

PROPOSED PERFORMANCE MEASURES AND STATE RESPONSES: ANALYSIS AND NEXT STEPS

Prepared for the
Integrated Performance Information System Project Meeting
Seattle, WA

Christopher T. King
Sarah Looney

June 2004



Lyndon B. Johnson School of Public Affairs
The University of Texas at Austin
3001 Lake Austin Blvd., Suite 3.200 Austin, TX 78703
512.471.7891

This paper was prepared for the National Governors Association Center for Best Practices and Washington State's Workforce Training and Education Coordinating Board with funding provided by the U.S. Department of Labor's Employment and Training Administration. The views expressed are those of the authors and do not represent the official positions of NGA, Washington State, or the University.

BACKGROUND

There have been numerous efforts to improve the way performance is measured for workforce development and related programs in recent years, most of which have focused more on programs rather than systems. Over a decade ago, amendments to the Job Training Partnership Act fostered more comprehensive measurement for job training programs, even as the Government Performance and Results Act (GPRA) mandated that all federal agencies develop and implement broader measures of their strategies, among other efforts. In 2002, the Office of Management and Budget (OMB) directed key federal agencies responsible for administering education, employment and training programs to use a number of “common measures” of performance. The status of the OMB common measures effort is somewhat uncertain, in that some agencies and programs have embraced them more fully than others. In the meantime, a number of leading-edge states — and some local areas — have continued to work toward more comprehensive measurement and management frameworks.

The Integrated Performance Information (IPI) Project seeks to provide the U.S. Department of Labor (DOL) with input from the states on what is required to support integrated information on the results of an array of workforce investment programs and the One-Stop Career Center system.¹ At the first IPI Project meeting held in New Orleans in early January 2003, multi-program policy teams from the participating states² identified seven key areas of interest to policy and program leaders, as follows:

- ❑ Outcomes for employers and the economy,
- ❑ Labor market outcomes for program participants,
- ❑ Social welfare outcomes,
- ❑ Customer satisfaction,
- ❑ Skills gains,
- ❑ Return on investment, and
- ❑ Subgroup and comparative information.³

Technical staff from states participating in the IPI Project met in Miami the first week of April 2004. In response to the information needs identified by policymakers during the January IPI meeting, the state technical teams proposed 30 performance measures and then prioritized these measures through a nominal voting process (see Table 1). After the meeting, participants took these measures back to their home states to solicit feedback and further analyze them.

¹ See Ellen O’Brien Saunders and Bryan Wilson, *Integrated Performance Information (IPI) for Workforce Development System Planning, Oversight and Management*, Olympia, WA: Workforce Education and Training Coordinating Board, November, 2003.

² The participating IPI states are Florida, Michigan, Montana, Oregon, Texas, and Washington State. Louisiana, Minnesota, Missouri, New York, and Pennsylvania also have attended as observers.

³ See Dan O’Shea, Sarah Looney and Christopher T. King, *The Alignment of Workforce Performance Measures and Policymaker Needs in Ten Leading-Edge States*, Austin: Ray Marshall Center, LBJ School of Public Affairs, The University of Texas at Austin, March, 2004.

TABLE 1: PROPOSED WORKFORCE PERFORMANCE MEASURES

	Measure	Definition	Votes	Vetoes
Outcomes for Employers and the Economy	4. Market Penetration / Employer Participation	Number of employers who use the system/All employers	30	
	20. Productivity (with a base – employer group/sector)	Gross revenue for employers/Number of employers	7	
	24. Skill Match	New hires in target industry/Vacancies in target industry	5	
	25. Labor Market Match	Number of job vacancies/Number of employed plus number of vacancies	4	
Labor Market Outcomes for Program Participants	1. Employment	All those with earnings in the 1 st quarter after exit/All served who exit	33	
	7. Entered Employment	All those with earnings the 1 st quarter after exit/All unemployed at registration who exit	21	
	16. Percent of Cohort Employed	Customers employed in the 2 nd quarter/New customers receiving services in the 1 st quarter	11	
	2. Retention with Same Employer	Number of individuals employed with the same employer two quarters after exit/Number of individuals exiting the system who are hired	33	
	8. Retention	Those employed the 5 th quarter after exit/All employed in the 1 st quarter after exit	19	
	26. Employed Third Quarter After Exit	Those employed minus those enrolled in higher education/All exits	2	
	12. Wage Level	Total earnings in the 1 st quarter after exit/All those employed the 1 st quarter after exit	16	
	15. Earnings	Earnings in the 2 nd quarter after exit/Number of people employed in the 2 nd quarter after exit	12	
	5. Earnings Gain	Average earnings in the 5 th quarter after exit/Average earnings in the 1 st quarter after exit	26	
	11. Earnings Change	Post program average quarterly earnings in 3 rd quarter minus pre-program average quarterly earnings one quarter prior to registration	15	
Social Welfare Outcomes	3. Decrease in Public Assistance	Percent of exiters who receive public assistance during 3 rd quarter after exit/Percent of exiters who received public assistance in the 2 nd quarter prior to registration	32	
	19. Social Welfare	Post-program public assistance minus pre-program public assistance/Cost of the program	8	
Customer Satisfaction	9. Employer Satisfaction (Repeat Customers)	Number of employers who return/Number of employers who receive any service in a given quarter of the year	19	
	9. Employer Satisfaction	Number of employers satisfied/Total number of employers responding	12	
	21. Mystery Shopping	Number of customers at a Center/Total number of customers	6	
Skill Gains	10. Literacy/Numeracy	Number of individuals who increase one or more education skill level/Total number who exit plus all who have completed one year in the program.	16	1
	18. Skill Gains for Exiters	Number of successful completions/Number of participants exiting the program who have not been in the program for three quarters	8	
	27. Skill Gains	Number of successful completions in all programs/Adult population in the jurisdiction	1	
	22. Skill Gains	Number of successful completers/Total number of participants enrolled	6	
Return on Investment	12. ROI	Tax revenue gains from extra wages/Cost of the program	12	
	6. Cost Savings	Net income taxes and public costs avoided/Total cost	25	
	28. Cost savings to Employers – survey information	<i>Not yet defined</i>	1	1
	29. ROI	Wages of cohort minus wages of control group/Cost of the program	1	1
Other	17. Participation Rates, 18-60	Number of students in post secondary/Number of total population who could be involved in post secondary	9	1
	30. Percent of Eligible Enrolled in the Program	Number enrolled in program X/Number eligible for program X	0	

Source: *Integrated Performance Information Project Academy Meeting*, Miami, FL, April 2, 2004.

State comments on the 30 proposed measures are summarized in this report, providing advantages, disadvantages and brief discussions for each. Researchers from the Ray Marshall Center for the Study of Human Resources at the University of Texas at Austin's LBJ School of Public Affairs provide additional analysis of each proposed measure. We considered a variety of factors in reviewing each of the proposed measures, including:

Rationale: What information is the measure designed to provide? How does this information help policymakers and other stakeholders? Does the value of the information provided by a measure balance with the amount of effort required to calculate it?

Feasibility: Are data readily available to support the proposed measure? Are these data expensive or labor-intensive to retrieve?

Accuracy: Does the indicator accurately and dependably provide the information it was designed to produce? What factors could limit the accuracy of the measure? Does the measure create incentives to "game" or "cream" program participants to produce more desirable outcomes?

Validity: Does the measure or indicator predict the desired near- or longer-term impact?

Systems Focus: Is the measure applicable to all of the programs in the workforce system? Is it scalable? Does it treat diverse customer populations equitably?

Time Frames: Time frames – including both the length of the measurement period and the point in time at which measurement occurs – can dramatically affect outcomes. Also, what is the value of near-term v. long-term measures?

Definitions: Definitions are the final ingredients of well-designed performance measures. Definitions must be exact, especially when used for comparative purposes. Key terms that need to be more clearly defined for this effort include participant; exiter; completer; public assistance; and satisfaction.

In the next section, we introduce each of the proposed measures, summarize their pros and cons, and provide brief discussion. In the following section, we address key measurement issues and gaps, after which we present a series of possible next steps.

ANALYSIS OF THE PROPOSED MEASURES

This section discusses the proposed measures grouped according to the needs articulated by state policymakers at the first IPI meeting in January 2004 in New Orleans. Their definitions are shown in italics. For the most part, the presentation of advantages and disadvantages reflects comments made by participating states; the discussion that follows each measure is a mix of state commentary and interpretation by Ray Marshall Center staff. A matrix containing all 30 proposed measures, their definitions, and a detailed summary of their advantages and disadvantages is provided as an appendix.

Outcomes for Employers and the Economy

MEASURE #4: Market Penetration/Employer Participation (30 votes)

Number of employers who use the system/All employers

Advantages:

The sole advantage cited by states was that this measure indicates the extent to which employers are aware of and find workforce services of sufficient value to use.

Disadvantages:

States expressed concern that this measure would be inaccurate because the denominator includes employers who are not presently hiring, that it would mainly reflect changes in supply and demand, rather than changes in the quality of service, and that it is a process not an outcome measure. They also pointed to measurement difficulties and geographic identification issues.

Discussion:

Some suggested that a better penetration measure might be: the number of employers who use the system/Number of employers who report an increase in number of employees on UI tax records. Others suggested imposing a minimum employer size (based on employment, not revenue), since small businesses may never require workforce services. A market penetration measure would contribute at least one employer-related measure to those now favored by DOL. Some states (e.g., Missouri) are already moving towards use of market penetration measures as part of a “scorecard” approach to show the extent to which workforce services are reaching employers in their states and workforce areas. It is one of the easier indicators to measure accurately and consistently via administrative records. With 30 votes, market penetration is also a measure that has broad support among the states.

MEASURE #20: Productivity (with a base – employer group/sector, 7 votes)

Gross revenue for employers/Number of employers

Advantages:

This measure may be useful for tracking sectoral strategies.

Disadvantages:

Disadvantages of the productivity measure focused on the fact that it was not scaleable, that it would vary mainly due to factors outside the control of workforce programs, and that it was not relevant to workforce development generally. One state pointed out that a median rather than an average would be more insightful.

Discussion:

This measure might be difficult to measure accurately and consistently. It could also lead to confusion on the part of employers and policymakers because “productivity” typically is defined by the Bureau of Labor Statistics in terms of output divided by man-hours. It would provide at least one employer-related measure. This measure received only modest state support as indicated by the number of favorable votes.

MEASURE #4: Skill Match (5 votes)

New hires in target industry/Vacancies in target industry

Advantages:

The only advantage cited was that this provides some measure of whether the demand for new employees in target industries is being met.

Disadvantages:

States were concerned that this only tracks *hiring*, not whether the target industry can hire the *skills* it needs and that it would be difficult to measure accurately and consistently.

Discussion:

As defined, this is not a measure of “skill match,” but more a sectoral or target hiring efficiency measure. The existing data series for vacancies are not as well developed as those for employment, unemployment and other variables. This measure received little support among the states.

MEASURE #25: Labor Market Match (4 votes)

Number of job vacancies/Number of employed plus number of vacancies

Advantages:

This measures the number of job vacancies relative to the size of the economy.

Disadvantages:

This measure can be affected by macroeconomic trends: the number of vacancies falls in recessions and rises when labor demand increases. They are also difficult to measure accurately. Labor exchange captures some but not all vacancies. A survey would be necessary. It may be difficult to measure the number employed and the number of job vacancies using the same data source, definition, and time period. Are people holding more than one job counted twice in the denominator as employed, or just once? Is the observation made at a point in time or over time (e.g., quarterly)?

Discussion:

As defined, this measure does not capture “match” but rather the degree (percent) of unmet demand in the labor market. It is unclear what this measure would contribute to measuring the performance of state and local workforce systems.

Labor Market Outcomes for Program Participants

MEASURE #1: Employment (33 votes)

All those with earnings in the first quarter after exit/All served who exit

Advantages:

This measure is easily understood, relevant to most workforce funding streams, has a relatively short time lag, and provides information wanted by policymakers.

Disadvantages:

Disadvantages cited include that first quarter outcomes are not good (valid) indicators of long-term outcomes, especially for career and technical education, that the measure is biased in favor of work-first over skills development programs, and, as defined, it includes people employed at registration, so it does not prove a causal relationship between programming and obtaining employment.

Discussion:

Employment is one of the central or core measures of workforce program performance and is one of the OMB common measures. It should not unduly favor one service strategy over another (e.g., work-first). It also resonates with policymakers, administrators and the general public alike and is easily and reliably measured. Researchers have questioned the validity of termination-based placement measures for predicting long-term impacts, but little research has been done using first-quarter employment.⁴ This measure also better reflects patterns of use by those seeking work, many of whom are already employed but seeking better or at least different jobs. This measure had the highest support among the states with 33 votes.

MEASURE #7: Entered Employment (21 votes)

All those with earnings the first quarter after exit/All unemployed at registration who exit

Advantages:

This employment measure was cited as being useful for programs that primarily attempt to help the unemployed quickly obtain employment (e.g., “work-first) and scaleable, as well as being similar to the OMB common measure.

Disadvantages:

Disadvantages cited included that it is not a meaningful measure for programs with many participants who are employed at registration, not useful as a consumer report measure to inform the general public attempting to choose a training program, and creates a misleading impression that the program(s) *caused* employment. In addition, first-quarter follow-up is too short to capture employment outcomes, especially for participants in educational programs.

Discussion:

This measure is similar to measure #1, which computes employment based on the universe of participants served, not just those who were unemployed at registration. It is less reflective of participants and services in today’s labor markets than measure #1, though it is an OMB common measure. Support among states for this measure was only about two-thirds as strong as for the more encompassing employment measure (21/33).

⁴ Barnow, Burt. S. (2000), “Exploring the Relationship between Performance Management and Program Impact: A Case Study of the Job Training Partnership Act,” *Journal of Policy Analysis and Management* 19(1): 118-141.

MEASURE #16: Percent of Cohort Employed (11 votes)

Customers employed in the second quarter/New customers receiving services in the first quarter

Advantages:

The main advantage cited for this employment measure was that it minimized time lags relative to others.

Disadvantages:

States pointed out that this measure is poorly defined, biased toward strategies emphasizing quick job placement, penalizes participants in programs lasting more than 2 quarters, and is logically problematic in that it measures two cohorts and is not tied to an exit quarter unless programs are comparable in length. Some pointed out that adult learning “cohorts” may take seven years to earn a certificate.

Discussion:

This measure received less support from states and is unlikely to resonate as well as other employment measures with policymakers, administrators or the wider public. It attempts to capture employment for new short-term participants in immediate placement strategies, but it is unclear whether it offers advantages over measures such as #1 and 7.

MEASURE #2: Retention with Same Employer (33 votes)

Number of individuals employed with the same employer two quarters after exit/Number of individuals exiting the system who are hired

Advantages:

States stated only that this measure provides an indication of the value to employers of hiring program exiters.

Disadvantages:

Disadvantages highlighted included that this measure needs further definition (e.g., “hired” v. 1st quarter employment, multiple employers), does not acknowledge that employees change employers for both good and bad reasons, and is inherently biased against migrant and seasonal workers. It may also be difficult to identify employers due to changing employer identification numbers.

Discussion:

One state suggested that the measure be renamed “employer retention.” This measure is parallel to current employment retention measures except that it rewards short-term retention only with the same employer. CETA, JTPA and WIA retention measures, as well as the OMB common measure have not imposed the “same employer” criteria, better reflecting the fact that individuals change jobs for good as well as bad reasons and that, as research suggests, the key to longer-term success is *employment*, not *employer* stability. However, this measure garnered the highest support of any measure, tying support for employment measure #1 with 33 votes.

MEASURE #8: Retention (19 votes)

Those employed the fifth quarter after exit/All employed the first quarter after exit

Advantages:

Advantages cited for this measure were that data (UI wage records) to capture it are readily available, it provides some indicator of the long-term (i.e., 5th quarter) result for an important outcome, the fifth quarter-first quarter comparison provides a logical, simple one-year follow-up period, and avoids differences due to seasonality.

Disadvantages:

Disadvantages cited were that this measure necessarily has a long lag time, misses those individuals who became employed later, and the combination of using entered employment and retention as measures may create a misleading impression of success. For example, both the entered employment and the retention rates could be 80%, but this would mean only 64% of all exiters were employed the fifth quarter after exit.

Discussion:

One state suggested the measure be renamed “participant retention,” while another questioned whether both a short-and longer-term employment retention measure should be maintained. One preferred this to measure #2, though support for this measure was less (19 votes) than for other similar measures.

MEASURE #26: Employed Third Quarter After Exit (2 votes)

Those employed minus those enrolled in higher education/All exits

Advantages:

Key advantages cited by states were that this measures employment in a follow-up period of sufficient length to indicate that the outcome is lasting and attempts to not penalize educationally oriented programs as some other measures do.

Disadvantages:

Disadvantages included that the measure lacks a clear rationale, has a long lag time (i.e., the 3rd post-exit quarter), and undercounts employment since many people hold jobs and go to school at the same time.

Discussion:

While the intent of the measure is clear about not penalizing programs for participants who are enrolled in higher education rather than employment, the language is not clear. It probably makes sense to exclude those enrolled in higher education from the denominator of this measure as well. Programs that prepare students for further education (like high school vocational programs or WIA youth programs) could be placed at a disadvantage under either definition of the denominator. Use of the National Student Loan Clearinghouse, a national data repository, might make identification of higher education enrollments possible if it can be made available to all programs/systems. Note that this measure garnered almost no support from states, receiving just 2 votes.

MEASURE #12: Wage Level (16 votes)

Total⁵ earnings in the first quarter after exit/All those employed the first quarter after exit

Advantages:

Advantages cited for this measure were that it is a simple, easily explained measure that policy makers want, and earnings level is probably the single most important result to program participants. In addition, the measure is scaleable, has a minimal time lag, and can help indicate whether average wage is a “living wage.”

Disadvantages:

States were concerned that the median would be a better measure than average because it is not subject to outliers and indicates the approximate earnings that most people receive. In addition, they felt the first quarter after exit was too soon to indicate lasting results, creates a bias toward programs that emphasize quick job placement services rather than training, and was a particularly poor time period for secondary career and technical education participants.

Discussion:

States also questioned what standard should be used to identify whether the wage level is acceptable or not. This is an unambiguous, policy-relevant measure with high resonance among a wide array of stakeholders. Since average quarterly earnings are measured only for those employed, it communicates relevant information to participants, policymakers, administrators and the wider public. As John Baj, a researcher at Northern Illinois University’s Center for Governmental Studies, has pointed out, earnings are more stable measured over multiple quarters rather than just a single one. The first post-exit quarter may in fact be too soon to capture stable earnings for participants, especially those in education and training activities. The measure received moderate support from states with 16 votes.

MEASURE #15: Earnings (12 votes)

Earnings in the second quarter after exit/Number of people employed in the second quarter after exit

Advantages:

As with the previous wage measure, this earnings measure has the advantages that it is easy to explain, salient with policymakers and the public, and is applicable to a wide range of workforce programs.

Disadvantages:

Disadvantages cited were that the median rather than the average would be a better measure, since for many workforce programs only about a third of participants earn at or above the average, so the average is misleading.

Discussion:

States suggested publishing breakouts by earnings level (e.g., earning less than full-time minimum, between full-time minimum wage and full-time at or above \$9/hour. There was

⁵ There was a suspected typo or misstatement in this measure. It read “average” rather than “total.”

concern that it was unclear what standard should be used to identify whether the wage level is acceptable or not. This appears to be a compromise earnings level measure, reflecting earnings in the 2nd post-exit quarter for those employed in the quarter. It garnered moderate support with 12 votes.

MEASURE #5: Earnings Gain (26 votes)

Average earnings fifth quarter after exit/Average earnings first quarter after exit

Advantages:

The sole advantage cited was that this measure provides some indication of longer-term results on a key outcome, i.e., earnings gain.

Disadvantages:

States expressed concern that individuals who stay employed typically experience earnings increases, so this measure could be significantly above 100% without the program(s) having induced any positive net *impact*. The measure also implies a norm of raises during the first year of employment post-program and fails to demonstrate effectiveness in restoring pre-unemployment earnings, since it focuses only on the post-program period. Additionally, not all earnings are readily available (e.g., Department of Defense, Office of Personnel Management, U.S. Postal Service), and this measure may create disincentives to serve older workers, students, and persons seeking part-time employment.

Discussion:

States suggested this gain measure is inappropriately named and should be modified to note that this it only applies to those employed. States again suggested using median instead of average earnings gains and conducting research to determine which time periods are best for this kind of measurement. This earnings gain measure received very high levels of support from states with 26 votes, greater than any other gain/change measure. As pre/post (or even immediate-post/longer-post) measures, *all* earnings/wage gain measures are problematical, in that they attribute all gains/changes to participation despite the fact that most individuals will return to their permanent earnings trajectory sometime after exit. This measure may be preferred to simple pre/post measures that use the quarter immediately prior to registration for the pre-program earnings level.

MEASURE #11: Earnings Change (15 votes)

Post program average quarterly earnings in third quarter minus pre-program average quarterly earnings one quarter prior to registration

Advantages:

Two advantages were cited for this earnings change measure, that data should be readily available and that it offers policymakers something they want to know: how much did earnings increase?

Disadvantages:

States listed many disadvantages for this measure. Not only it does not answer the question, how much do people make, but the quarter prior to registration is often atypically low in earnings due

to unemployment, creating a false impression of the extent of the gain and likely promoting “creaming” in participant selection. It is also difficult to judge the value of a pre-post earnings gain without a context in terms of earnings *levels*. Moreover, pre-post measures imply that programs caused the change, but a net impact analysis (with random assignment or a quasi-experiment) is necessary to measure this accurately. Finally, a pre-post earnings gain measure is not useful as a consumer report measure for individuals who want to know how much they are likely to make after training when trying to decide what course of study to take.

Discussion:

The listing of disadvantages is relatively complete. Measuring earnings gains from the first quarter prior to program entry will produce the greatest gains but they will be falsely measured due to the well known pre-program “earnings dip” documented by economists starting with Ashenfelter (1978). This is one of the least appropriate gain measures as a result though it garnered moderate support from the states with 15 votes. It is also close to the OMB gain measure.

MEASURE #23: Earnings Gain (6 votes)

Post program average quarterly earnings at the third quarter/Average quarterly earning the second quarter prior to registration

Advantages:

This gain measure offers the same advantages as the others, providing policy makers something they believe they want to know—how much did earnings increase—and data are readily available to measure it.

Disadvantages:

States expressed many disadvantages for this measure as they did for other gain measures. Changes over time in the results on pre-post measures of earnings are greatly affected by changes in earnings during the pre-program period, so it is difficult to judge the value of a given pre-post earnings gain. In addition, pre-post measures create a false sense that the programs caused the change and are not useful as consumer report measures for individuals who want to know how much they are likely to make after training when trying to decide what course of study to take.

Discussion:

This gain measure is similar to other measures (i.e., #s 11 and 15), but is for a different time period. While all gain measures are problematical, using a more distant pre-program quarter is an improvement over the other earnings gain measures, but only slightly. Moving to the third quarter post-program is also an improvement. This measure did not garner significant support from the states with only 6 votes.

Social Welfare Outcomes

MEASURE #3: Decrease in Public Assistance (32 votes)

Percent of exiters who receive public assistance during third quarter after exit/Percent of exiters who received public assistance in the second quarter prior to registration

Advantages:

This measure addresses an issue salient with policymakers and the public.

Disadvantages:

States cited many disadvantages for this measure, including that reduced public assistance is not relevant for all programs/populations, that the measure is misnamed, actually measuring the percentage of exiters who *continue* to receive public assistance, and that it would be strongly influenced by eligibility policies that vary by state. In addition, “public assistance” is not adequately defined, possibly including TANF, Food Stamps, housing and other forms of assistance. It is unclear whether the numerator only includes persons who received public assistance at the time of program enrollment or also includes new recipients during the program. Using this measure could discourage “public assistance” programming that helps individuals obtain employment such as child care and transitional medical care.

Discussion:

States suggested that the measure could be reworked as a “reduction-in-cash-assistance” measure and might be better stated as a ratio of absolute numbers rather than percentages. This measure would be good for TANF as second quarter would get us past the 4-month income disregard, but other services do not fit well. Receipt of public assistance fluctuates with the business cycle, declining in good times and increasing in bad times. In addition, state assistance and workforce targeting policies both have substantial effects on assistance entries and exits that have very little to do with the effectiveness of workforce services.⁶ States and local areas might be induced to (further) “cream” among eligible assistance recipients with such a measure. This measure received very strong support from the states with 32 votes.

MEASURE #19: Social Welfare (8 votes)

Post-program public assistance minus pre-program public assistance/Cost of the program

Advantages:

This measure has the advantage of recognizing the value of reduced dependence on public assistance, as does the previous measure.

Disadvantages:

Disadvantages cited were that the measure was not applicable to all programs, creates a false impression that programs caused the pre-post difference, and is difficult to measure. It also could discourage day-care, temporary medical assistance, etc. unless better defined to only include cash assistance programs. In addition, the cost of these programs is very difficult to determine, making the denominator hard to measure.

Discussion:

This measure is not well named: “social welfare” does not reflect what this measures, namely the pre/post reduction in cash welfare payment expenditures for the cost of workforce program

⁶ See Rebecca M. Blank (2002), “Evaluating Welfare Reform in the United States,” *Journal of Economic Literature*, Vol. XL (December), pp. 1105-1166.

services. This might also be viewed as a return-on-investment measure focused on welfare costs. The measure garnered little support among states with only 8 votes.

Customer Satisfaction

Measure #9: Employer Satisfaction (Repeat Customers) (19 votes)

Number of employers who return/Number of employers who receive any service in a given quarter of the year

Advantages:

States felt this measure of employer satisfaction provided some *objective indication* of the extent to which employers find workforce services of sufficient value to use again.

Disadvantages:

Disadvantages cited for this measure were that it may be misleading since the government has a near monopoly on many workforce services (e.g., labor exchange) and because an employer may return for service because it was unsuccessful the first time. Also, unique employer identifiers are not readily available in all states, the time frame for return is not defined in the measure and small employers may return infrequently.

Discussion:

One state indicated that customer service measures should be used at the point of service, suggesting some support for a more qualitative satisfaction indicator. Note that far more states voted for an objective measure than the typical qualitative (#14).

MEASURE #14: Employer Satisfaction (12 votes)

Number of employers satisfied/Total number of employers responding

Advantages:

States indicated that this measure provides insight into employers' satisfaction with the workforce system.

Disadvantages:

States also listed many disadvantages, including that satisfaction was subjective, costly to make scaleable across programs and to the local level, assumes consistent survey methodology to achieve comparability across programs and states. Also, satisfaction data may not be easy to obtain. Satisfaction surveys generally have low response rates, also indicating dissatisfaction but not captured in this measure.

Discussion:

This more traditional measure of employer satisfaction parallels usage under WIA, though is somewhat different in that it would report the share of employers who were satisfied rather than an average satisfaction score on a standard instrument, e.g., the American Customer Satisfaction Index now used with workforce and related programs. As defined, it is unclear whether this is intended for use as a point-of-service measure as well as one for area and state use. The measure also received moderate support with 12 votes.

MEASURE #21: Mystery Shopping (6 votes)

Number of customers at a Center/Total number of customers

Advantages:

No advantages were listed for this measure.

Disadvantages:

Several disadvantages were noted including that this measure was subjective, not applicable to many workforce development programs, was difficult to measure consistently across programs, labor intensive and expensive to collect. Also, one stated noted that the numerator should be “number of mystery customers satisfied” and the denominator the “total number of mystery customers”.

Discussion:

This measure is problematical in part due to the fact that few states have any “history” or “track record” with it, which likely contributed to the absence of any advantages. In addition, the measure received very little support with only 6 votes.

Skill Gains

MEASURE #10: Literacy/Numeracy (16 votes, 1 veto)

Number of individuals who increase one or more education skill level/Total number who exit plus all who have completed one year in the program

Advantages:

States noted that this is the only direct measure of skill gain on the list, and that skill gains are essential to developing a skilled workforce. They also noted that this could be a good youth measure as well as demonstrating the value of WIA Title II programs. It is also an OMB common measure.

Disadvantages:

The noted disadvantages were that a literacy/numeracy measure was irrelevant to programs that do not provide basic skills instruction (e.g., work-first programs), and that persons taking a semester off could negatively impact this measure. There were also questions raised about data availability in light of recent interpretations of the Federal Educational Rights and Privacy Act (FERPA).

Discussion:

One state suggested the denominator be changed to be the number of participants registered during the program year. This is one of the only measures to gauge skill gains and thus respond to some of the concerns raised by team members from secondary and postsecondary educational agencies as well as those workforce policy makers focused more on education and training oriented strategies. As noted, it is also one of the OMB common measures for youth and will be used by DOL for WIA and related programs. This measure received moderate support from the states (16 votes) but also drew one of only four vetoes.

MEASURE #18: Skill Gains for Exiters (8 votes)

Number of successful completions/Number of participants exiting the program who have not been in the program for three quarters

Advantages:

The sole advantage states cited for this measure was that policy makers are very interested in (skill?) retention rates.

Disadvantages:

Disadvantages cited by states included that this measure was not clearly defined (e.g., “completion”), that completion does not equal skill gain, and that it is more important how many newly prepared workers there are than this measure of retention. They also noted that the logic of the denominator is not clear, e.g., why not include all exiters?

Discussion:

This measure lacks clarity and would need significant work before states were comfortable with it. It received low support as well with only 8 votes.

MEASURE #27: Skill Gains (1 vote)

Number of successful completions in all programs/Adult population in the jurisdiction

Advantages:

The sole advantage cited for this skill gains measure is that it provides some measure of the number of skill credentials granted relative to the size of a participant population, i.e., adults in an area.

Disadvantages:

As with the previous measure, states cited many disadvantages, including that “successful completion” is difficult to define for many programs, that it is more important how many newly prepared workers there are, that completion counts may differ from counts of newly prepared workers, and that this measure could have scalability problems to the extent that completions may occur in different jurisdictions than the participants work or receive their training. They also noted that the denominator is too broad or difficult to measure and that workforce programs are not funded to serve the entire population, arguing against its use as a metric.

Discussion:

The measure as defined is ambiguous, though these problems could be cleared up with better definitions. As a more population- or systems-oriented measure for addressing broader accountability purposes, it has considerable appeal. However, the measure received only 1 vote.

MEASURE #22: Skill Gains (6 votes)

Number of successful completers/Total number of participants enrolled

Advantages:

The sole advantage cited was that policy makers are very interested in (skill?) retention rates.

Disadvantages:

States noted numerous disadvantages, including that definitions of key terms such as “completion” (as skill gain”) and the time period for “enrolled” were unclear, that retention is less important than skill gains, and that people still participating in the denominator but not the numerator. In addition, states pointed out that program cuts, which reduced new enrollments, would increase this rate in the short run and that a measure of skill gains not based on an exit cohort is biased by the length of the program, creating an uneven playing field.

Discussion

One state suggested considering the use of *all* exiters as the denominator for this measure. As with the preceding skill gain measure, if ambiguities in its definition could be clarified, this has considerable appeal as a measure more geared towards measuring systems performance and addressing broader accountability issues. However, with only 6 votes, it only received limited support.

Return on Investment**MEASURE #12: ROI** (12 votes)

Tax revenue gains from extra wages/Cost of the program

Advantages:

States cited the fact that this ROI measure attempts to measure the rate of return to the *taxpayers*, something that is very salient to policy makers and the public.

Disadvantages:

The two disadvantages noted were that data are not readily available to measure it and that defining “extra wages” would be difficult.

Discussion:

The intent of this measure was to capture the *net impact* of workforce participation on tax revenue gains, which will vary from state to state depending on a state’s tax structure. There are readily available statistical models (from state comptrollers, business research bureaus, etc.) for estimating taxes paid based on earnings, location and family size. Figuring out “extra wages,” that is the amount of increased earnings attributable to participation, is the harder part. ROI is one of the more popular measures with elected officials and the wider public, in that it tells them what they are getting back for their investment. While the social rate of return is more relevant from an economic perspective, taxpayer return is the one that policymakers and the public focus on. This measure only factors in part of the returns to taxpayers, without also including reductions in public assistance or UI payments. As such it is only a partial taxpayer ROI measure. More research is needed to determine the best methodology to use for gauging taxpayer ROI.⁷ This measure received moderate support with 12 votes.

⁷ For more on measuring ROI for workforce services, see Christopher T. King and Dan O’Shea, *Estimating Return-on-Investment (ROI) for Texas Workforce Development Boards: Lessons Learned & Next Steps*, Austin: Ray

MEASURE #6: Cost Savings (25 votes)

Net income taxes and public costs avoided/Total cost

Advantages:

States pointed to two advantages, that this attempts to measure the rate of return to taxpayers, something that is very salient to policy makers and the public and that it is a measure that is easily communicable.

Disadvantages:

In addition to the name of the measure being somewhat misleading, disadvantages included that this does not measure the earnings gain for participants, and that it is difficult to measure costs fully, accurately, and consistently across programs as well as to quantify “avoided costs.” They also noted that results will differ based on differences in state tax law (e.g., not all states have income taxes) and that data are not readily available, especially from the IRS.

Discussion:

This might be thought of as a “back-door” ROI measure, one focused largely on taxes and cost avoidance. Definitions need to be clarified and made more explicit, and the same issues that affect the preceding measure are relevant here, especially estimating net impacts on participant earnings. It appears to capture both tax increments and reductions in public assistance and UI payments (as avoided costs) that the preceding measure ignored. States indicated that this was an interesting idea to pursue but needs more development. This measure received very strong support from the states with fully 25 votes.

MEASURE #28: Cost Savings to Employers – survey information (1 vote, 1 veto)

No definition was provided for this measure.

Advantages:

States cited no advantages for this measure, probably due to the fact that no definition was provided.

Disadvantages:

The disadvantages cited included that it was unlikely that consistent, objective data could be obtained, that it would be costly to measure, and that it was subjective and not defined.

Discussion:

No further discussion is offered for this measure, except to note that 1) it would require an additional survey of employers, which would be expensive and unpopular with a key workforce customer, and 2) the measure received only 1 vote but was vetoed by another state. It is probably not worth pursuing this measure further.

MEASURE #29: ROI (1 vote, 1 veto)

Wages of cohort minus wages of control group/Cost of the program

Advantages:

States cited the advantage that this is the only measure that really gets at the heart of the desired outcomes: that the program(s) make a difference in earnings above and beyond the cost of the program. In addition, ROI information helps make the case for program funding.

Disadvantages:

Disadvantages noted were that ROI can be difficult to measure, that many accounting systems do not capture program costs in a way that makes it easy to measure the denominator correctly, and that control or comparison groups are typically difficult to create.

Discussion:

Program costs typically occur in a relatively short time period, while benefits accrue over long periods of time. One must either wait a long time for results or estimate long-term wage results based on a short period of observed wage differences. Past cost benefit studies have produced a variety of estimates, some of them contradictory. Assumptions about the longevity (or persistence) of wage impacts produce most of the measured ROI. There are many things that we do not know, such as whether ROI varies depending on economic conditions (recession versus boom) and how this should affect our investment decisions. This measure would be expensive to calculate and difficult to calculate frequently. It is worth noting that, as with the previous one, this measure drew both 1 vote and 1 veto.

Other Measures

MEASURE #17: Participation Rates, 18-60 (9 votes, 1 veto)

Number of students in postsecondary/Number of total population who could be involved in postsecondary

Advantages:

States pointed out that this participation measure forecasts skill gains and could be a good youth measure. It could also be useful for continuous improvement.

Disadvantages:

Disadvantages noted were that the measure is not defined clearly and that data may not be readily available to measure participation due to FERPA, and, if available, would be expensive and labor intensive to collect. The measure also assumes post-secondary is the best option for all participants.

Discussion:

States suggested an alternative measure, namely the number of older youth in post-secondary/Number of older youth served. This is another measure that has appeal as a population- or system-oriented measure. It would be useful for addressing broader accountability concerns. Somewhat similar measures have been adopted by states for use with

their emerging “scorecard” approaches. This measure received limited support (9 votes) but also drew 1 veto.

MEASURE #30: Percent of Eligible Enrolled in the Program (0 votes)

Definition:

Number enrolled in program X/Number eligible for program X

Advantages:

States cited the advantage that this provides one measure of the percent of needs met by workforce programs in the state/community.

Disadvantages:

The only disadvantage cited was that it is very difficult to measure the number of people eligible for some (many) programs. To the extent that the eligible population must be defined using census-type data, it will be difficult to measure appropriate denominators for states and sub-areas as frequently as desired.

Discussion:

This measure would have appeal as a population- or systems-oriented measure but did not receive a single vote of support.

OBSERVATIONS, GAPS AND ISSUES

Observations and Gaps

The list of proposed measures that evolved from the second IPI meeting in Miami appears to address most of the policy information needs identified in the earlier meeting in New Orleans, though there are some noteworthy areas where more attention could be focused. The three core areas for adult and youth workforce programs are addressed—employment, retention, and earnings—and measures for skills and skill gains are incorporated as well. Employer measures also received some attention, as did measures for special population groups (e.g., public assistance) and return on investment.

These measures essentially can be sorted into three broad groups within the categories of need identified by policymakers: those with substantial support (21 or more votes), those with moderate support (11 to 19 votes), and those with low or no support (fewer than 11 votes). The last group can and probably should be dropped or at least consigned to a “further research and development” group. The state IPI teams should probably focus their efforts on refining the remaining two groups of measures, especially those that received substantial support from the states.

This observation is reinforced by our recent examination of experiences with developing non-federal measurement and management approaches in leading-edge states, which was conducted

for NGA and the U.S. Department of Labor.⁸ States were generally moving in the direction of 3-tiered performance measurement systems for workforce development, with common or system measures in the highest tier, program measures in the middle, and day-to-day management indicators in the bottom tier. Developing the upper tier measures required the greatest amount of effort collectively and had the longest lead-time for design and implementation. The IPI states appear to be at a stage where they can agree on basic outcomes for employers and program participants but are not quite on the same page in terms of defining social welfare, skill gains, and customer satisfaction. Maybe this work should be envisioned as part of stage two. Stage three would include working on the most complex and challenging measures, e.g., ROI, and probably should not become a priority until the previous stages are well established and functioning as designed.

Thus, the following measures appear to be the top candidates for adoption as part of the IPI Project:

Outcomes for Employers and the Economy

Measure #4: Market Penetration/Employer Penetration (30 votes)

Labor Market Outcomes for Program Participants

Measure #1: Employment (33 votes)

Measure #2: Retention with Same Employer (33 votes)

Measure #5: Earnings Gain (26 votes)

Social Welfare Outcomes

Measure #3: Decrease in Public Assistance (32 votes)

Return on Investment

Measure #6: Cost Savings (25 votes)

Customer Satisfaction, Skill Gains, & Other

N/A

These measures require further refinement and clarification in all cases, addressing some of the disadvantages cited by states as well as responding to some of the problems raised in the subsequent discussions.

States also suggested several additional measures for consideration, as follows:

- ❑ *Workforce Services Duration*, defined as the median number of days to exit all workforce services.
- ❑ *A Workforce System Completion Rate*, defined as the total number of completers divided by all exiters.

⁸ See Dan O'Shea, Sarah E. Looney, and Christopher T. King, *Non-federal Workforce System Performance Measures in the States: Overview*, Austin: Ray Marshall Center, LBJ School of Public Affairs, The University of Texas at Austin, December, 2003.

- *Continuing Education.* Louisiana considers continuing education to be a positive outcome for its workforce development system and measures those “continuing education and employed,” those “continuing education not employed,” and those “continuing education or employed.” They view the last is the most relevant measure.

Issues

Important policy and technical issues remain to be fully addressed, and a number of them have been around for some time. These include:

Time Frames. Time frames – including both the time period an indicator measures and the point in time at which measurement occurs – can dramatically affect measured outcomes. For example, measuring for pre/post wage gains by comparing the quarter immediately preceding program entry to the quarter immediately following program completion is likely to demonstrate a greater wage gain than comparing three quarters prior to program entry to the quarter immediately following program completion or exit. Since many participants enroll when they lose their jobs, it is highly likely that a participant’s wages during the quarter immediately preceding program entry will be uncharacteristically low (see Table 2 for an illustration). Several states expressed a desire to further explore issues related to time frames in order to identify the best strategies for addressing this issue.

Table 2		
Wages Q-2	Wages Q-1	Wages Q+1
\$4,895	\$3,120	\$4,912
Wage Gain:		
Formula A: $\$4,912 - \$3,120 = \$1,792$		
Formula B: $\$4,912 - \$4,895 = \$17$		
Difference: $\$1,775$		

A related issue is the value of short-term v. long-term measures. Research indicates that the value of participation in many workforce programs cannot be fully captured for several years. Yet, at the same time, policymakers demand near-term feedback for accountability purposes and program administrators desire short-term feedback to monitor the performance of their programs. An emerging practice in some leading states is the use of a tiered system, which measures outcomes at different times for different audiences.

Data and Measurement Challenges. There are continuing data and measurement challenges that involve key elements and aspects ranging literacy and numeracy to employment and earnings. First, we have not yet reached anything approaching a consensus on how best to measure skills or skill gains. This must be part of a continuing policy research agenda. Second, given the number of labor markets straddling state borders — especially along the Atlantic Coast, but elsewhere as well — and given increasingly dynamic labor markets and reliance on temporary employees, enhancing our capacity to measure employment and earnings reliably and consistently must be a top priority. This means at a minimum strengthening the Wage Records Interchange System (WRIS) and expanding participation in it to capture work and earnings across state lines. Third, there continue to be varying cost bases and accounting systems used by

the relevant federal and state workforce development programs that make up the “system.” These differences must be addressed if we are to seriously entertain designing and implementing anything beyond ad hoc ROI measures specific to particular states and local areas.

Relation to OMB Common Measures. The OMB common measures initiative is likely to be an issue in the immediate future as well. DOL has embraced the common measures at least in part, even if other agencies and programs have not done so as enthusiastically or wholeheartedly. The four common measures that apply to participants in adult programs are:

Entered Employment – Percentage employed in the 1st quarter after program exit.

Retention – Percentage of those employed in the 1st quarter after program exit that were still employed in the 2nd and 3rd quarters after program exit.

Earnings Increase – Percentage change in earnings: (i) pre-registration to post-program and (ii) 1st quarter after exit to 3rd quarter after exit.

Efficiency – Annual cost per participant.

Whatever results from the IPI Project will need to be cognizant of the OMB effort as it continues, whether or not WIA, Perkins and other reauthorizations are successful in formally adopting these measures. DOL intends to proceed by regulation if not legislation.

Integrated Information Systems. Finally, there are outstanding issues concerning effective and efficient ways to design and institute integrated information systems to support such performance measures that are not addressed in this paper. These must be articulated clearly and addressed in the future at the federal, state and local levels.

NEXT STEPS

The next steps remaining for state policy teams include the following:

- ❑ Reach consensus on a small handful of core measures that encompass the policy needs groupings identified in New Orleans, adopting the measures with substantial support for employment, earnings and employers for immediate use and identifying others for further research and development.
- ❑ Refine the definitions and preferred data sources for the core system measures.