verification and reporting requirements. While most site visit participants had issues with the current federal work participation measure, some feared that changes to the measure, especially if they included increased or different reporting requirements, could increase the burden on states with little improvement in client outcomes. New York City officials said they were not interested in federal requirements expanding beyond the work participation rate, because the philosophy at HRA is that engagement in work, except in unusual circumstances, is preferable to engagement in non-work activities.

Implications of Hypothetical Changes

In our site visit discussions, we identified a number of hypothetical changes to the federal TANF performance measurement system and asked the discussion participants to comment on their implications. These hypothetical changes included adjusting federal targets for state conditions; measuring improvement over time rather than comparing performance to a fixed target; ranking states without financial penalties; and measuring child well-being, poverty, or take-up of TANF among eligible families. The initial response was that there must first be agreement on *what* to measure before any of these changes in *how* to measure would matter. In other words, ranking states or measuring their improvement is only as meaningful as the underlying goals and measures of the performance measurement system.

Adjusting Targets by State Economic or Other Circumstances

Discussion participants in several of our site visit states agreed that adjusting federal targets for state economic conditions, and possibly for demographic differences, would be useful, especially for evaluating performance on outcome measures like job entries or earnings. However, one state official argued against such adjustments, saying they were a way to hide poor performance.

Measuring Improvement over Time, Rather Than Comparing Performance to a Fixed Target

Discussion participants noted if states were measured on how well they help families progress, it would create positive incentives for helping the most challenged families, in contrast to the current incentive (with the work participation rate) to not serve these families. On the other hand, officials noted some challenges inherent in measuring performance over time, such as the potential for gaming and the need to account for states starting at a high level of performance. In addition, some noted the need to account for deteriorating external conditions, such as a recession.

Ranking States on Performance without Imposing Financial Penalties

States would always like to avoid financial penalties; however, they had nuanced responses to the possibility of disconnecting funding from performance. One issue was whether states would continue to respond to a performance measure that did not have the threat of a financial penalty. A number of officials argued that while some state governors care about rankings and would respond to a performance measure even without penalties, other state governors would stop responding and the measure essentially would be ignored. A second issue centered on the validity of the rankings themselves. For the most part, officials argued that the states differ so greatly on program rules, political climates, economic conditions, and other factors, that fair comparison would not be possible. They argued that rankings would promote gaming and a focus on the unfairness of the rankings rather than on the end goals.

Adding Measures Related to Poverty, Access, or Child Well-Being

States had mixed reactions to the idea of adding measures related to poverty, access (take-up), or child well-being. Several people mentioned that there would first need to be a debate over whether poverty reduction and child well-being are or should be goals of TANF. Across the states, much of the opposition to the idea of adding measures of poverty and child well-being centered on the argument that TANF alone can only do so much to affect these outcomes in the face of racial inequality, failing education systems, and low-paying jobs, especially with fixed funding. On the other hand, some people said that adding these broader measures would create positive incentives for state TANF programs to work more closely with other programs to achieve broader social outcomes, and that an increased focus on support for

families and children would be nice, including for child-only cases. Nonetheless, some states felt the state should (and in some cases does) monitor these factors but they should not be added as performance measures.

VI. Conclusions

The states participating in this study believe that performance measurement is an important element of their state TANF programs. From multiple states and respondents we heard that measurement leads to increased focus on what is measured —“measurement matters”—and that this focus can be usefully targeted to promote action toward program goals. The study states are focused particularly on employment outcomes. The flip side of this, mentioned in many places as well, is that measurement can draw negative or unintended focus, taking attention away from other goals that are not being measured or putting too much focus on how to meet the details of a measure—the letter of the law—but perhaps missing the intention. These must be balanced in any successful system.

It is clear from our discussion with states that most do not feel it is possible for other states to implement their system entirely without changes. For example, Minnesota described the process necessary to build stakeholder buy-in for their adjustment measure. Other states remarked on the need to have a culture that embraces data and performance measurement among staff, and on the key role of leadership in shaping that culture. However, several elements of these states’ systems could provide useful lessons and examples for other states. In particular, the study states provide examples of how to use data from existing systems to measure performance and how to structure an active management process that keeps attention dynamically focused on desired outcomes.

Overall, this study shows a number of states have implemented performance measurement systems in their TANF programs well beyond the federal WPR measure. The states we visited unanimously believe their measurement systems are helping or have helped them actively manage performance to better reach their goals for employment and other outcomes. The lessons learned from these states provide important information for other states considering expanding performance measurement in their TANF programs and for the federal government.

Appendix A. Expanded State Summaries

This section describes in detail the performance measurement system in each state visited.

Maryland

Performance management system history. Maryland's TANF program (Temporary Cash Assistance or TCA) is supervised and administered by the state through county social services departments that may use vendor contracting. Maryland includes outcome- and process-based measures through reports originally developed when the state shifted to a universal engagement model in 2003, anticipating the Deficit Reduction Act. To support this model, the state adopted a now-suspended performance management process based on New York City's JobStat model, with monthly, county reports scoring performance on a range of indicators and a combined score.

Summary of performance measure types. Despite the suspension of the formal JobStat system, the state still tracks the measures originally reported in the now-suspended JobStat templates. Employment outcome measures include percentage of increased earnings over time for employed individuals, total number of job placements, job retention rate (defined as the percentage of individuals who obtained employment in one calendar quarter and remain employed in the following quarter), paid internship/apprenticeship placements, full-time \$10-per-hour job placements, and specific submeasures of the work participation rate (such as a 24-month and 60-month activity participation rate).

Maryland also has a performance standard for measuring universal engagement that tracks federal- and state-defined work activities. Caseworkers at the time of application give nonexempt participants a schedule of agreed upon activities dubbed an independence plan. Local departments determine the range of work activities offered, and work-eligible adults are required to participate in for a minimum of 30 hours per week unless they have children under age 6. The universal engagement rate is the percentage of participants meeting this broader set of activities requirements, with the goal for each county set at 100 percent.

Other measures focus on diversion assistance and administrative-related efficiency. State officials pull reports that track the number of payments made as part of Maryland's diversion program, called Welfare Avoidance Grants. There are no defined performance expectations based on the WAG reports, but the data are sent to each department. TCA application timeliness rates (the percentage of cases processed within the set limit) and TCA payment accuracy rates both were tracked and compared against goals in the JobStat report templates. Currently, these

rates still are tracked, but with the suspension of JobStat they are not compared against goals or presented in dashboard-type reporting system.

Performance measure implementation. Prior to the JobStat phaseout, the report template included administrative- and outcome-based indicators. Staff set goals and assigned a weighted distribution of points for meeting expectations on each indicator; these points were tallied in a combined score on a 100-point scale. State staff had regular meetings with local offices to discuss their performance, areas that needed improvement, and ways to improve. The state placed work participation specialists in local offices to help work with data, provide technical assistance on understanding the measures, and improve performance.

With the JobStat template phasing out, officials still track the performance indicators and set goals but no longer calculate or disseminate an overall score. They no longer hold regular meetings with counties to address performance on the work measures, with the exception of counties that continue to struggle. County officials, in addition to seeing state indicators, are able to see how their county compares statewide. State officials said they use the performance indicators to identify areas in which counties are struggling and offer assistance.

Within the past two years, state and local TCA staff began tracking a new collection of outcome-focused indicators centering on employment—full-time \$10-an-hour job placements and internship/apprenticeship placements (sometimes referred to as subsidized employment placements) furnished through the Maryland RISE initiative. Maryland RISE is connected to the StateStat system introduced by Governor Martin O’Malley and tracks the payment accuracy rate, federal work participation rate, and year-to-date job placements, among other administrative numbers.

For a time, the local allocation of work program funds was at least partly tied to performance on a number of outcome-focused employment indicators, including job placements, universal engagement rates, and work participation rates. Counties were grouped in one of three tiers, with different incentives tied to each tier and with metropolitan counties eligible for a greater share. Funding incentives tied to performance were discontinued due to several factors: the system was overly complicated, money ran out for the contract to assess county performance, and counties achieved marked progress statewide on the work participation rate.

Administrators’ assessment of effective use. State and local administrators said JobStat, especially the employment outcome indicators, likely played a significant role in Maryland improving its work participation rate along with the hiring of 24 work participation specialists assigned to individual counties, whose sole focus was to improve participation. Also, Maryland RISE and the shift in focus to higher-paying jobs with defined career advancement likely helped improve the long-term outlook for TANF families and placed some recipients in better jobs, administrators said, although advocates did not necessarily agree.

With JobStat discontinuing, Maryland is becoming less focused on the work participation rate, instead almost exclusively focusing on application and payment error rates (especially in the past year), a local TCA manager said.

Key challenges and elements of success. Working with an outdated legacy data system is a key challenge, as are extended data lags, for indicators such as job retention and earnings gains, state and local officials said. Also, tying a positive correlation between recipients' outcomes and improvements to the work participation rate is a challenge, due partly to data shortcomings, local officials said.

State administrators said highlighting universal engagement and participation in an approved activity from application instead of at 24 months of benefit receipt was a positive force behind improved participation and participant's outcomes. Hiring work participation specialists also played an important role in that effort.

Advocates of some of the populations typically served by TCA said there still may be a disconnect, despite the state's emphasis on universal engagement and other outcome-based measures. For example, an advocate said the decision to assign participants to work activities at application sometimes fails in implementation because employment counselors still tend to decide whether a client is exempt based on what's best for the work participation rate instead of what serves the client best.

Additionally, advocates identified several challenges to implementing the performance system, including reductions in Family Investment Administration staff coupled with an increase in caseload that generally stretches resources too thin, and staff focusing on completing administrative paperwork to the detriment of case management. In addition, TCA clients are not the only ones eligible for Maryland RISE positions and are not always informed of or included in this effort.

Minnesota

Performance management system history. Minnesota's additional performance measurement began in January 1999 when it created a county performance reporting system that included six measures it considered outcome-focused: (1) percentage of the caseload with budgeted earnings, (2) percentage employed and receiving only the food allotment from MFIP²⁰, (3) percentage leaving MFIP during the quarter, (4) median placement wage rate, (5) federal work participation rate, and (6) countable months remaining before the 60-month limit.

However, these measures were never tied to county funding. Citing deficiencies in the above six measures and a desire to better gauge county performance by accounting for different demographic and economic characteristics, Minnesota began overhauling its performance measurement system in 2001, resulting in the current system in 2003. Since 2003, Minnesota has tied a portion of county TANF funding to two performance measures. One is the county's performance on the Self-Support Index, a measure based on the percentage of adults three years from a base quarter who meet a positive outcome, defined as either (1) working 30 hours or more per week (and still receiving cash assistance) or (2) no longer receiving cash assistance (including Supplemental Security Income) but not because of a time limit or sanction. The other is the county's performance on a measure that parallels the federal work participation rate.

Performance measure implementation. For the Self-Support Index, Minnesota establishes a "range of expected performance" retrospectively, taking into account actual conditions in the county both currently and three years prior. State administrators compare a county's actual performance (that is, the percentage of adults who meet the success criteria) to a confidence interval ("a range of expected performance") developed using a logistic regression model and additional statistical techniques. The model calculates an expected range of values for a county's success rate on the Self-Support Index, controlling for about 30 independent variables relating to economic and demographic characteristics. The model currently controls for the following variables:

- | | | |
|--------------------------|-----------------------|-------------------------|
| • Marital status | • Mental illness | • Chemical dependency |
| • Needing an interpreter | diagnosis | diagnosis |
| • Moving within past 3 | • Coming from another | • SSI adult in the case |
| years | state | • County population |
| • SSI child in the case | • Student status | • County unemployment |
| • Race/ethnicity | • Two-adult case | rate |
| • Amount of child | • Age of the youngest | • County child poverty |
| support | child | rate |

²⁰ Minnesota has a waiver from the USDA Food and Nutrition Service to combine the cash and food benefits in their MFIP program, which is why they measured those receiving only the food portion of MFIP.

- Number of children in case
- Months of shelter subsidy
- Proximity to urban center
- Immigrant status
- Education level

Program managers and analysts derived the 30 independent variables through large workgroup meetings that met regularly for about six months. Administrators developed an initial list of about 100 potential variables imagined to possibly affect a county's success. From the list, workgroup participants began winnowing the variables for three main reasons: (1) data didn't exist, (2) variables were highly correlated, or (3) variables were otherwise deemed redundant. The ensuing 30 variables were then adopted in the model. "We didn't rule anything out that we had data on that made sense," administrators said.

Since officials engineered the model to adapt to the future addition of independent variables, the list at first changed frequently, nearly from quarter to quarter, but has stabilized in the past three or four years. The frequent turnover in variables during the index's early years prohibits comparisons of performance over time for values before 2005. However, the model's stability since 2005 enables year-to-year performance comparisons after this later date.

The state's second performance measure, the work participation rate, is Minnesota's estimate designed to match as closely as possible the federal work participation rate, but at the county level. Minnesota sets the target for its work participation rate at 50 percent, adjusted for county caseload declines. Prior to the Deficit Reduction Act, Minnesota's calculation of the state's work participation rate measure included all TANF-eligible participants (including some cases not included in the federal work participation rate calculation) but now includes only cases that would be included in the federal calculation, as well as county caseload reductions. State and county administrators said individual caseworkers may be held accountable to the county for meeting the work participation rate. The work participation rate data are calculated monthly but made available quarterly, to allow time for data corrections.

Additionally, despite not being tied to funding, the other measures included in the MFIP management indicators report allow for comparisons among counties. For example, each county's terminations are shown in absolute numbers, rates, percentage points from the area mean, and percentage points from the state mean. Further, counties may use these and other indicators to assess the performance of employment service providers or caseworkers, local and state officials said.

Data management. Most of Minnesota's data used in the state's performance management system are derived from frontline workers' coding in the TANF eligibility system, save for some data drawn from employment services, other programs such as Medicaid and child care, and demographic data on poverty, unemployment, and urbanicity, state administrators said. For example, the regression model behind the Self-Support Index uses data from these other

programs. Much of the data are self-reported by applicants, with some items verified by caseworkers such as recipients' Social Security numbers, paystub records, and citizenship status. After collection, the data are routed to a data warehouse where analysts may use them with statistical software packages.

Administrators' assessment of effective use. Analyses of the demographic control variables included in the Self-Support Index have been helpful in highlighting persistent racial and ethnic disparities. The regression model showed that race and ethnicity explained much more of the variation in the defined success rate than did county economic conditions. This awareness has targeted specialized services to African American and Native American populations, state officials said.

A key benefit to controlling for demographic and economic variables in the Self-Support Index measure has been the ability to more fairly evaluate counties against each other, according to state staff. For example, two counties may have Self-Support Index averages drastically different from each other on an absolute scale (even by tens of percentage points), but by controlling for variables judged beyond the counties' control, their adjusted performance may be similar. By comparing each county's performance with a statistically derived range of how it "should" be doing, some counties previously thought to be performing poorly actually were found to be exceeding expectations, and some counties with high success rates were deemed not meeting expectations. As a result, some counties with seemingly poorer performance have been tapped to share their strategies for overcoming adverse conditions and exceeding expectations.

More broadly, state and local administrators said that the Self-Support Index measure helps focus attention beyond the federal work participation rate, which they criticized for leading caseworkers to focus on meeting the rate at the expense of serving their clients' needs.

Administrators said that for some counties, the financial penalties for not meeting the measures may be insufficient for them to develop or implement policies necessary to bring the performance up to acceptable levels. Similarly, some counties apparently determine that the resources needed to submit a corrective action plan to the state, which would restore the withheld funding, are not worth the effort. Nonetheless, one county administrator reported that developing a corrective action plan has helped improve her county's performance.

Key challenges and elements of success. State and local TANF officials and a TANF advocate from Minnesota said a significant challenge of the Self-Support Index measure is its complexity. The statistical/econometric formulation of the range of expected performance makes the measure hard for some people to understand. Its three-year retrospective nature (requested by counties to allow time to work with participants) also limits its usefulness for managing the TANF program at the caseload level, a state administrator said. Further, implementing the Self-Support Index measure requires statistical expertise, and the definition of success can be a

political one, an advocate and some state workers said. Other critiques and challenges of the Self-Support Index measure include the imprecision of some control variables and the possibility that a county's demographic characteristics may be taken as a reason for a county to accept the status quo.

A key to the Self-Support Index measure's success has been collaborative development involving state, county, and provider representatives during the formative stages and on a standing advisory committee. According to respondents, the diverse stakeholder involvement has fostered a sense of buy-in at all levels. Another significant success factor has been the state's commitment to and capacity for collecting and analyzing data to inform policy, though the state still relies on a legacy eligibility system. The ability to link data across programs and the availability of earnings data from participants' pay stubs have been critical, data staff said. Finally, the evaluation of racial and ethnic disparities has motivated work to decrease the gaps between racial and ethnic groups and has greatly helped the Self-Support Index measure succeed, according to state staff.

New York City

Performance management system history. New York City's Human Resources Administration (HRA) implemented its performance measurement system shortly after welfare reform.

Performance measure implementation. On JobStat, indicators are reviewed and set yearly. The Office of Data Reporting and Analysis posts JobStat reports online monthly, with data at the Job Center, region, and city level. These reports include an index score that represents a weighted composite of all JobStat performance indicators. The weights for each indicator are also reviewed annually. For each indicator, HRA sets a goal, also reviewed and set annually. HRA recognizes the differences in the populations served by each Job Center. Area and caseload demographic information are included in each JobStat report; depending on the population served, some centers do not use certain indicators and some centers are not expected to perform as well on certain goals. Staff from each Job Center come to HRA for regular JobStat meetings. At these roughly quarterly meetings, Job Centers discuss issues that have surfaced on the JobStat report with HRA staff, agreeing on plans for corrective action. Job Centers are expected to come into these meetings with plans to correct poor performance on indicators. Initially HRA tied performance on the JobStat report to management bonuses, but has since switched to using gift certificates and noncash awards to reward high performance. Several HRA staff mentioned the important role of public reports and competition between Job Centers in motivating high performance. Centers can request additional or disaggregated data (at the client or caseworker level) to identify problem areas and guide management.

For VendorStat, HRA holds performance-based contracts with seven human services organizations to provide job readiness, training, and placement services to TANF applicants and recipients. Like with the JobStat report, HRA reviews and sets performance measures and goals for Back to Work vendors yearly. The Office of Data Reporting and Analysis posts a monthly VendorStat report including performance data at the site, vendor, and all-vendor levels. Payments to the vendors are tied to performance on several key VendorStat report indicators, including the creation of an employment plan, 30-day placement in employment, and 90- and 180-day retention in employment. In addition to operating under performance-based contracts, HRA holds weekly VendorStat meetings, with each vendor the focus of the meeting every seven weeks. The meetings include key staff from HRA and the vendor organization and provide a forum to address any issues that arise on the VendorStat report. HRA staff also spoke about the role of competition between vendors, who can see each other's reports, in motivating performance. HRA staff saw management of the vendors through these meetings in conjunction with the VendorStat reports as key to making this performance measure system for vendors work.

In addition to JobStat and VendorStat, to get a full picture of the engagement of recipients, additional data tracking reports are generated including an engagement report, an employment report, and reports for each relevant HRA program area. TANF cases that are permanently or temporarily not employable are served through other NYC TANF programs, such as WeCARE (Wellness, Comprehensive Assessment, Rehabilitation and Employment), for which separate data reports are kept.

Data management. Data for all of HRA's programs are collected by the Management Information Systems (MIS) department. For TANF two main systems are used, with information coming from front-line staff. These systems are uploaded on a weekly basis into the Enterprise Data Warehouse (EDW). The Office of Data Reporting and Analysis (ODRA) generates its TANF-related reports, including JobStat and VendorStat largely from EDW, although other MIS databases also contribute to the reporting of data on TANF-related programs. The warehouse also contains historical data back to 1999. (Back to Work program vendors are paid through another data system, the HRA Payment and Claiming System (PaCS). HRA's investment in the operational databases, the EDW, and the multiple computer programmers and analysts at ODRA allows the accountability process to be so data intensive.

Administrators' assessments of effective use. HRA sees JobStat and VendorStat as tools to work toward their goal of moving individuals off of cash assistance and into full-time employment. HRA assesses the effectiveness of the JobStat and VendorStat systems by the extent to which these goals continue to be met. HRA staff said they believe that performance measurement has created local accountability for client outcomes.

HRA believes that performance measurement allows them to identify weak performing Job Centers and Back to Work vendors and to work collaboratively to correct issues. HRA staff report that Job Centers feel held accountable at JobStat meetings and see the system as a useful way to know how they will be evaluated and to get an objective measure of their performance. HRA staff also mention that Job Center regional staff use the JobStat reports to manage locally, to learn where they are underperforming and where to target additional resources.

Vendor staff also talked about the performance measurement system and the VendorStat reports and meetings as a valuable forum for bringing together all parties interested in program outcomes to create solutions for underperformance. Vendors felt that communication has been enhanced and problems cleared up faster under the VendorStat system. HRA has weighted the VendorStat measures to push vendors toward rapid engagement with clients, and they reported vendors respond to the goals communicated through performance measurement. Vendor staff also mentioned that the introduction of performance measures and performance-based contracting changed the tone of service delivery from a social work model to a more corporate model, for both workers and administrators.

Key challenges and elements of success. Many of the challenges officials cited center on qualities inherent to performance-based contracting. For example, the practice requires large changes in business process, including more analytical and business-minded managers. Also, contractors struggle to cover the cost of assessment and engagement since the bulk of payment comes when a client meets the 30-day placement milestone and the 90- and 180-day retention milestones.

Additionally, using administrative data can be challenging for performance management. Workers don't always use codes consistently, not all outcomes that should be tracked are available, and decisions about how to calculate certain measures can pose problems. Accounting for the needs of special demographic groups and for redundant data entry also poses problems. The challenges of serving certain populations that are disproportionally represented in specific local offices, such as homeless persons, people with criminal records, or refugee populations, are difficult to reflect in the goals to which each vendor and center are held. HRA adjusts targets to account for the special nature of these populations. Since vendors maintain their own data systems, they must key in client information twice, once into the vendor system and once into the HRA system.

From an upper- management perspective, successes of the performance management system include greater flexibility, performance comparison, and program analysis. The flexibility to revisit and revise indicators, weights for each indicator, and yearly goals allow HRA to refocus the measures to align with broader program goals. A limited number of measures are central to focusing program administrators and workers, and to making incidences of high and low performance easier to identify. Additionally, participation of high-level agency staff in management meetings increases the sense of accountability, and upfront investment in systems and analytic staff builds the ability to track, analyze, and interpret performance data.

Changes brought on by the system also positively affect vendors and local offices, administrators said. First, the system uses data already entered as part of the regular course of business, reducing the burden on vendors and local offices and tying measures closer to the reality on the ground. Second, combining performance measurement with active management (JobStat and VendorStat meetings) based on the performance data fosters both working relationships between the agency and the local office/vendor staff and collaborative efforts to address low performance. Also, performance-based contracting allows vendors greater flexibility on budget decisions than line-item contracting. Finally, having analytic staff at the central office to implement measures and participate in performance measurement meetings means local staff can be easily trained to understand and use performance data.

Utah

Performance management system history. Utah began redesigning its TANF program in April 2010 with a Work First model that included an overhaul of its TANF performance measurement system. Before the redesign, performance measures for caseworkers focused on the federal work participation rate measure. While the WPR countable activities are still included among Utah's performance measures, the state has broadened its focus to include additional performance measures that officials say better help move customers into employment. The new performance measurement system, still in transition, focuses on TANF customers' outcomes, including not only employment and earnings but also educational attainment and case closures for positive reasons. The new performance system ties in to Utah's existing computerized dashboard that tracks performance for each caseworker. Full implementation of the new system will depend on how quickly the underlying computer programming can be completed.

Summary of performance measure types. Standards include measures for increased earnings and other income, recipients who transition off assistance, partial credit for hours in countable activities toward the work participation rate, high school and GED completion, compliance, quality, and completion of a new job club program.

- *Increased earnings* measures customers increasing income through entering employment, job retention, and increased wages. Caseworkers continue to maintain contact with customers for six months after case closure to help address transition issues and retention.
- *Positive enrollment closures* tracks recipients who transition off assistance for employment-related reasons including earning increased income that makes them ineligible.
- *Increased income* tracks customers with increased income from child support, marriage/reunification, or SSI.
- *High school and GED completion* credits caseworkers for customers' educational attainment, though caseworkers are not necessarily held accountable when this doesn't occur.
- *Compliance and quality* automatically notifies a state compliance review team when certain actions, such as case sanctioning, are taken. Supervisors review a sample of each caseworker's cases quarterly and assess whether the actions taken were most appropriate for the customer.
- *Completion of job club* measures state job club counselors on the percentage of customers who successfully complete the job club within two to four weeks.

Performance measure implementation. A mismatch between the federal work participation rate and positive outcomes for customers spurred state administrators to augment Utah's TANF performance indicators. Under the leadership of the executive director of the Utah Department of Workforce Services, Utah embarked on a formal assessment of its TANF program

goals, processes, and constraints, drawn partly from a longitudinal study of Utah TANF customers conducted by the University of Utah.

The TANF redesign began with thinking about what the TANF program goals should be and revising the program and performance measurement system to promote those goals. Executive staff, regional managers, and local supervisors and caseworkers crafted the new indicators in a series of work groups. Many of the measures still are in development, as are their possible application and consequences in some cases.

No formal penalties or incentives are connected to the performance measures. Staff at each level (i.e., executive officials, service-area managers, local managers, and team leaders) may see their own performance and that of the administrative units below them. However, peers may not see each other's performance. In lieu of state officials adjusting the measures for regional or local factors such as high unemployment, regional managers will have the flexibility to set targets within their own area.

Data management. Utah's data management system allows for a broad collection of indicators and variables, although officials do not always have the capability or resources to robustly analyze everything that is collected, administrators said. "When it comes to data we're at a pretty good advantage because we get a ton of it," one worker said. The state's data warehouse draws from the TANF eligibility and case record management systems; sources include auto-population of some fields, with employment counselors directly entering much of the information as they meet with clients. All the tasks the employment counselors perform are documented in this system.

Officials track, analyze, and use the data to implement the TANF performance measures with an IBM-published business intelligence and performance management software package dubbed COGNOS. The software is not a report in itself, but instead manages and reports the data from the warehouse. The software package draws data from the TANF eligibility and case management systems, and allows users to view reports in a dashboard-type layout. As of March 2011, staff were moving reports from a legacy software package to the new system, and several features such as the ability to view reports within the department on the division level were only recently introduced.

Administrators' assessments of effective use. Representatives from every level of management said that implementing standards beyond the work participation rate has already begun to shift their focus toward improving TANF customers' outcomes. Several discussion participants observed that before the measures broadened, caseworkers were focused on the best outcome to affect the work participation rate and not necessarily the best activity for the customer. Despite the new measures not being fully implemented, officials said that based on limited point-in-time data samples, the shift appears to be resulting in more appropriate

placements (including greater transition to unsubsidized employment), improved staff morale, and better customer case management services.

Regionally, several managers credited the new measures with recent improvements in moving customers to unsubsidized employment. A senior manager said there are very positive expectations for the new measures' effectiveness and results. However, firm data-backed results are not available because of the short time since the implementation of the new measures.

An early negative result attributed to the switch is a noticeable decline in the work participation rate, mentioned as a concern by most of the discussion participants. Officials reported that they expected the work participation rate to rebound as the new system and programs (e.g., job club) move out of the transition phase. At the same time, discussion participants repeatedly referred to the commitment that the executive director has made to promote the best outcomes for recipients regardless of the federal work participation rate.

Key challenges and elements of success. An early challenge has been tempering concerns among caseworkers after seeing the drop in the work participation rate, which senior officials so far have stressed will not affect dedication to the new measures, and which to some extent was expected.

A challenge for senior administrators and data managers has been balancing the data needs for the state's outcomes-focused performance measurement system and for federal reporting requirements. In addition, data lags for employment wage data, which is delayed at least a quarter, have been a significant issue for measuring customers' increased earnings.

Most state administrators said it is too early to tell if the transition to broader performance measures is or will be successful (beyond early, anecdotal positive evidence related to outcomes). However, several officials said just the act of changing the focus to improving customer case management and moving recipients to employment is a benefit.

Several discussion participants identified the executive director's leadership and support as key to the success of the revised performance measurement system.

Performance management system history. Performance measures beyond the WPR have been around in Wisconsin since the implementation of TANF. Since the beginning of the program, Wisconsin selected local W-2 (TANF) agencies competitively. Starting from the first contract, the Department of Children and Families looked at Job Opportunities and Basic Skills Training Program (JOBS) performance, and agencies that had done well with JOBS were given the opportunity to run the program without having to compete, except in Milwaukee. Wisconsin's supplemental measures evolved from this beginning with each successive two-year contract with local W-2 agencies. While many of the performance standards at the very beginning of W-2 were process-based, there has been a shift over time to outcome-based indicators.

Summary of performance measures. Standards include measures for job entry, earnings stabilization, success of job skills training, educational attainment, and customer satisfaction survey results.

- *Job entry* measures the percentage of unduplicated W-2 individuals who secure full- and part-time jobs expected to last 30 days or more, who obtain a job with higher gross wages, or who obtain an additional job to increase total wages.
- *Earnings stabilization* measures the number of individuals who have entered employment and have quarterly UI reported earnings equal to or greater than the benchmark for that area.
- *Success of job skills training* measures the percentage of individuals who obtain jobs within 90 days of successfully completing job skills training.
- *Educational attainment* measures the percentage of W-2 individuals who improve skills (i.e., moving up at least one grade level in basic education, literacy skills, or English as a second language), pass at least one GED or high school diploma equivalency test, or successfully complete a job skills training course.
- *Customer satisfaction survey results* include clients' opinions on caseworkers' responsiveness and the clarity of the caseworkers' explanations.

Additionally, in Milwaukee, a separate agency type is responsible only for W-2 eligibility and assessment. These agencies have performance standards related to timeliness of interviews, W-2 placements, and completion of screenings and assessment, in addition to the customer satisfaction survey.

Performance measure implementation. For a description of the state's contracts with local providers, please see Setting Performance Targets or Benchmarks on page 22. Benchmarks for the targets are set mostly in percentages; however, the customer satisfaction survey indicator is gauged in an average score on a five-point scale. The measures are compared across local offices in a monthly report dubbed a W-2 Summary Report, which visually displays according to a color

scheme whether an office is failing, meeting, or exceeding expectations, and compares performance from one month to the next. The report, a monthly local office comparison, was referenced extensively by both state administrators and a local agency official as being especially important to monitor and track performance. This information eventually is made public.

Top-level managers of the Department of Children and Families review the performance of each W-2 agency on these measures at monthly KidStat meetings, where they also review performance for the other programs the department oversees. At the KidStat meeting, the division administrator presents charts showing each W-2 agency's performance against the benchmarks for each performance standard.

Wisconsin uses the performance measures beyond the WPR to identify issues among struggling local W-2 agencies, identify agencies for technical assistance, and to create an incentive for competition in future contracts. On this latter point, if an agency meets acceptable standards, it receives preference in a points system to be used in a subsequent W-2 contract. Agencies that meet or exceed the standards also share a \$1 million incentive earnings pool. Any agencies that meet or exceed their performance standards will share the pool, according to a formula that accounts for agency size and other metrics. The agencies can use the funds to reinvest or innovate in any ways they choose.

Data management. The majority of the data used in Wisconsin's standards are pulled from the state's eligibility and case management system, which is generally derived from frontline caseworkers directly entering information. However, data related to some measures is cross-matched with other programs' systems, such as unemployment insurance data for the earnings stabilization measure and SSI receipt for the Milwaukee agencies whose main responsibility is assisting families with applications for SSI and SSDI. Customer satisfaction survey results are another source of data for some standards. The state's eligibility and case management mainframe system dates to 1994 and was adapted to fit the TANF program after being originally developed for AFDC, data administrators said. Staff are modernizing the case management side of the system, which officials said is expected to streamline data entry. Most reports drawn from the mainframe may be drilled down to the participant level.

Administrators' assessments of effective use. The performance measures have affected the identification and facilitation of struggling agencies or agencies requiring technical assistance. KidStat and other reports are central to this effort, state officials said. The system has aided in refining and fine-tuning performance measures to be included in future contracts. Also, the measures have naturally focused attention at the local level, and have shifted the focus from process- goals to outcomes, which state officials believe better relate to desired goals. State officials said the local agencies use the increased data availability to better target deficiencies or drill down to the individual caseworker or team level. From Department of Children and

Families staff regularly looking at KidStat, local providers have been more cognizant of their performance, DCF staff said.

Department of Children and Families staff said a dedicated study has not been conducted recently to gauge impact, but they have observed improvement in the rate of participant job entries. By inherently focusing on outcomes rather than process measures, DCF said agencies are better focused on the W-2 program goals. Also, DCF has been encouraged by a high correlation of actually finding problems within a provider when the measures and reports suggest the provider is not meeting acceptable standards.

Key challenges and elements of success. The central challenge cited by state administrators to implementation and effectiveness of Wisconsin's performance measures was data lags. For example, firm job entry data are not available for at least 60 days after a reporting period, which has affected how the state targets the related measure. However, administrators did not describe these lags as a major obstacle, nor did the lags prevent the new performance measures' use or affect perceived future effectiveness. Another challenge officials mentioned was keeping the measures simple and not overwhelming in number or technical complexity—in other words, making them understandable and usable.

But because the standards have evolved in response to changing needs, the data are not conducive to longitudinal comparisons—a frustrating limitation, DCF staff said. Nonetheless, the data are informative on whether an agency performed poorly (or well), both on the prior contract's standards and the current contract's standards.

A key success cited by administrators was the flexibility and evolution of the state's performance measures. While not necessarily met with enthusiasm by local agencies, Wisconsin's willingness to evolve the measures has made agencies more responsive to the program and data changes. KidStat and the W-2 summary reports have helped the state better identify struggling local agencies in need of technical assistance, and to identify and share best practices among providers.

In addition, Wisconsin officials see the graduated grading system (as opposed to the prior pass/fail system) as promoting improvement among all agencies.