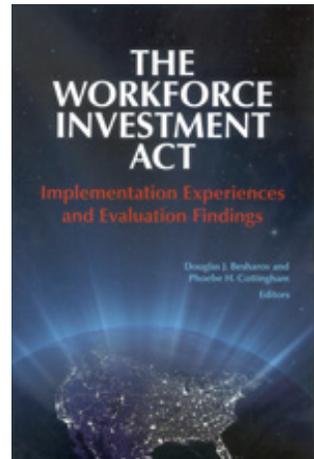

Upjohn Institute Press

Introduction

Phoebe H. Cottingham
U.S. Department of Education

Douglas J. Besharov
University of Maryland



Chapter 1 (pp. 1-46) in:

The Workforce Investment Act: Implementation Experiences and Evaluation Findings

Douglas J. Besharov and Phoebe H. Cottingham, eds.

Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2011

DOI: 10.17848/9780880994026.ch1

The Workforce Investment Act

Implementation Experiences and Evaluation Findings

Douglas J. Besharov
Phoebe H. Cottingham
Editors

2011

W.E. Upjohn Institute for Employment Research
Kalamazoo, Michigan

Library of Congress Cataloging-in-Publication Data

The Workforce Investment Act : implementation experiences and evaluation findings / Douglas J. Besharov, Phoebe H. Cottingham, editors.

p. cm.

Includes bibliographical references and index.

ISBN-13: 978-0-88099-370-8 (pbk. : alk. paper)

ISBN-10: 0-88099-370-7 (pbk. : alk. paper)

ISBN-13: 978-0-88099-371-5 (hardcover : alk. paper)

ISBN-10: 0-88099-371-5 (hardcover : alk. paper)

1. Occupational training—Government policy—United States. 2. Occupational training—Government policy—United States—Evaluation. 3. Occupational training—Law and legislation—United States. 4. Employees—Training of—Law and legislation—United States. 5. Vocational guidance—Law and legislation—United States. 6. United States. Workforce Investment Act of 1998. I. Besharov, Douglas J. II. Cottingham, Phoebe H.

HD5715.2.W666 2011

370.1130973—dc22

2011010981

© 2011

W.E. Upjohn Institute for Employment Research
300 S. Westnedge Avenue
Kalamazoo, Michigan 49007-4686

The facts presented in this study and the observations and viewpoints expressed are the sole responsibility of the author. They do not necessarily represent positions of the W.E. Upjohn Institute for Employment Research.

Cover design by Alcorn Publication Design.

Index prepared by Diane Worden.

Printed in the United States of America.

Printed on recycled paper.

1

Introduction

Phoebe H. Cottingham
U.S. Department of Education (retired)

Douglas J. Besharov
University of Maryland School of Public Policy

Over a decade ago, Congress initiated a major shift in federal workforce policy through the Workforce Investment Act (WIA) of 1998. WIA aimed to consolidate and modernize disparate workforce programs, and to assure that job seekers and employers benefited from a more open and effective utilization of federal funds. No single study has examined all aspects of the act. There are many studies of its program features, as well as efforts to estimate the economic outcomes for those receiving WIA services.

This volume examines WIA's objectives and the evidence on program performance and impact. The chapters originally were commissioned for a meeting held with staff of the European Commission for a discussion of WIA lessons and the implications for future workforce programming in the United States as well as Europe.¹ The chapters are organized into five general areas:

- 1) understanding WIA,
- 2) program implementation,
- 3) performance management,
- 4) impact evaluations, and
- 5) future evaluation choices.

The 2009 congressional appropriation for WIA was over \$15.9 billion, including \$3.3 billion for three WIA employment and training programs that replaced prior Job Training Partnership Act (JTPA) programs.² WIA funds are allocated to states based on five-year plans. States are responsible for using the funds for services operated under

local entities. The European Social Fund of the European Commission (EC) allocates approximately \$70 billion across seven-year cycles (roughly equivalent to the annual WIA appropriation) to EC member states for workforce services, of which occupational training is a major component.

The chapters in this volume focus exclusively on the U.S. experience, framed to help the European Commission staff in its deliberations on workforce programming to understand how the WIA performance management systems function, as well as the role of evaluations assessing workforce programs.

UNDERSTANDING WIA

WIA's main purpose, as set forth in the 1998 legislation, was broad: "*to consolidate, coordinate, and improve employment, training, literacy, and vocational rehabilitation programs in the United States.*" Indeed, WIA introduced extensive changes in the nation's publicly funded workforce programming. As summarized by Dianne Blank, Laura Heald, and Cynthia Fagnoni in Chapter 2, "An Overview of WIA," public workforce programs had become "fragmented"—an "uncoordinated patchwork of programs and agencies" suffering from "inefficiency, duplication of effort, and confusion for the job seeker."

The solution in the 1998 WIA legislation, Blank, Heald, and Fagnoni note, was to decrease the previous focus on income eligibility as the only basis for accessing services (as well as the focus on job training as the primary means for getting a job) and increase the focus on assessment and marketing existing skills. These changes placed more emphasis on personal responsibility, self-service, and consumer awareness in choosing options. The consolidation of services was to take place locally, through a new system of WIA One-Stop centers, guided by state and local entities to assure service coordination and customer access as required by WIA.

Whether WIA indeed produced the efficient streamlining of funds into open access systems as intended is a question that Blank, Heald, and Fagnoni believe is not fully answered. There are 25 reports by the U.S. Government Accountability Office (GAO) over the decade responding

to questions raised by members of Congress about WIA (see Chapter 2). (*Editor's note: The Government Accountability Office changed its name from the General Accounting Office in 2004. For readers' ease and consistency, we use the current name in the text and references throughout the book.*)

By being made available at the local level through one entry point—the One-Stop centers—the 16 different federal programs (see Table 2.1) would no longer require potential applicants to go to different offices to apply for services. For fiscal year 2009, Congress appropriated over \$15.9 billion for the 16 mandatory programs, including \$3.3 billion for three new WIA programs (Adults, Dislocated Workers, and Youth) replacing prior JTPA programs.³ The federal program offices were expected to work with the U.S. Department of Labor (USDOL), as the department was given overall responsibility for administering the provisions of WIA.⁴

WIA also initiated major changes in how funding for training services is distributed, by mandating that training funds be sent through individual training accounts (ITAs) to the training providers chosen by the WIA participants eligible for training services. Under JTPA and the Comprehensive Employment and Training Act (CETA), training funds went directly from the governmental entity at the state or local level for training service contracts with providers who were also responsible for recruiting trainees, typically from other local agencies. A second change under WIA required a process for establishing eligibility of WIA training providers. The training providers who are to receive ITA funds for training services to WIA participants need an established track record of positive outcomes that meet or exceed each state's performance criteria. WIA required states to establish eligible training provider lists (ETPLs) of providers and approved training course offerings that have met and continue to meet the state's performance criteria.

Blank, Heald, and Fagnoni point out that in 2001, the GAO found the requirements on training providers to be overly burdensome because so few people were referred under WIA. Subsequently, the USDOL began to provide waivers of the ETPL requirements, and 40 states obtained such waivers (see Chapter 6 for more on ETPLs).

In terms of governance, WIA required state governors to set up a state Workforce Investment Board (WIB) to oversee WIA implementation at state and local levels, with local WIBs organized to oversee

the One-Stop center operations. Governors decide how many members will serve on the WIBs, and they are required to assure that a private sector representative is named to chair each board, and that those representatives make up the majority of board members. Blank, Heald, and Fagnoni report that WIBs average 40–60 members, and that in 2007 there were 1,850 One-Stop centers, under the jurisdiction of a WIB (a regional WIB can be given several centers in its jurisdiction).

A key question in many GAO reports on WIA is whether the new performance management was being established. GAO reports repeatedly urge the USDOL to move more quickly to establish the requirements in clear, unambiguous terms. In particular, WIA mandated that the USDOL establish performance measures on five outcomes to be used by all states, and that the USDOL negotiate with each state on their minimal performance levels for each measure. (States may add measures or set higher levels for particular jurisdictions.) Central in the WIA performance system is the congressional requirement that states use Unemployment Insurance (UI) records for three of the five WIA performance measures—job placement, retention, earnings—with other sources for measuring skill attainment and customer satisfaction.⁵

The early years of WIA showed how challenging it was for states to develop new performance systems based on the UI records. Blank, Heald, and Fagnoni note that overall, the use of UI records for timely analysis and reporting at the operations level in WIA has proved less successful than hoped for time-sensitive management functions. In response, the USDOL has allowed states to use “supplemental data” to fill gaps in the UI wage records and collect job placement outcome information from sources other than UI records, or the “supplemental data sources.” In 2004, over 75 percent of local areas reported that “they directly follow up with participants after they leave the program . . . to help fill gaps until the data are available from the UI wage records.” In some cases the supplemental data are viewed as interim indicators to manage WIA programs or predict WIA performance outcomes. Blank, Heald, and Fagnoni believe the GAO recommendation to allow continued use of supplemental data is sensible.

Ultimately, WIA did advance the linking of WIA and UI record systems to record the placement and earnings results, but states, USDOL, and researchers use these files more for monitoring overall progress. The USDOL operates a nationwide, computerized WIA Standardized

Record Data (WIASRD) system for states to input data on aggregated counts or averages, based on what states collect from the One-Stop centers, providers, and employers.

The central and more difficult part of the performance management challenge emanating from WIA was the required establishment of definitions of participant status to be used across the state-run WIA system to measure performance at the local and state levels. In 2005, the U.S. Office of Management and Budget called for common measures—the harmonization of performance measures across a larger swath of federal workforce and training assistance programs managed not only by the USDOL but other federal agencies, according to Blank, Heald, and Fagnoni. In general, progress was slow, although USDOL eventually made advances to improve the accuracy of performance data and settle performance measures, beginning in 2005.⁶

The WIA performance management system was further complicated by Congress asking that the performance goals be set through negotiations between individual states and the USDOL. Blank, Heald, and Fagnoni summarize the criticism of negotiation without a standardized and uniform procedure for establishing what are reasonable performance goals. Also, many expressed concerns that without adjustment procedures, the system discourages One-Stop centers from providing services to those who appear less likely to get and keep a job. Most recently the USDOL has used a regression model to set national performance goals, based on data on job seekers in local labor markets, using the WIA database and other data.⁷ Another concern the authors describe is that only a small proportion of job seekers who receive services at One-Stops are actually reflected in WIA outcome data. In the 2004 GAO study, only about 5 percent of job seekers who walked into a One-Stop center were registered for WIA and tracked for outcomes. The self-service customers, those seeking information on their own, are actually the largest group served under WIA. Blank, Heald, and Fagnoni point out that the GAO has recommended that the USDOL consider ways for states to track all job seekers coming into One-Stop services, but this presents problems when self-served customer results are combined with other WIA customers who obtain more intensive services, especially training.

Blank, Heald, and Fagnoni conclude that there still is not a uniform national practice for tracking registrants in WIA, undermining accuracy

of performance data and the ability to compare states equitably. They also note the inherent tensions between local entities, state entities, and the federal government.

PROGRAM IMPLEMENTATION

Many of the WIA issues noted by Blank, Heald, and Fagnoni are explored in depth in other contributions to this volume. All draw on federal studies of WIA's implementation, especially during the first half of the 2000–2009 decade, when attention focused heavily on the new role of One-Stop centers, the performance management system, and what features appeared to be more successfully implemented than others.

In Chapter 3, “The Use of Market Mechanisms,” Christopher T. King and Burt S. Barnow summarize the early implementation challenges in eight states in 2003–2005. They draw conclusions similar to those of Blank, Heald, and Fagnoni concerning the central importance of a potentially stronger performance management system in WIA than existed under JTPA. Under JTPA, the federal government did not attempt to set performance standards at all, leaving it to local service delivery areas (SDAs) to set their own performance standards with approval by the lead state office. Under WIA, the top level of the performance management structure is now at the federal–state level, after the startup phase when states had less performance system direction from the federal level.

King and Barnow conclude that there was considerable tension associated with the implementation of the new WIA performance management system. Having local and state layers of government adjust to operating with a set of standards derived from a higher level brought resistance from some at the local level. The focus was on state WIA leaders first “negotiating” standards with federal officials and then translating them to local areas, often with differing rules according to state policy. The principal complaint about this approach centers on states having to apply the negotiated state performance standards for all One-Stop centers within their state. King and Barnow find that in the eight states studied, there was considerable variation across the states in their performance standards under WIA. Both state and local staff

disliked WIA performance measures and standards, believing JTPA performance management worked better. According to King and Barnow, performance standards are seen as arbitrary numeric goals, with no allowance or adjustments for serving more distressed areas (as many claimed the JTPA adjustment allowances provided).

Another concern arising during the transition from JTPA to WIA was who collected the data to measure program outcomes and how accessible it was for program operators. JTPA performance measures of program outcomes relied on job placement reporting by providers who would conduct their own follow-up contacts with program participants. WIA deliberately sought to shift the reporting responsibility to state entities, requiring the utilization of Unemployment Insurance administrative records to document who was employed. The plan was that state entities first collect participant information from One-Stop center providers and then match the individual participant records with reports submitted by employers to the state Unemployment Insurance office.

King and Barnow describe the extent to which this intended change in who collected outcome data brought extensive deliberations and contentions over how program entry and exit status would be defined in the WIA performance management systems. One-Stop center operators understood that who counts as a WIA participant would form the base count for establishing their performance record for judging how many of the WIA participants succeeded in finding jobs. Naturally, program operators want to include as WIA participants those most likely to succeed in finding jobs, and exclude those who are likely to fail in achieving the outcome. After considerable negotiation at the federal level, it was finally agreed to exclude job seekers who do not utilize the core WIA services, focusing on those using the second and third service tiers, called “intensive” and “training.” This shifted attention to defining what were core services. Similar disputes arose over defining when a WIA participant has exited WIA, and thus is countable as a successful job placement or not.

In response, the USDOL established a reporting system to be used by states and One-Stop centers, containing a standardized set of definitions, extensive documentation, and technical explanations. More recently, the USDOL commissioned work on how to introduce adjustments to performance standards to take account of state and local economic conditions and job seeker characteristics (see Chapter 9).

Another key change in workforce programming under WIA was the requirement that there be universal access at entry to the One-Stop centers, meaning no barriers on entry to core services of the job search and information assistance. Users of One-Stop centers do not have to pass eligibility requirements that in the past limited core services to low-income persons or the long-term unemployed. This was a major change from most of the JTPA job training programs that had requirements to keep training just for the economically disadvantaged or long-term unemployed.⁸

Some believe that disadvantaged populations have had less access to job training under WIA than under JTPA because of the open access and no low-income eligibility requirement. In their state case studies conducted during the early phase of WIA, King and Barnow find tension over spreading WIA funds in a way that may not be for those who would gain the most, the disadvantaged. Studies looking at very large samples of WIA participants in training do not necessarily support the presumption that disadvantaged populations have lost out on access to training through WIA due to open access policies. (See Chapter 13 for a summary of an extensive study of WIA participants.) It appears that disadvantaged populations are the beneficiaries of WIA adult training and the youth programs. While the total number of disadvantaged people in WIA training may be less than was true under JTPA, the most recent study, across 12 states, finds that adults in WIA training are, on average, disadvantaged in prior earnings, employment, and education. Dislocated workers receiving WIA services, as was true during the JTPA decade, overall have less disadvantaged backgrounds, and reflect the general population in terms of education, work experience, and prior earnings.

Another major change under WIA is the funneling of job training funds through ITAs, essentially vouchers tied to the job seeker, not job training providers. States and localities have some flexibility in setting the dollar value of ITAs, and within a One-Stop center can tailor the value to fit with individual customer choices. The main purpose of ITAs is to provide job seekers who need and/or want training with many choices among eligible providers and training courses, rather than be limited to just a few training options or slots determined by contracts with a few providers to provide training for groups of job seekers.

King and Barnow find that ITAs have been well received and appear to be working because there is flexibility in ITA values, as determined by states and localities. They report that in most centers, policies allowed ITAs to be adjusted by center staff based on the job seeker's needs.

During the JTPA and CETA program years, many focused on the seemingly ineffective job preparation programs for the disadvantaged. In Chapter 4, "Customized Training," David A. Long explains why traditional job training was unpopular with many employers, noting the trade-off that companies see between retraining their own employees for new skills needed in the business versus finding new employees trained by others, especially training paid for by government as a way to reduce unemployment or help the disadvantaged. Long explores why customized training may be more effective than the training focused exclusively on raising general skills of the unemployed or disadvantaged, typically for jobs at the lower end of the job market. As the economy changed, traditional job training became disconnected from job-specific skills needed by employers in high-growth, new markets. Long defines customized training as "the provision of particular employee skills needed by specific firms in their current and new workers."

Customized training is done outside the firm, by intermediaries who take on the role of recruiting and screening applicants for customized training when partner employers are looking to hire new skilled workers, which Long notes also allows programs to give priority to low-income and disadvantaged groups. He believes customized training should also be distinguished from "off the shelf" training provided by vocational education. Customized training is responsive to the needs of specific local employers in filling particular skilled work positions, with a commitment by the employer to employ some or all successful completers of the training (or continue employing incumbent workers) and share the costs of the training.

Long also examines research on incumbent worker training, that is, in-house training of employees by employers. Several nonexperimental studies used large 1990s data sets to look for a difference between the productivity (wage growth, performance ratings, and career advancement) of employees who reported receiving in-house training provided by employers and the productivity of employees who did not have the training. The three studies produced varying estimates of the average

rate of return (from 17 to 50 percent) to the firms from in-house training, but it is not clear how well these studies controlled for selection biases.

Long recalls early efforts under WIA and by private foundations to fund demonstration projects or partnerships that engaged business and training entities to align WIA-sponsored job training with private sector employers, with a focus on sector training. He summarizes why the idea of customized training is even more popular under WIA, and reports on a recent study that rigorously examines customized training impacts in three sites in a foundation-supported project. He says that participants earned 18 percent (about \$4,500) more than controls during two years from baseline during the 2004–2008 period (Maguire et al. 2009). The three sites were located in urban areas with trainees who appeared little different from WIA training program entrants and therefore may be relevant for WIA efforts to develop customized training.

Studies limited to a few sites where a particularly impressive program has been established, while useful to learning if something quite unusual actually is making the differences claimed, need further testing through replication. This is where the news can go sour, as rigorous replication studies often fail to find the same effects as in the original program site where strong vision and commitment may produce a charged-up staff with specific synergy with customers and businesses. Some believe original sites may have built fortuitous partnerships with particular employers that are most difficult to replicate in new sites. Long recalls how the USDOL replication study of the San Jose CET (customized) model training program, which was very successful in the 1990s, could not find similar net impacts. Where moderate impacts appeared, they were in replication sites that seemed to have more “faithfully implemented” the original training program. This indicates how difficult it is to turn highly promising, even effective training programs into large-scale franchises or regular program practice. Similar problems are present in education, where rarely if ever have successful particular program models been “scaled up” into major service systems that reproduce effects anywhere near those found in an original small-scale study. Moreover, the costs are great to introduce change by extensive scaling up, as opposed to testing changes in practice within the existing system where the change may be more doable, or more gradual expansion of a program started in one site.

Looking to the future of customized training, Long notes that there are several issues. One is the matter of curriculum, and how to know that customized training meets the needs of both employer and potential worker. He cautions that some customized training initiatives tend to build from past training curriculums, which presents problems if new business sectors are being targeted or when new skills are needed. Another issue is the effort that training providers or intermediaries have to put forth to prescreen potential trainees to make good career matches, as well as provide ongoing support.

Input from the private sector and active engagement is an important WIA objective. The WIBs are considered the key entity that brings in private sector perspectives to One-Stop centers. There are also state WIBs, whose key responsibility is advising on the state's performance standards and the policies governing eligible provider lists. The WIBs are supposed to bring input in from employers, business groups, and other stakeholders on policies and operational plans for local One-Stop centers as well as the state. Reports are spotty and generally are not very convincing that WIBs have brought the strong private sector engagement intended. Some conclude that the WIBs have too many stakeholders, and that many of them are conflicted or compromised by WIB member associations with WIA, either as major training providers or contractors to WIA.

David Heaney considers the role of the private sector in managing One-Stop centers in Chapter 5, "One-Stop Management and the Private Sector." He notes that WIA intended private sector engagement to be welcomed in operating the centers, not just on the WIBs. WIA placed "a high premium on employer-driven strategies and integrated service delivery through colocating key providers under one roof . . . to effectively leverage the strengths of [a] diverse set of partner organizations operating side by side." Business and employers should determine the content of programs for preparing the workforce.

Heaney offers a critical perspective, however, on whether the private sector is engaged in WIA. He believes that active participation of the private sector has been stymied in One-Stop center operations. He argues that the history so far has been progressively less employer-business input as WIA implementation progressed. Heaney notes that at first those serving as One-Stop center operators included a healthy distribution across private for-profit, nonprofit, and public sector man-

agers. Over time, the procurement process appears to have narrowed the distribution, with far fewer for-profits or business entities. In fact, fewer entities now compete for the One-Stop center operating contracts.

Without more effective private employer input, Heaney is concerned that job seeker selections made through customer choice may be out of sync with the realities of the existing market, or a true employer-driven service delivery system. Training providers do not necessarily adapt effectively to market realities. Heaney urges consideration of policies that would attract a greater number and more diverse set of qualified bidders from all sectors for One-Stop center management, including allowing risk/reward tolerance levels in the pricing and design of contracts. He believes operators should have flexibility to refine and change practices much as takes place in typical company staffing operations, rather than be restricted to overly prescriptive practices that have unknown effectiveness (for example, caps on administration costs and profits discourage private sector involvement in WIA).

Heaney agrees with the widespread criticism of WIB representation requirements that seem to burden these important advisory groups with too many interested parties and decrease WIB effectiveness. He sees a parallel burden or inefficiency in One-Stop centers because center staff attempt to handle too many interested agencies and customer calls. He favors giving center managers more authority to make decisions on performance, quality, and corrective actions. He also urges that more evaluation should be done, however, before attempting to change the performance management system.

The overarching watchword of the WIA system is accountability: accountability for results but also assuring that data is fully used so those involved in the WIA system know what is going on, and how the various responsible parties are performing. In addition to the new performance management system established at the federal-state level, information that helps the clients and staff in the system make wise choices is essential. WIA managers also are accountable for how the public funds are spent on training opportunities, to assure that the most effective training opportunities are identified and funding goes to the effective trainers. This requires management review of performance records of training providers, and designation, based on performance standards, of who is eligible to receive WIA training funds through the ITAs. Public access to the performance records is an essential part of an

accountability system to assure that providers of training services make available performance records so customers can see the potential payoff in a job and earnings by previous trainees.

In Chapter 6, “Eligible Training Provider Lists and Consumer Report Cards,” Carl E. Van Horn and Aaron Fichtner report on their study of four states’ progress in developing performance records and publicly available information. They find evidence that the new accountability system requirements under WIA for the provision of training services have been implemented to some degree, thereby demonstrating that accountability systems that meet the 1998 WIA vision are possible. Therefore, they point out, what is most important is not the more limited training that some find being funded under WIA—it is the WIA requirement that states set up new workforce systems for deciding who needs, receives, and provides training. Those eligible for training are required to have the opportunity to review and select from lists of training courses rather than be assigned to a training course by program administrators. The lists of training courses and providers appear on the state lists of eligible training providers—those who have verifiable records of results based on previous trainees that have been deemed by states as meeting state performance standards. The information is translated into state consumer report systems to disseminate the performance training outcomes for each provider and program, so that ITA holders and others can view the training options meeting the standards.

As Van Horn and Fichtner point out, under JTPA, training services were typically procured directly by local government agencies that selected the occupational concentrations and the service providers. Basically, an annual plan would select providers who would offer what was thought to be suitable occupational training for local populations. The government administrators would procure set numbers of training slots to be filled during the year, and then use up the slots by various entity referrals. This system could not assure that those who would benefit most got training, or that the training provided was necessarily effective in helping the job seeker secure a new job. It was convenient for government administrators, but likely led to waste or inappropriate or ineffective training purchases. Job seekers were simply placed in particular training slots because the slots were already committed under the contract. Some believed that better outcomes were possible if job seekers were provided much better information on the labor market in

general, on the occupations (and their requirements) most likely to be in demand, and to have some understanding of the employment results that others had achieved by taking a particular provider and occupational training track.

Van Horn and Richtner conclude that the UI records as mandated by WIA are being used to varying degrees to build statewide listings of eligible training providers and customer report cards that give summaries of provider performance records. The delayed availability of UI records noted by others has not prevented the creation and updating of eligible training provider lists in some states.

Based on their research conducting interviews in four states during 2009, Van Horn and Richtner find progress in recent years. In the four states, there are provider lists and performance measures called for under WIA that have surprisingly deep repositories of in-depth information. In the four states examined—Florida, New Jersey, Texas, and Washington State—these efforts actually began before WIA and were funded by the USDOL as part of their pilot initiatives prior to WIA. It took years of work to reach the data accumulation now available. All states now have online performance reporting systems in use, which is an achievement. These efforts required pulling together data from the state entities managing WIA, UI, and education and training organizations, and calculating average performance levels. The information also encourages the training institutions to provide explicit details on the particular training or career certification courses available and the placement results obtained by their course takers.

For New Jersey, as an example, Van Horn and Fichtner report that the online eligible training provider list contains more than 600 education and training providers, offering more than 3,000 training programs. Performance data are not yet available on all courses, as only one-third of the provider files contain performance averages that are outputted to consumer report cards on every course and/or provider. The report cards have to be based on WIA-supported students, and for some courses there are too few WIA course takers to produce an average (in these cases the placement information is averaged across all courses sponsored by a training provider).

It is important to note that, in contrast to much earlier reports by others finding inability or ineffectual use of UI records by states, state capability has grown and the picture may be different. At least in the

four states reported on by Van Horn and Fichtner, and relying on their recent interviews, all now have state-run, fully developed ETPL systems with consumer report card systems, and all utilize UI records to calculate outcomes.

Van Horn and Fichtner also report that the ITA system has stimulated the large providers, such as postsecondary education institutions, to help WIA trainees secure other public funding sources for which they may be eligible, thereby adding to the ITA funding. It is quite likely that training choices are influenced not only by the ITA voucher value, but the potential to tap other sources of support for trainees that training institutions can provide.

PERFORMANCE MEASUREMENT

Accountability for using public funds to achieve the public objectives is widely expected. In workforce programming, being able to assure accountability by measuring performance appears to be a straightforward process, because the end result, a job and increased earnings, should be easily measured. Under WIA, establishing the benchmarks for measuring performance, with allowance for state negotiation to reflect economic conditions, brought two advances over the past: 1) a common set of definitions was established, permitting national assessment of overall achievements and comparisons of state and local assessments; and 2) full computerization of administrative records allowed more timely reporting of results.

In Chapter 7, “The Challenges of Measuring Performance,” William S. Borden recommends establishing standardized definitions before a program is initiated to assure comparability across geographical and function units. Borden looks at WIA performance measurement issues as a case example of the complexities inherent in creating, maintaining, and using performance systems for management. He speaks from his experience in helping government agencies design and implement performance management and data validation systems, not only for WIA but other programs in the USDOL and the Department of Health and Human Services. These performance systems are clearly necessary, says Borden, because “tracking and measuring customer flow, services,

and outcomes is inherently desirable and even necessary to managing any modern organization. Therefore, the question is not whether we should track customer flow and services and measure performance, but whether and how we should use the data to determine funding, incentives, and sanctions.”

Many mistakenly believe performance can be monitored through relatively simple systems that capture seemingly obvious goals and processes. Alas, as Borden notes, there are “challenges that are little known except to the state and federal staff managing the performance systems, and that are often not clearly understood. There is very little that is easy and straightforward about measuring program performances. Seemingly simple concepts . . . are actually very complex . . .” The many conceptual and operational issues raise significant questions about motivation, state–federal political power sharing, and the management of government programs.

Borden makes clear that he agrees with others who believe that program evaluation and performance management derive from different sources and motives and that keeping them as separate functions is warranted. WIA, he notes, “has shown that it is difficult to measure performance well, and that using inaccurate performance data to drive policy and incentives leads to misallocated resources.” Administrative data are needed to accomplish both functions: to understand and monitor program operations, and to carry rigorous evaluations using randomization of applicants.

Standardized definitions, according to Borden, are critical and must be established before developing system software and validation checks that provide information essential for program managers to keep on top of the complex systems. In short, definitions must be agreed to by those engaged in various levels of operations, enforceable, and support consistency checks so essential for building the performance system. Arriving at standardized definitions challenges programs with shared governance structures. Those working within the structures develop stakeholder interests, and are typically more concerned about meeting their goals than improving their results. Consequently, Borden notes, they tend to have “somewhat exaggerated reaction to the burdens imposed by performance systems.” He points out that in WIA, a diverse system, “forces of fragmentation and inconsistent data are so great that only a very strong and standardized performance management system

can overcome or at least neutralize” the stakeholder pressures. Despite the progress made in WIA in developing measurement definitions and performance systems, there remain the inherent tendencies by some program operators to resist any seemingly externally determined system that may challenge their perceptions of performance. Borden suggests a number of ways to get “buy in” from program operators to a system. It is very important, he says, to focus initially on building strong data capacity through effective performance management tools and methods rather than on the punitive aspects of performance management.

Borden sees wide variation among states, grantees, and local program operators in their level of sophistication and case management data they collect. Many, he says, collect far more detailed performance data than anything imposed by the USDOL. Federal efforts should focus on the key data validation component, to raise every state and grantee to a minimum acceptable level of data management and data reliability.

On the issue of the impact of performance management on customer selection, Borden finds a conundrum because barriers to success tend to be subjective and unreliable, and consequently very difficult to measure. He suggests that computing performance separately for different classes of customers based on barriers still provides the clearest information to program operators. If performance is adjusted after the fact using regression models, results should be similar. The problem with using negotiation under WIA to obtain flexibility and avoid the complexity of regression-based adjustments is the overall absence of systematic and consistent performance goals across states.

An effective management system does have costs to establish, but there are also considerable costs to allowing states to administer their own programs and make their own rules—no usable national data can come from this type of devolution. In the case of WIA, where Congress tried to confront an overall system that was highly fragmented and turn it into the One-Stop system, with seamless access for the customer, there are obvious challenges to building a successful reporting process. For example, there are still fragmented funding streams coming into the One-Stops, with requirements for data collection and reporting to many programs and agencies with varying and even conflicting definitions of customer characteristics. The challenge is to acknowledge that specialized programs may be more effective in serving difficult populations,

but there is still a need for reliable and consistent data across all the states and local areas to improve the entire performance process.

Even more important, if program operators see performance as a game, not a management tool, they are tempted to manipulate reporting their outcomes. Borden reviews the ways outcomes are distorted in WIA, such as who is actually enrolled and how services are defined, and of course manipulating exit dates. He concludes with this summary: “Do not attempt to measure something you cannot define or validate, and make sure the calculations are reliable and well tested.”

Borden’s main concern, echoed in other chapters as well, is that the accuracy of management system data is likely to be compromised if the performance data is used for funding decisions, and for assigning financial rewards or sanctions or incentives in general.⁹ Incentives encourage program staff to pick those considered most likely to succeed and recruit them for services (“creaming”), a selection bias at odds with the program goal to get the most net benefit for costs. Distorted data provide inaccurate counts. Borden believes most data systems simply do not have the accuracy required for discerning true performance differences. He recommends that performance management systems be kept out of incentive systems, and operate as they are intended: to assist managers in watching how their systems are operating.

A similar view to Borden’s is presented by Burt S. Barnow in Chapter 8, “Lessons from the WIA Performance Measures.” Barnow appraises the role of performance management measures as contrasted with measures from impact evaluations. He notes that evaluators see performance management as a kind of offshoot of their process or implementation studies. “Process studies document what happened while the program is implemented, impact evaluations assess what difference in outcome measures was due to the intervention, and cost-benefit analyses assess whether the benefits of a program exceed the costs.” Barnow believes the key differences between performance management and evaluation activities are “matters of depth of analysis and causality.” Performance management relies on “easy to collect data on inputs, activities, and outputs.” Functions important to impact evaluations are not included and would be too expensive and even irrelevant in any event within a performance measurement system. This includes evaluation functions such as tracking long-term outcomes, and of course establishing and tracking a control group created out of the applicant pool.

Hence, performance management systems typically utilize some judgment about what a program should or could accomplish in job placements and earnings of participants—the program performance standards. These judgments at the local level are inferred from the placement and/or earnings of the last cycle of participants, or averaged performance records for what seem to be similar programs and participants. These performance judgments can be reasonable or unreasonable, but they are not impact evaluations.

Another purpose of performance management systems is to establish a feedback process that gives signals back to those who are responsible—accountable—for obtaining results, and that applies rewards and sanctions on work units or individuals involved. Government performance management systems typically build on the bureaucratic system, or the bureaucratic process pipeline, as the production process in a program. Systems establish measures of what should be accomplished at various points in the pipeline, such as success rates in recruiting customers, proportion eligible for services who were offered them, response patterns of customers, retaining customers for the desired time period, and status at exit. Evaluations, however, view such pipeline checkpoints as implementation variables whose purpose is to describe what the program looks like. Evaluators direct their attention primarily on whether program completers or exiters succeeded in some externally determined outcome (e.g., a job) relative to what they might have accomplished without the program.

Barnow concludes there are three central issues in the debate over performance management and evaluation in the workforce area:

- 1) Does performance management influence, indeed negatively distort, the service system itself in ways not intended by the program designers? Barnow believes there is evidence of too much distortion, and it is negative rather than positive.
- 2) Does attaching incentive systems to performance systems bring the results intended? Barnow believes there should not be large rewards and/or sanctions, as there is not evidence that these have markedly changed management practices, consistent with conclusions in Wandner and Wiseman, as well as in Borden.
- 3) Can adjustments reduce distortions created by the performance management system? Barnow favors adjustments more for im-

parting “fairness” and psychologically reducing the tendencies to distort measures or game the system than for overall effectiveness, but cites no particular studies on the question.

Theories or explanations are plentiful regarding why one should expect that performance management systems can and do distort the behavior of agencies subject to the performance measures. Program operators, Barnow notes, respond to performance management systems by spending resources “trying to look good rather than doing good.” This includes modifying the timing of entry and/or exit, or “more pernicious effects, as when programs engage in ‘cream skimming’ and serve those less in need to receive better performance scores.”

Studies find that such service changes are identifiable as direct responses to the performance management system rather than responses to the clients. Barnow reviews the reasons why this happens, why managers display “selection biases” in choosing or helping the customers. He notes that studies comparing the characteristics of WIA enrollees versus JTPA enrollees find that WIA enrollees shifted upward toward individuals with few barriers to employment. WIA also had reduced levels of enrollment, and researchers concluded that selective registration was the reason.

Barnow further points out that studies comparing estimates of short- and long-term program impacts obtained in rigorous evaluations with the measured outcomes on the same program units as captured in the performance management system clearly show there is very little relationship between the two. Either the correlations are nonexistent or very weak. Thus, Barnow concludes that performance management system results are by design short term, and do not capture very strongly program impacts.

Barnow also concludes that performance management systems should not deploy large rewards and/or sanctions, as these efforts are very weakly related—if at all—to program impacts and encourage data distortions. Program management, in short, is not nearly as important as sound evaluation in guiding overall policy directions, and has limited support as an assurance that a program is achieving the central objectives of the policymakers.

Given that performance management systems risk distortions in who is served away from program purposes, does adjusting perfor-

mance standards reduce incentives to torque the program service and customer mix? Here there is another debate. Those who look at the wide differences between local populations and economic conditions confronting program managers argue that programs should not be penalized for performance outcomes in more difficult conditions. Thus, adjustments to performance standards are a reasonable approach to level the playing field. Those opposed to adjusting performance standards argue that setting lower expectations for some programs than others perpetuates inequities. Barnow approves of adjusting performance standards to take account of particular program goals, participant characteristics, and environmental conditions, and thus, to judge different programs in different circumstances appropriately.

Arguments in favor of adjustments to WIA performance standards regained momentum during the 2000–2009 decade. The GAO and others recommended that the USDOL develop procedures that could be used by states and localities for making adjustments for local economic conditions and client characteristics. Until such adjustments are implemented, it is difficult to say whether the gaming and resulting shifts in populations served are reduced when adjustments in performance standards are introduced.

A set of adjustment techniques for WIA developed by Randall W. Eberts is presented in Chapter 9, “Recent Advances in Performance Measurement.” Eberts created the adjustment system for the USDOL, so his chapter is designed to help one understand what can be done in the WIA context, using the much richer data sources now available than under JTPA. Eberts’s objective is to develop procedures that can be used to adjust state and local WIA performance targets for factors that affect performance outcomes but are outside the control of state and local administrators. The intent is to level the playing field by making the targets neutral with respect to the observed characteristics of WIA participants and of the local labor market conditions in which they seek employment.

As noted earlier, the lack of adjustments in setting performance outcomes has been a major complaint about WIA, especially since the predecessor program, JTPA, had allowed particular statistical adjustments (derived from regression estimates) to be employed by SDAs in setting their performance standards with the states. WIA, in contrast, called for performance outcomes or standards to be set through negoti-

ated standards between federal and state offices, with no allowance for particular adjustments.

Eberts's techniques require two adjustment procedures to reach a general adjustment model: first, national performance targets are adjusted for changes in the unemployment rate using regression estimates, and second, state and local performance targets are adjusted for differences in local market conditions and personal characteristics of WIA participants. This results in adjusting each state's targets according to the extent a state's participant and local labor market characteristics differ from those at the national level.

The weights used to adjust the values are estimated by using data on outcomes of individual participants of workforce programs from the WIASRD rather than the aggregated local data used under JTPA adjustment formulas. Thus, this adjustment procedure for WIA relies on direct estimates of the effects of unemployment rates on performance measures for various programs at the local level using the data on individuals in the three programs within WIA: Adults, Dislocated Workers, and Youth. Further, the adjustment framework assures the targets for local workforce areas and state targets add up to the national target.

The tables in Chapter 9 provide the estimating models and results. They present the variation in unemployment rates at the local (county) level nationally from January 2000 to November 2008, as well as the estimated relationships between participant characteristics and the five WIA performance measures (entered employment, retention, average earnings, credential, and employment).

Eberts also demonstrates how the adjustments from the statistical model compare, for each state, with negotiated performance levels and actual performance levels, with a wider spread observed for the statistical model. The results using data from 2006 (Table 9.6) reveal that adjusting the performance standard for a state (e.g., percent entering employment) could increase the difference between actual performance levels and the adjusted performance standard versus the prior difference between the actual and the negotiated standard. These increases go in different directions. For some states, the procedure brings the adjusted standard closer to the actual; for other states, the adjusted standard moves even higher than the unadjusted (negotiated) standard. The impact may not be favorable for all states; some do better without the adjustment of their negotiated standard.

Even with adjustment procedures, a performance management system still has to incorporate how the results will be used. Will the results be linked to incentives or rewards in support of meeting or exceeding goals, as well as penalties of some kind? In Chapter 10, “Financial Performance Incentives,” Stephen A. Wandner and Michael Wiseman review the use of incentive awards, called high performance bonuses (HPBs), in three major federal social programs: 1) WIA, 2) Temporary Assistance to Needy Families (TANF), and 3) Food Stamps. Overall, they urge caution on attempting to manage through incentives, finding that generally field operations operate with the immediacy of decision making on matters that are poorly reflected in performance measures and goals. In the three major federal social programs, Wandner and Wiseman find no evidence that incentive (or bonus) schemes in public sector social programming matter in the ultimate public policy outcome because there are counterfactuals that provide a comparison. Since all administrative units (e.g., states) are placed under the same HPB, there are no counterfactuals. The authors therefore deploy case study methods to draw conclusions.

Wandner and Wiseman describe the patterns of WIA HPB grants awarded from 1999 to 2004 (funding for these grants was dropped beginning in 2005). They point out that these HPB awards were based on the negotiated agreements between states and the regional offices of the USDOL that set performance levels to encourage state and local setting of performance levels that fit with local conditions. Wandner and Wiseman observe that federal negotiators had to measure and weigh local factors on their own without uniform methodology that assured equitable treatment among states and regions. They conclude that the patterns of HPBs vary widely by state and region. Overall, during the nine years of WIA incentive funds, five states received 31 awards, or 25 percent of all awards, and nine states received no awards. “The variation is so great,” observe Wandner and Wiseman, “that it appears that USDOL has been, in part, rewarding behaviors that attempt to game the system . . .” and that “whole regions of states garner[ed] a significant number of awards.” In addition, the WIA monetary incentives were very small, so one would not expect a strong relationship between WIA incentives and WIA programs. Indeed, state plans on intended uses of incentive grants show that states used the funds for new programs or increases in

services rather than individuals involved in frontline service—there has been no incentive for staff to provide more effective services.

TANF provides another example, say Wandner and Wiseman, of “no evidence” that an HPB in a federal program affected state policy or program effectiveness. The size of TANF HPBs was small, averaging overall \$200 million per year, or less than 1 percent of total outlays, they report. Also, in contrast to WIA, the program was voluntary for states (yet, most states participate in the competition), and capped at 5 percent of a state’s TANF block grant.

Wandner and Wiseman offer several observations on the TANF experience with HPBs. Since the TANF HPB program was based on information not available to state- and local-level program managers, it could not provide any real feedback to program operators. Further, the Department of Health and Human Services did not link or publicize possible best practice lessons that might have been discernable among states winning the top awards. Wandner and Wiseman also note that, as with WIA, the TANF HPB system did not allow one to decipher what particular performance areas drove higher rewards (too many indicators were part of the process). While top state managers may have welcomed the public recognition of the bonus awards, they used the funds for special projects rather than individual staff recognition. The TANF HPB operated from 1998 through 2004, ceasing in 2005 in the TANF reauthorization.

The third case study on financial incentive results in federal programs focuses on the Supplemental Nutrition Assistance Program (SNAP), formerly called the Food Stamp Program, administered by the Food and Nutrition Service of the U.S. Department of Agriculture and operated locally by state governments or by county governments with state supervision. The HPB for SNAP, conclude Wandner and Wiseman, is better designed and operated, but the program’s small size and universal availability make its impact difficult to assess.

Wandner and Wiseman point out that the HPB application to the Food Stamp Program operates in a different program environment. Most important, the SNAP benefit is delivered by electronic benefits transfer into a special credit card for recipients to use to purchase food, with eligibility determination monthly. While this federal program has much larger outlays (\$37.7 billion in the 2008 fiscal year) than either TANF (\$25 billion) or WIA (\$16 billion), it is an entitlement and has

clear eligibility rules. The policymaker issues tend to be assuring that those eligible are reached and that error rates are kept to some tolerable levels. The authors note how much attention went into quality control of the benefit errors prior to the 2002 initiation of performance measures and HPB payments, thereby clearly establishing the measures.

While the HPB awards paid out under SNAP are small (\$48 million in 2008), Wandner and Wiseman find that nationally SNAP had increasing averages on performance measures. It is impossible to conclude that this would be due to the HPB, but it is consistent with the intended effect. Wandner and Wiseman believe that HPB may be working in SNAP to improve performance because the HPB is based on the direct connection with what is done and what should be monitored at the ground level. Not only is the HPB directly tied to local operations, it can be audited and has very good statistical inference, meaning the precision of the estimates is calculable, with confidence intervals around the point estimates. Finally, the Food and Nutrition Service has made efforts to link the HPB results with promising practices.

Wandner and Wiseman also report on other studies on performance pay within differing agency environments and conclude that the same finding is repeated in each study: agency staff react by selective reporting; frontline staff tend to cream skim. They also point to significant difficulties encountered by federal agencies in managing federal–state performance bonus systems. Wandner and Wiseman sum it up: “High-performance bonuses in government programs [are] an inefficient use of federal resources.”

IMPACT EVALUATIONS

Impact evaluations of federal programs are now regularly mandated in Congressional authorizations and appropriations. In Chapter 11, “Ten Years of WIA Research,” Paul T. Decker summarizes the most influential evaluation research on workforce programming, and relates it to the state of evaluation knowledge that has been available so far on WIA.

Decker first looks at findings from implementation studies of WIA during the early WIA years, through 2006. He examines whether the

seven key principles of WIA were fulfilled. Overall, he finds that most implementation studies are in agreement on which WIA program principles seem to have worked relatively well, and which are more mixed. This reflects the particular new WIA operational features a study was designed to address. None of the implementation studies were able to look at what was going on across all the states; most were designed to look at WIA principles expected to produce problems in the early part of the decade within selected states.

For service coordination, a key change sought by Congress, Decker finds that it generally succeeded through local One-Stop service centers, but that there have been challenges, including the fact that WIA's mandatory partners have made only limited financial contributions, and that conflicting goals impede partnerships. Decker also cites the lack of common data systems. On prioritizing customer choice, Decker concludes that “. . . local workforce investment agencies have enthusiastically embraced customer choice by offering a wide range of core and intensive services and establishing ITAs to facilitate customer choice of training,” but there are weaknesses due to incomplete cooperation by training providers in providing information to meet the ETPL requirements.

Decker notes that strong positive responses are apparent over the decade to the universal access principle of WIA, and he marks it as an area of great progress. However, tensions exist between core and intensive services for a wide range of customers with a smaller group getting the more extensive training. The emphasis on performance management as a driving force for effective service delivery is marked by mixed success. The 17 performance goals were too numerous and complex, the data used to measure performance were of uncertain reliability and received too late by agencies to use in managing the program, and local agencies tended to focus on managing the performance system to make the numbers. Decker finds that progress was made by the USDOL in 2005, in response to demands for common measures.

The WIA principle of close connections to and with the private sector is another area with mixed results. Decker points to the substantial local variation in how much state and local WIA agencies have connected with the private sector. WIA remains a public policy area with examples of success and examples of disappointment in connecting with the private sector. Decker says that while youth programs have

been implemented, it is still challenging to find eligible providers, find and retain at-risk out-of-school youth, and establish WIA eligibility.

Setting the stage for considering WIA evaluations, Decker briefly recalls the results from large-scale evaluation studies of federal employment training programs prior to 1995. He believes that the national study of JTPA, the predecessor to WIA, was “a critical turning point in the creation of evidence.” Not only did the study use random assignment of applicants to a treatment group offered JTPA services or to a control group denied access to JTPA, but by design the study sought to obtain a nationally representative study sample—a first for an experimental study in employment training program evaluations. The JTPA study found that overall, men and women obtained equivalent net benefits per enrollee. Subsequent longer-term follow-up analyses conducted by the GAO (using Social Security earnings records) looked at impacts five to six years later, finding sustained earnings gains among both men and women.

Decker summarizes as well two experiments testing dislocated worker interventions that were conducted during the JTPA period. These tests of changes in dislocated worker programming focused on similar populations but had different program conditions. The conclusions from both interventions were comparable and had considerable impact on policy. It was shown that by using job search assistance only treatments with dislocated workers, the workers speeded up in the timing of their reemployment and had increased earnings versus the dislocated workers who had no job search assistance offered or required. Although the impacts were short lived, the benefits outweighed the program costs, so taxpayer funds invested in this strategy had a payoff. Also, in both demonstrations, those that offered training on top of the job search assistance had no greater outcomes than those in the job-search-only group. The findings stimulated changes in state UI programs, specifically, the use of statistical recipient profiling to identify UI recipients likely to face long unemployment spells, and to direct UI recipients to mandatory reemployment services as a condition of continued benefit payments.

Further studies of mandatory job search assistance for profiled UI recipients in the 1990s confirmed the earlier findings from the Texas and New Jersey demonstrations. Decker notes: “In contrast to the substantial body of evidence on JSA’s effects for dislocated workers, the

effects of more intensive classroom training or of job training have not been fully tested for dislocated workers using an experimental design.”

WIA replaced JTPA in 1998. The first USDOL experimental study of WIA impacts, now in field operations, is comparable to the JTPA study in design and focus, and uses a nationally representative sample design. Decker reports that the study uses random assignment of applicants to a group that has access to all WIA services that will be compared to one or more groups with limited or no access, again seeking a nationally representative sample. Decker explains that the only random assignment study of WIA prior to the new national study focused on a program implementation issue that arose early in WIA’s history. This was the question of how best to provide ITAs, the vehicle for funding training under WIA. Due to the changes made under WIA in how training opportunities were accessed, it was decided to determine if it mattered how WIA center staff offered the ITA vouchers. The study compared three alternative methods of administering ITAs: guided customer choice, structured customer choice, and maximum customer choice. Decker concludes that the ITA experiment “supports the widespread use of the ‘guided choice’ model by local agencies in the current [WIA] environment.” The study is now in a long-term follow-up phase, tracking outcomes six to seven years.

Decker notes that a sequel to the ITA experiment—the personal reemployment account (PRA)—extends the training voucher question to the dislocated workers. It was designed to test vouchers offered to UI recipients as an alternative to participation in WIA. It took place in seven states in 2004. Findings echoed the ITA experiment reports. Finally, a third USDOL study on training vouchers began in 2006 and is ongoing. Career advancement accounts (CAAs) rigorously test how best to structure training vouchers, and test this new type of ITA by offering it to spouses of military personnel in 18 military installations in eight states.

These three studies—ITA, PRA, and CAA—are all rigorous, in-depth investigations of what happens if public sector funds are funneled through voucher programs under WIA into support for adults seeking training to find a new or better job. The results will undoubtedly be useful for the next decade, as the expansion of WIA as an entry point to not only jobs but education and training raises issues about the most cost-

effective strategies to accelerate preparation for skill-specific needs in the economy and competitive positioning. Decker summarizes the findings from two nonexperimental studies of WIA (see Chapters 12 and 13) and compares them with the earlier JTPA and dislocated worker study findings, adjusting all earnings estimates into 2005 dollar equivalents for ease in comparisons.

Kevin Hollenbeck reports on nonexperimental estimates of WIA impacts in Chapter 12, “Short-Term Net Impact Estimates and Rates of Return.” Hollenbeck utilized data from studies conducted independently of each other in response to issues within particular states, so the WIA study samples were determined by particular programs or services that were each study’s foci. As noted by Hollenbeck, the studies used the entire universe of program exiters in selected years in three states with varied time periods. Further, each study examined a slightly different set of workforce development programs covering different time periods, and thus each study selected slightly different population groups drawn to the particular workforce programs of interest for the studies. In most cases, the program service population for the WIA adult and WIA dislocated worker groups could be identified within the state study, and thus the findings across states for these groups could be combined. The results focus on the programs offered under WIA for job training in order to compare with JTPA impacts.

Hollenbeck examines the earnings and employment impacts and hours of work and wage rate impacts from participation in WIA Adult, Dislocated Worker, and Youth programs, including how these key outcomes changed over time post program. Hollenbeck also estimates the benefits and costs, incorporating estimates of impacts on fringe benefits, tax payments, and income-conditioned transfers, to arrive at rates of return from the programs for the public and society as a whole, or the rates of return for individuals served by the programs, for state taxpayers, and for society as a whole.

Hollenbeck combines the program administrative data in the WIASRD system with state UI records and state Employment Service (ES) records. Comparison groups were constructed using propensity scoring to statistically match individuals who had not participated in WIA within each state to the WIA participants in the state. The matching relied on the administrative records available through WIA, ES, and UI systems.

The study states covered in Hollenbeck's analysis were Washington State (two studies, one very early, 1998–2000, and the other later, during 2002–2004); Virginia (2004–2005); and Indiana (2005–2006). The state study samples were constrained in Washington State, as WIA records did not include the date of entry, only the date of exit. Therefore Hollenbeck conducted his impact estimations across all the study states using quarter of the date of exit from WIA as the starting point for the follow-up analysis. Hollenbeck assembled administration data at the individual level for the treatment and comparison group samples receipt of transfer income from UI benefits, Food Stamps, Medicaid, TANF benefits, and fringe benefits and taxes on earnings, as well as the employment and earnings outcomes. These allowed estimating not only employment and earnings during the preprogram and postprogram outcome time periods but the ways WIA participants might have received both positive and negative benefits by participating in WIA, the benefits or costs to taxpayers, and the overall social benefit-cost estimates that combine both the program participant gains or losses and the taxpayers' benefits or costs.

Hollenbeck concludes there were strong and positive results on the post-WIA earnings for adult WIA populations in that all appear to have statistically significant earnings and employment impacts from participating in WIA although of varying magnitudes and trends over time depending on the state study. The point estimates of average quarterly earnings for the WIA Adult program show gains in earnings on average, beginning in the short-term time period (two to three full quarters after program exit) of \$146–\$711 per quarter. The WIA Adult earnings gains during the long-term follow-up time period (from 4 quarters to 12 quarters after program exit) average \$455–\$463 per quarter.

The results for youth in WIA are less positive. Hollenbeck estimates that short-term earnings gains among youth in WIA are near zero and not significant. The long-term average earnings gain among WIA youth was mixed—in one state study it was significant at \$325, but in another state it remained not significant and near zero. It should be recalled that under JTPA, estimates of youth earnings gains from training were negative and statistically significant in the short term, and near zero and not significant in the long term.

For two of the state studies, Hollenbeck was also able to decompose the net impacts into employment, wage, and hours impacts, finding

positive net impacts and returns on investment for virtually all of the programs. He also finds very strong and positive, statistically significant impacts on earnings for the dislocated workers who participated in WIA in the short term and the long term, varying between \$410–\$784 and \$310–\$771, respectively. Because of the baseline for the studies at program exit, the opportunity costs or forgone earnings experienced by dislocated workers from entry into WIA are not incorporated in the short- and long-run earnings and employment estimates.

Hollenbeck's benefit-cost analysis estimates that the discounted net benefits to participants over the first 10 quarters after exit range between \$3,500 and \$5,000 over all three groups. There are important differences between the groups. The costs to dislocated workers of participating in WIA (the forgone earnings) are so large that the net benefit return for dislocated workers is consistently negative. Hollenbeck estimates the losses to those entering WIA by tracking through ES and UI files the preprogram earnings and employment. He also projects from the estimates for the first 2.5 years after exit from WIA, all the private and public benefits and costs over the first 25 years after program exit, and over the working lifetime.

These projections show that for dislocated workers and youth in the states studied, the private and public benefits and costs from their participation in WIA produced overall negative rates of return. In the case of dislocated workers, the main explanation of the negative rate of return is the loss of earnings they experience, on average, by participating in WIA, and essentially delaying reemployment and the earnings benefits obtained that the comparison group obtains. In addition, the public costs for training of dislocated workers outweigh the public benefits obtained in Hollenbeck's estimates. However, employment gains still exist and are strong enough so that Hollenbeck suggests considering policies, such as a stipend, for dislocated workers in the training programs to offset the forgone earnings.

The youth population also has a negative social return that outweighs the marginal economic gains in Hollenbeck's benefit-cost analysis. The earnings gains for youth were essentially zero, so the program costs easily exceeded the benefits of WIA serving youth, at least in the one state study undertaken in Washington State by Hollenbeck and Huang (2003). The earnings and employment gains estimated for the first 2.5 years after the adult population exited WIA training were enough to

outweigh forgone earnings, and allow Hollenbeck to conclude that the overall social benefits were greater than the social costs.

A national nonexperimental study of WIA impacts was undertaken by a team led by Carolyn J. Heinrich with Peter R. Mueser, Kenneth R. Troske, Kyung-Seong Jeon, and Daver C. Kahvecioglu. The study is summarized in Chapter 13, “A Nonexperimental Evaluation of WIA Programs.” The objective was to reach the national WIA participant population in a study for the USDOL, but Heinrich and colleagues caution that they could not obtain a truly representative national sample. Heinrich and colleagues use the point of program entry to begin the study observation period. The study evaluates two WIA programs: the Adult program, serving largely disadvantaged individuals, and the Dislocated Worker program, serving those who have lost jobs.

Heinrich et al. draw a number of conclusions regarding the most immediate or short-term (immediately after WIA entry) earnings impacts WIA participants obtain, contrasted with their longer-run patterns of gains (up to four years after WIA entry); differences by gender; differences between the Adult WIA program participant gains and the Dislocated Worker WIA program participant gains; and possible interactions of earnings gain patterns with various selection bias considerations, such as measured differences (and the unobservable differences) between the participant and comparison study samples at baseline and preprogram and the variance in participation patterns in WIA.

In discussing the conclusions on short- and long-term impacts, Heinrich and colleagues emphasize how different the results are for the Adult versus the Dislocated Worker programs. They also stress that by examining the likely long-term benefits of training—the benefits estimated for the last 11–16 quarters (generally the fourth year after program entry)—one obtains some gauge on whether the WIA programs pass a benefit-cost standard.

Earnings for men and women in the WIA Adult program increased during quarters 11–16 after WIA entry that average annualized earnings gains of 26 percent for women (\$2,363) and of 15 percent for men (\$1,676). The employment rate increments estimated are 12 percent for both men and women, or employment rates rising by about 6.5 percentage points. The WIA Dislocated Worker program estimate by Heinrich and colleagues presents annualized earnings gains in quarters 11–16 after program exit that are very small and not significant. Employment

rates, however, appear to increase by 4–5 percentage points, or 7–8 percent gains, and are statistically significant. They point out that dislocated worker populations are, on average, those who have strong work histories and higher wages so their entry into services such as training or extensive career counseling remove them from the successful job finding process evident in control groups in experimental studies of dislocated workers. There are diminished earnings and employment for dislocated workers during program participation, with about four years needed to recoup and return to the “normal” pattern, and eventually show some earnings gains.

For the Dislocated Worker program, the earnings impacts would need to be long lived to exceed costs, and earnings gains for dislocated workers who are men are basically not discernable, meaning benefits do not exceed costs. The study estimates that the WIA Adult program clearly satisfies a benefit-cost standard for both men and women if the earnings impacts continue for a period of just two or three years. For the Dislocated Worker program, the evidence is much less clear.

Heinrich and colleagues describe the latitude in WIA that states have used to structure the One-Stop system to reflect local preferences, under direction of the local agency, the WIB, stressing that there are wide variations across localities. They note that the sequential service mandate may cause “negative selection into training” because one must have been unsuccessful in obtaining employment through core and intensive service sequences to be eligible for training. On the other hand, they also note that it has been found in most sites that “as many as a third of those who participate in WIA have a particular training goal prior to program entry (they are often referred to WIA by the training provider), and, in general, WIA staff make an effort to accommodate them.” It is also expected that the performance measures encourage positive selection of those perceived to be most successful in the labor market for WIA services.

The authors point out how there is no simple picture of what services a customer receives under WIA. For example, a recent study finds that nationwide, about one in five WIA participants received only core services, and about two in five were coded as receiving training services (Social Policy Research Associates 2006). Of those receiving training, up to 10 percent received on-the-job training and another 5 percent received basic skills training, with the remainder receiving occupational

and other training. It is also interesting that half of all training recorded was funded by ITAs, with two-thirds of those in training receiving some kind of credential. Somewhere between a half and a third of WIA participants exited the program in less than 26 weeks, the balance remaining in WIA and then exiting were in the program for at least a year. The impact estimates in the study could not consider how all these different service patterns might shift the overall average earnings and employment estimates per participant in the Adult or Dislocated Worker programs, but there is considerable study discussion of how these patterns could be theorized to influence such estimates.

Overall, the main conclusions are the consistent and significant gains obtained by women and men in the WIA Adult program, in contrast to the lack of such gains, in general, among those in the WIA Dislocated Worker program, and that conclusions regarding program effectiveness should not be based on the very short-term recorded earnings and employment within the first year after program exit. To ignore the more significant gains for Adult program participants, which emerge by years three and four, misses a potential public sector investment payoff rarely found through solid, rigorous evaluations. Heinrich and colleagues urge investing random assignment studies that can offer findings that are more confirmatory. Some of the study findings appear consistent with the earlier random assignment study of JTPA, especially in the ineffectiveness of providing training as the main service for the Dislocated Worker population. The Heinrich team could not replicate the important studies of the effectiveness of incentives that encourage swifter reattachment to jobs, but they note that these strategies appear to be a more efficient approach. The Adult program findings, on the other hand, support those who believe job search assistance and training services are effective when dedicated to those with weak employment experience or attachment to the mainstream economy.

FUTURE EVALUATION CHOICES

The term *evaluation* encompasses many different efforts to assess the effects of a policy, program, or particular practice. Whatever the focus, central to evaluation standards is how well the counterfactual

produces reliable impact estimates. The next set of papers reviews the conditions that need to exist to implement particular evaluation designs, noting which conditions reduce or increase reliability.

In Chapter 14, “Nonexperimental Impact Evaluations,” Haeil Jung and Maureen A. Pirog review the history of employment and training program evaluations, focusing on the Manpower Development and Training Act (MDTA), which began in 1962, and CETA, which began in 1973. Evaluations of CETA were nonexperimental and drew samples of participants from a longitudinal manpower survey that tracked participants to compare with those not in CETA, drawn from the Current Population Survey. Jung and Pirog recall that these evaluations produced widely varying findings that drew serious examination of why the findings varied, and stimulated beginning efforts to use experimental designs, which had begun to test other employability programs outside of the CETA system. Evaluation specialists began to compare findings from experiments with efforts to replicate the experiment using comparison groups such as one would draw from the Current Population Survey and other sources.

Jung and Pirog describe the outgrowth of the replication studies, an ongoing methodological concern with what data conditions and which nonexperimental methods might be the second-best fits to what would be obtained if an experimental study could have been undertaken. Accompanying this were efforts to define and measure the types of selection bias that produced estimates from nonexperimental studies that did not get close to experimental estimates. They point out that, although there may be logistical difficulties encountered in implementing a random assignment experiment, these difficulties must be weighted against the likelihood of giving bad advice to policymakers, the likely result of applying nonexperimental techniques in many situations without taking account of the assumptions.

Jung and Pirog aim their discussion to those who, it is assumed, are not able to use experimental methods given the constraints from program operations such as mandatory participation and thus must choose among nonexperimental methods. They first observe the various types of questions that experimental methods address, and then discuss the sources of selection bias that an econometric estimator might correct for. Jung and Pirog point out that there are different types of selection bias in training programs, and the challenge is to understand how the

sources of the selection biases might operate in particular training program contexts.

Four key conditions that nonexperimental evaluations need to have in order to reduce the measurable sources of bias, according to Jung and Pirog, are 1) comparison groups drawn from the same local labor markets as the programs studied, 2) the same instrumentation used to collect data from the treatment and comparison groups, 3) the same range of values for the observed characteristics of the comparison group's members as for treatment group's members, and 4) the same distributions of the values across the ranges of the treatment and comparison group.

Jung and Pirog also advise caution in specifying the policy question, noting that the policy question addressed in intent to treat evaluations is different from the question addressed in treatment on treatment evaluations. The former includes the applicants assigned to the program who may drop out after baseline and thus addresses whether the program overall had a desirable impact on the eligible population. Evaluations that focus on those who received the treatment are aimed at the effects for a subgroup of those eligible. Thus, these evaluations cannot capture the overall policy effectiveness of expenditures on treatments offered to a much larger group of eligible people. The authors review the modeling methods used to work on estimates for groups who somehow do get treatment (the subset of the intent to treat population), relying on the observables captured in data sets. These techniques include difference-in-differences extensions on matching, regression discontinuity design, and the marginal treatment effect using local instrumental variables, and are summarized along with earlier modeling methods such as regression estimators, propensity score matching, and difference-in-differences.

In Chapter 15, "Designing Reliable Impact Evaluations," Larry L. Orr, Stephen H. Bell, and Jacob A. Klerman give an overview of the directions taken in evaluation science over the 40 years of efforts to evaluate job training programs, pointing out how over time the standards become more certain and focused on experimental designs. Orr, Bell, and Klerman note that for the first job training studies of MDTA during the 1960s, evaluations started out with simple before-after methods looking at whether postprogram earnings improved over pre-program earnings.

Important work by economists pointed out that preprogram earnings are simply a marker of the reason why one comes for job training:

one has lost a job so earnings are nil; then one finds a job, whether through a job training program or other strategies. Rebounding from a job loss naturally leads to most eventually finding another job. This Ashenfelter dip, the natural drop of earnings due to job loss–job recovery before one needs the services of a job training program, means one has to have some comparison of similar people who are also going through the job loss–job gain cycle. During the 1970s, evaluations focused on using data about people who looked similar in that they too had suffered loss of a job.

By the 1980s, economists analyzed how well methods using such data sources (typically national longitudinal data) could replicate the findings from some of the first national experimental studies of workforce programs, such as the Supported Work evaluation by Manpower Demonstration Research Corporation. Recognition that past, nonexperimental studies could not produce scientifically valid program impact estimates brought congressional requirements for more rigorous evaluations. A prime example was the serious investment in the National JTPA Study, using experimental methods to assign over 20,000 applicants to job training or control groups in sixteen local programs and study the outcomes for an extended period. Other workforce programs received rigorous evaluations as well.

Orr and his colleagues stress that experimental methods (using random assignment to allocate applicants to the program or to a control group) are not only scientifically accurate, but they avoid the methodological debates that accompany presentations of nonexperimental results. The lack of comparability between nonexperimental evaluation methods versus the experimental method is the fact that by randomly assigning eligible applicants for a training program into two groups—the treatment group allowed to enter the program and the control group that is not allowed to enter the program—the two samples, due to random assignment, have by chance the same set of background factors represented in them. Most important, they have the same set of unobservable characteristics, motivations, and experiences that are unknown. If, for example, one were to select a comparison group of nonparticipants as the counterfactual, one cannot ever be assured that the factors are taken into account that turned them from potential application and participation into a nonparticipant.

The authors review efforts to improve nonexperimental methods to bring them closer in credibility as experimental studies. This question stimulated many methods studies, testing how close the results from these methodological developments are to those from a study done with experimental methods. They note that it is particularly the case in workforce program evaluations that the nonexperimental methods are “not well-suited” to econometric modeling of job loss–job gain and the accompanying loss of earnings. Those who have come the closest have the benefit of data sets of large numbers of individuals that have extensive longitudinal data on the employment and earnings on the individuals.

Orr, Bell, and Klerman also emphasize that a major problem with using nonexperimental methods is that, before a program evaluation is put into the field, one has no really viable rule that specifies what will produce the estimate closest to the estimate obtained in an experimental evaluation. Thus, while after an experimental study is completed, one may check out which nonexperimental method applied to the experimental data appears to come closest, before the study one has no way of judging this. Design parameters are critical for estimating study costs, and are more indeterminate if a nonexperimental method is used. With an experiment, there are much stronger estimates possible of what will be required to conduct the study, so both the costs and the likelihood that the study will find significant effects if they exist are firmer and provide more assurance to policymakers that funds are being wisely applied.

Advances have been made using experimental methods, as Orr and his colleagues explain. They should be recognized and their use encouraged, for they demonstrate that it is possible to apply experimental methods to a variety of program conditions. The authors provide examples, including spreading control groups over many sites, decreasing the allocation percentage, as well as allowing program operators to increase the odds of assignment to treatment for preferred applicants or testing greater levels of intervention versus the standard services, and using administrative data instead of surveys.

The latest plans for a national random assignment study of WIA are discussed in Chapter 16, titled “Neither Easy Nor Cheap,” by Sheena McConnell, Peter Schochet, and Alberto Martini, who are leading the new WIA experimental study. The authors remind us of the ways evaluations have affected policy and programmatic decisions, new demands

on experimental studies, and how future evaluations may need to be tailored.

The first evaluation of the Job Corps (which actually was nonexperimental in design) found it cost-effective. This was the first time policymakers had a major, national longitudinal study assessing a job program, and the findings led them to increase the funding and sustain the program. Job Corps has remained popular, with the longest “life” next to Head Start. When the JTPA evaluation findings for youth were released much later showing there was no significant earnings gain, and even hints of negative impacts for some subgroups, the response was major programmatic changes and reduced funding as well. One experimental test of reemployment services for jobless people on UI led to changes in UI services, requiring worker profiling and reemployment services for recipients likely to exhaust benefits before finding employment, basically requiring those recipients to receive services much as welfare benefits are conditioned on participation. Another experimental study led to the creation of the Self-Employment Assistance program for those on unemployment compensation, to help the unemployed start their own small business.

McConnell, Schochet, and Martini urge planners of future evaluation to pay attention to lessons learned from over 30 years of experiments in workforce programs. First on their list of key lessons is the careful development of the evaluation questions a study will take on. (A rigorous study can only accommodate a few questions.) The question dilemma first arises by confronting whether to evaluate the whole program or which components of a program are effective. The authors believe the second strategy is advantageous in workforce program evaluations. Whole programs these days have too much complexity, too many moving parts. One may learn more by focusing on particular program areas aimed to help identified target groups. Target groups, however, need to be clearly identified in the evaluation plan, as they are typically broken into subgroups. One is constrained by evaluation costs to a limited number of subgroups. Trying to collect a lot of demographic variables and then fishing around later, testing out different subgroups, is considered “milking,” and reviewers will ask what the original plan was for testing the key hypotheses driving the program design and the subgroup characteristics specified *a priori*. McConnell and her colleagues also remind evaluation planners that the procedures

and data to be used to draw the counterfactual are also critical to assuring that one can obtain the sample(s) needed for the questions.

On the question of evaluation design, in situations where a “no service offered” control group cannot be established within the existing program, the authors urge trying an experimental design that uses “randomized encouragement,” in which all the eligible participants are assigned to services, but the random assignment sets up a lottery assigning an encouragement to some participants to voluntarily use more services. This incorporates random assignment into program operations, with the assignment being to the encouraged group or to the regular service menu.

McConnell, Schochet, and Martini also provide a summary of non-experimental designs, noting the difficulties involved. They strongly recommend that the regression discontinuity or propensity score methods be used rather than less credible methods. They conclude that the regression discontinuity approach has the most promise when experimental methods are not viable. The selection rule for receiving the treatment is fully known under the regression discontinuity approach. The propensity score approach has to rely on modeling using observable baseline data, so one cannot know for sure whether the unobservables are introducing substantial bias into the findings.

The authors note the methodological efforts to use propensity score methods that match program participants to a comparison sample and appraise how successful these methods are at getting estimates that are stable and similar to experimental results. McConnell and colleagues conclude that while some estimation techniques appear more successful than others in replication studies focused on particular experimental data sets, they believe that it is the data available that can be used for the comparison group that determines the validity of the estimate, rather than the nonexperimental technique. In other words, the things that matter are whether the data contains an extensive and good set of matching variables for modeling the participation decision, such as extensive pre-program earnings histories, samples possible from the same geographic areas as the experimental study group, and same follow-up data sources available for treatment and comparison groups.

McConnell and colleagues also summarize the issues that arise when using administrative records, especially the state UI data that studies on WIA have used. They point out how important it is to have data

on service receipt among program participants and the control group. Typically the latter requires special surveys. Having data regularly collected in administrative systems has been a great advantage in studying programs (for both experiments and nonexperimental studies). And, they argue, an experimental design will pay back dividends. The costs of conducting experiments are not nearly so high as once believed, with the availability of systems tracking employment and earnings that can be linked to the program records, and by doing the assignment within the workforce system, comparing different alternatives for important populations, policymakers have very credible estimates about whether more costly alternatives matter, as well as the realities of who is attracted to these programs, who stays, and who seems to obtain real benefits.

Turning to lessons from evaluations of U.S. workforce programs to the evaluation of Active Labor Market Programs (ALMPs) in the European context, Jeffrey Smith in Chapter 17, “Improving Impact Evaluation in Europe,” draws on the North American experience and contrasts it with European practices. He makes the case for three particular features in the North American context that he believes would improve policymaking in the European context: 1) greater use of random assignment methods, 2) recognition and adoption of serious cost-benefit analysis, and 3) greater attention to developing and using evaluation industry entities such as takes place in North America. Smith points out that there are European practices that could or should be considered as improvements within the North American context, including the rich, well-maintained, and accessible administrative data and specific data elements (e.g., caseworker ratings of clients, the emphasis on documenting sanction regimes for benefit recipients). Further, he notes that European countries vary greater in their research and evaluation capacities being applied to ALMPs.

In support of investing in experimental studies, Smith notes that the “key advantage [is] that their simple design makes them easy to explain and hard to argue with . . . giving them a policy-influencing power not enjoyed by even the cleanest nonexperimental designs.” Further, the high-quality data produced yield substantive advances in understanding labor markets. Smith acknowledges that there are limitations, including the questions that cannot be addressed through experimental methods. Some have difficulty understanding what use can be made of experimental results, noting the treatment dropouts and crossovers from

control groups, and the limitation on external validity, or generalization to people or areas that were not represented in the experiment. Social experiments that aim for “national representation” cost a great deal, because compliance of the scientifically drawn sample may require extra resources, but they satisfy those who want the most general assessment.

There are also policymakers and program administrators who will express ethical objections to random assignment. Smith comments, “In my experience, these objections nearly always represent a cover for simply not wanting to know the answer.” While it is empirically the case that “many, maybe most, programs fail when subjected to serious evaluation,” Smith counters that this may indicate that indeed it is “important constituencies, such as workers and agencies or firms that provide the treatments” who have an interest in a program’s existence in the first place. He notes that these constituencies also have an interest in low-quality evaluations or misleading performance measures in place of experimental evidence that compels others to challenge the program’s existence.

In response to those who express concerns about service denial, it is quite possible, where resources are not constrained, to design random assignment so that all receive some type of service, with multiple treatment arms. In cases where a program can only serve a portion of the presumed beneficiaries, the experimental design is an equitable way to allocate the scarce resources. There are also randomized encouragement designs, with the assignment being to an incentive to participate, where no one is excluded, and the incentive provides an exogenous variation in treatment status. Another type of design is “randomization at the margin,” where the group selected for randomization are those at the end of the list deemed most critically in need of services.

All in all, notes Smith, the experimental evaluations of workforce programming conducted in the United States have often led to policy changes, as he recalls a number of examples where shifts in policies and program funding were the results of an experimental study. He also explains why cost-benefit analysis is the most important contribution to arriving at “a direct policy conclusion.” The message is clear about the value of a program to the taxpayers who fund it. An important example is the cost-benefit analysis associated with the U.S. National Job Corps Study that has brought rethinking of a long-standing program.

Smith observes that “the modal European ALMP evaluation . . . contains no cost-benefit analysis at all.” While it is true that Europeans stress employment as the key outcome in workforce programming rather than earnings (one needs earnings to have a monetary benefit in cost-benefit analysis), he believes that recognition that employment impacts typically are negative or zero, discourages serious consideration of benefit-cost analysis. Smith also notes that the public sector in Europe does not believe it so important to cost out public services. Nevertheless, knowing the duration of program impacts is important as well, and a part of cost-benefit analysis.

Finally, Smith considers why there are such marked differences in the quantity and quality of workforce evaluations across countries. He observes how robust the evaluation sector is in the United States, the variety of entities engaged, and the very large size of the evaluation sector. In contrast to the European patterns, Smith believes the centering of the research market, the research activity, outside of government is critical in that it encourages and supports independent, objective appraisals of government programs. He recommends European governments consider how they might consolidate funds for evaluations, to increase the size of the European evaluation market, and open the market to a variety of research entities, both within Europe and from outside.

Smith notes that parallel with the U.S. growth in the research sector was the broad agreement that it is best for government not to undertake directly the evaluation of programs it funds and manages. He concludes by noting the imprint of neoclassical economics as possibly influencing the agreement regarding the role of independent research, as well as other broader and deeper differences across countries in individualism, deference to authority, the importance of social class, and average education.

CONCLUSIONS FROM THE CONFERENCE

WIA brought changes in the workforce training system because it shifted responsibilities between layers of government. States became principal actors in WIA, as they were given responsibility for determining the funding to local entities responsible for outreach and

coordination and making decisions about system operations and performance standards. This shifted these decisions away from the federal offices to local and state control.

Performance management became the central focus of the USDOL's management of WIA, and for states as well. WIA offers an excellent example of the kinds of changes that can accompany stronger accountability requirements in federally funded social and educational programs. The act's performance management mandates occurred at a time of extensive computerization of administrative records throughout government, and was able to take advantage of the new systems and efforts to link records across systems. Despite the expectations regarding easy transfer of performance systems into government, however, the systems may not be operating with the incentives expected, due to confusion and manipulations possible within the systems. It is unclear, as well, whether the investments in closer tracking of the users or beneficiaries of federal programs for the purposes of performance management actually helped job seekers gain faster access to jobs and achieve higher earnings. More likely, as these data files are made available to researchers, more will be learned about how the program works and what services are provided.

Implementation of WIA took time. Most states obtained waivers from the USDOL in the early years of the act to set up performance standards for the One-Stop centers, providers, and settled on the data requirements without operating fully the new accountability structures. States had to invest in large and comprehensive databases to create lists of effective providers deemed eligible to serve WIA applicants. New and better labor market and local economy information repositories were also created under WIA. One-Stop centers focused on encouraging individual "shopping" of WIA services and training options, with both staff and customers of One-Stop centers welcoming the shared responsibility.

Nonexperimental evaluations of WIA (at least the programs aimed at unemployed adults and at dislocated workers) have had mixed findings. Two studies find significant positive effects for adult job seekers who use WIA as opposed to those who do not. The findings are very different for dislocated workers. Here the two studies find no positive effects for dislocated workers. Comparison groups appear to move more quickly into jobs and thus dislocated workers in WIA suffer lost

wages while in WIA services. Whether dislocated workers eventually overtake the earnings obtained by their counterparts in the comparison group seems to depend on the length of follow-up. Estimates of the average longer-term net earnings gain among the dislocated workers appear to be less than the average social cost estimates of the WIA program for this group.

All of the evidence on WIA program effects on employment and earnings is suggestive, rather than confirmatory. Studies that produce confirmatory evidence need experimental designs to establish participant and control groups. The suggestive findings are based on methods that are careful and clear about the limitations of the impact estimates. It is fortunate that the effects of WIA, as it is now operating some 12 years after its creation, are finally being examined in new federal studies under way using randomized designs, but it will be several years before early results are known.

Notes

1. The meeting was a cross-national policy dialogue organized by the Center for International Policy Exchanges at the University of Maryland's School of Public Policy. This dialogue, one of a series with different entities (including the Association for Public Policy Analysis and Management), was organized by the University of Maryland Center and was held on November 7, 2009, in Washington DC. The topic, "Evaluation and Performance Management of Job Training Programs: What Can the European Social Fund Learn from the WIA Experience?" was developed in response to inquiries from the staff of the European Commission concerned with commission workforce programming. Patrick Tiedemann, research associate at the Center for International Policy Exchanges, assisted in organizing the meeting.
2. Nearly \$3 billion was added to the Adult, Dislocated Worker, and Youth programs under WIA in the American Recovery and Reinvestment Act of 2009 as a one-time increase to be spent by December 31, 2010.
3. If a job seeker doesn't get a job at tier 1, they move into tier 2, and then, again if no job, to tier 3.
4. Four federal agencies retained responsibility for distributing funds under each program to states, in most cases retaining existing formulas for distribution to the states, and other requirements and regulations from prior legislation, with amendments in WIA to accommodate the coordinated access to services entry points.
5. Blank, Heald, and Fagnoni report that their June 2004 GAO study report "estimated the cost of doing participant surveys, as was done under JTPA, at approximately

\$13.25 per participant compared with the cost of automated record matching to UI wage records, which costs less than \$0.05 per participant.”

6. At the same time, the USDOL is working on an enhanced data reporting system called the Workforce Investment Streamlined Performance Reporting system, to consolidate reporting requirements across several other USDOL programs, and lead to a single reporting structure that can track an individual’s progress through the One-Stop system.
7. See Chapter 9 for a description of how adjustments to the WIA performance levels can be carried out.
8. The WIA training program for youth, however, is targeted for the out-of-school youth.
9. See especially Chapters 3, 8, and 10.

References

- Hollenbeck, Kevin M., and Wei-Jang Huang. 2003. *Net Impact and Benefit-Cost Estimates of the Workforce Development System in Washington State*. Technical Report No. TR03-018. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- Maquire, Sheila, Joshua Freely, Carol Clymer, and Maureen Conway. 2009. *Job Training That Works: Findings from the Sectoral Impact Study*. Philadelphia, PA: Public/Private Ventures.
- Social Policy Research Associates. 2006. *2004 WIASRD Data Book*. Report prepared for the U.S. Department of Labor. Oakland, CA: Social Policy Research Associates.