SOCIAL POLICY RESEARCH
ASSOCIATES

Meta-Analysis of Voucher-Based Employment and Training Programs

For the Evaluation of the Career Advancement Account Demonstration

Completed January 2012

Prepared by:

Marian Negoita
Jeff Salzman
Andrew Wiegand

Prepared for:

U.S. Department of Labor/
Employment and Training Administration
200 Constitution Ave., N.W.
Washington, D.C. 20210

Project No 1355

1330 Broadway, Suite 1426
Oakland, CA 94612
Tel: (510) 763-1499
Fax: (510) 763-1599
www.spra.com

Acknowledgements and Attributions

The authors of this report wish to acknowledge the contributions and support of the U.S. Department of Labor, Employment and Training Administration staff, especially our Contracting Officer's Technical Representative (COTR) throughout most of the evaluation, Michelle Ennis. Additionally, the original COTR, Jonathan Simonetta, provided invaluable assistance and feedback. Their advice and steadfast support have been much appreciated.

As in any meta-analysis, the success of the results depends on the helpful cooperation of many individuals who conducted the original studies on which the meta-analysis is based. For this analysis, we are extraordinarily thankful to several individuals at Mathematica Policy Research who provided assistance and additional data for the various studies they had conducted. Specifically, for the Individual Training Account study, we would like to thank Sheena McConnell and Ken Fortson. For the Personal Reemployment Account evaluation, we would like to thank Gretchen Kirby and Margaret Sullivan. For the Evaluation of Military Spouse Career Advancement Accounts Demonstration, we would like to thank Linda Rosenberg. Additionally, Public Policy Associates (PPA) assisted us in gathering information and further data for two studies, the Evaluation of Lifelong Learning Accounts and the Career Management Accounts. In each case, we would like to thank Nancy Hewat of PPA for her gracious assistance and quick response. Finally, we would like to thank Richard West of Social Policy Research Associates (SPR) for his assistance with information from the Seattle-Denver Income Maintenance Evaluation. Finally, we would like to thank Betsy Becker of SPR for her wise and generous assistance with the technical aspects of the meta-analysis.

Each of these evaluations was informative and useful in their own right, and the authors of the reports did an excellent job of summarizing the study and providing results for use in the meta-analysis. Of course, any errors that derive from the meta-analysis are the responsibility of the authors of this report alone, and do not reflect on the work of those who conducted the earlier studies.

CONTENTS

I. INTRODUCTION

During the last two decades, voucher programs have become one of the preferred methods for delivering a wide array of social services, including public occupational training. Voucher-based programs have always been a presence in social service delivery, but the last few decades have witnessed an increase in the use of such programs in the United States and other countries, especially for publicly-funded training, which for a long time operated under different principles. Yet despite the rapid increase in the use of vouchers, surprisingly little is known about the training voucher programs' overall effect compared to alternative service delivery mechanisms (Barnow, 2009). Relatively few evaluation studies have been conducted, and the results have been inconclusive at best. Whereas some studies show a positive impact of voucher programs on two central outcomes—post-program employment and earnings—the resulting patterns are not uniform, and several studies have indicated little or no effect. In an era of serious budget constraints, the question of whether or not voucher-based programs work is of central importance.

The present report is an attempt to ascertain the overall impact of voucher-based training programs on labor market outcomes such as employment and earnings. This report has been commissioned by the Employment and Training Administration (ETA) of the U.S. Department of Labor to Social Policy Research Associates (SPR). The report has essentially two parts: (1) a qualitative review of existing evidence, and (2) a meta-analysis of existing programs based on statistical analysis. The report begins with a brief definition of voucher programs, followed by a review of the main theoretical perspectives on vouchers drawn from a variety of social sciences. Subsequently, the report provides an overview of the history of vocational training in the US, putting special emphasis on vouchers and their role in service delivery. The first half of the report ends with a detailed analysis of the known training voucher programs, summarizing the key details about each and reviewing the available evidence on their effectiveness.

The second part of the paper summarizes the existing evidence on the effectiveness of training vouchers using the rigorous statistical method known as meta-analysis. Meta-analysis allows the quantitative comparison of the effects of a variety of programs. It is akin to survey research, with the distinguishing feature that whereas in survey analysis the data set is comprised of survey respondents, in meta-analysis the main units of investigation are individual research studies.

After preparing the data from individual studies so that they can be compared, meta-analysis provides estimates of the effect of a specific type of intervention at the population level.

II. LITERATURE REVIEW

What Are Voucher-Based Programs?

The literature on social-service delivery shows that voucher programs can be categorized along several dimensions, often resulting in complex systems of categories (Valkama & Bailey, 2001). Despite the numerous differences among them, however, all vouchers have several common characteristics. First, they provide their beneficiaries with a sum of money or benefits capped at a certain value; second, beneficiaries have either complete or partial freedom to "spend" the money as they wish by obtaining goods or services directly from service providers (Burwick & Kirby, 2007). Thus, voucher programs differ from programs based on direct provision of social services, whereby the government delivers the good or service directly to the beneficiary (Barnow, 2009). They also differ from programs based on contracting out government services, since in the case of vouchers the individual recipients select the service providers.

Arguments In Support of Vouchers

The conceptual arguments for using vouchers for service delivery in general, and for training programs in particular, are tightly connected to the larger framework of New Public Management (NPM)—a social science paradigm that advocates reforming the public sector so as to make it more responsive to citizens' needs, to diminish wasteful spending, and to ensure better policy coordination. Before examining the claims typically offered in support of vouchers, it is necessary to briefly summarize some of the central NPM arguments.

The NPM paradigm emerged in the 1970s as economists, political scientists, and public administration theorists began to challenge longstanding principles in the organization of public administration by claiming that bureaucracies are often inefficient and wasteful. Many of the critics of bureaucracy expressed their views in the public choice literature, stressing that one of the main reasons behind public bureaucracies' failure was the limited amount of influence exercised by the beneficiaries, or clients, in the selection and provision of public programs (Savoie, 1994). According to the literature, bureaucrats as rational actors seek to maximize inputs (i.e., their budgets), to the detriment of their clients, meaning the taxpayers (Niskanen, 2007), (Ostrom, 1987).

Advocates of NPM argue that to counter this perceived problem, government agencies must adopt practices from the business sector (Osborne & Gaebler, 1993), (Kettl, 2002). This basic idea leads to several core principles. First, the public service should reflect entrepreneurial spirit and flexibility–when necessary, formal rules should be ignored in order to "get the job done" (Suleiman, 2003). Second, it is suggested that government agencies should develop a customer-centered orientation (Pollitt, Bathgate, Caulfield, Smullen, & Talbot, 2001). In addition, NPM supports the marketization of service delivery as a means to increase customer choice and reduce costs. In other words, although the government is always the *provider* of public social services, it need not always be the *producer* of services; sometimes, private producers can be more efficient. Thus, the split between customers, providers, and private service producers is at the heart of NPM thinking (Freedland, 2001). Third, the effectiveness of government units must be assessed using performance indicators. In establishing this principle, NPM promotes the transition from "process accountability" to accountability judged in terms of results or performance (Hood 1995). This principle was significant because it made privatization, contracting out, and competition between agencies performing the same function all viable options in the pursuit of increasing efficiency (Suleiman 2003).

Clearly, many of the arguments in favor of vouchers can be linked directly to an NPM-based perspective. In fact, experienced analysts of training vouchers frequently associate the increased use of vouchers with the advance of the "reinventing government" agenda, which in turn is a core component of NPM (Trutko & Barnow, 1999). Vouchers, after all, are a tool specifically designed to boost customer choice, which is one of NPM's central goals. What is more, within the NPM framework vouchers are in theory better suited to boosting customer choice than methods such as contracting out services, because, when the government contracts out services, customers must live with the government's choice of private producers. With vouchers, however, this is not the case; customers are more or less free to exercise their judgment and choose between competing providers (Hipp & Warner, 2008).

While increased customer choice is arguably desirable on its own, arguments in support of vouchers frequently hypothesize a positive effect of increased customer choice on outcomes. As a "quasi-market institution", vouchers are expected to induce efficiency gains associated with market competition (Lowery, 1998). If consumers are allowed to choose between competing providers of vocational training, not only will they access programs that best fit their needs, but the resulting competition between providers is likely to reduce costs and increase the quality of the available training programs (Ellis, 2001).

Arguments against Vouchers

Despite the many arguments in favor of vouchers, support for vouchers is hardly unanimous among academics and practitioners. As a noted analyst of training vouchers has remarked, “For almost every argument as to why vouchers are appealing, there is an argument to the contrary” (Barnow, 2009, p. 73). Following the pattern established in the previous section, we begin the section by reviewing the main criticisms of NPM, the conceptual paradigm behind voucher initiatives. We then summarize the drawbacks related to training vouchers in particular.

NPM-inspired theories make the fundamental assumption that markets always work better than alternative service delivery mechanisms. This assumption, however, is highly debatable. Economists have developed the concept of “market failure” to describe the kinds of situations where the overall impact of markets is less (and sometimes, far less) than optimal. These situations appear much more frequently than NPM supporters recognize. Monopolistic markets, for example, lack competition and therefore fail to lower costs; information asymmetries between buyers and sellers lead to suboptimal prices; and underinvestment in public goods results from the producers’ incapacity to extract full profits out of producing and selling such goods. Research conducted during the several decades since the first NPM-inspired reforms in social service provision were implemented has shown these “quasi-markets” to suffer from some of the typical market failures.

Among the “typical” failures, markets in social service delivery, including vocational training, seem to produce suboptimal results when they fail to generate enough market competition. This situation is so widespread that it almost constitutes the norm, rather than the exception. In turn, weak market competition negatively affects markets’ capacity to lower costs and improve service quality. In the United Kingdom, the privatization of social work in conditions of insufficient competition has led to a “race to the bottom” in which private service providers slashed costs by lowering the quality of the services and limiting the geographic availability of services (Carey, 2008). Typically, large urban centers and more economically developed areas tend to be reasonably well covered, whereas rural areas and less developed regions are less well serviced. The problem, of course, is that the latter usually have a greater need for services than the former. The same situation frequently obtains in vocational job training. In 2000, for example, whereas large relatively urban areas such as Indianapolis and Macomb-St. Clair were hosts to hundreds of providers, the whole state of Nebraska had only 100 vendors (O'Leary, Straits, & Wandner, 2004).

The second kind of market failure that tends to negatively affect social service delivery is information asymmetries. In other words, classic economics conceived of economic actors as fully rational and fully cognizant of market conditions. Research on decision-making and information processing has fundamentally challenged these assumptions, however, by showing that there are cognitive limits to humans' capacity to accumulate and analyze information. Economic agents, thus, possess not full, but "bounded" rationality. Because of cognitive and time constraints, decision makers often engage in "satisficing" by simplifying the available courses of action and choosing between them (Simon, 1997). The discovery of bounded rationality has had a deep impact on economics, forcing scholars to rethink the role of information asymmetries in market behavior. Most notably, if consumers lack complete knowledge of market conditions and may not be able to make optimal decisions, they will not always purchase the best goods and services at the lowest prices. In the market of vocational training, information asymmetries between buyers and sellers are even more pronounced because of the specific nature of the products offered for sale. First, the goods and services being exchanged are much more complex than those offered in typical market transactions; thus, consumers may lack the information that would enable them to select products that truly reflect their preferences (Lowery, 1998). Second, because of bounded rationality, many jobseekers may not know what kind of training is best suited for them and what qualifications are in demand in labor markets (Hipp & Warner, 2008). Third, training is a "post-experience good" (Ellis, 2001), meaning that it is often difficult for consumers to judge the quality of a training program even after they have enrolled in it.

Besides the difficulties stemming from market failures, various costly complexities arise from the new governance structure. More specifically, the producer/provider split has created numerous problems insofar as it creates additional costs associated with bargaining and monitoring. The initial expectation behind NPM-styled reforms was that, since much of the social service production will be transferred to private producers, government can be scaled down, resulting in budget savings. This expectation, however, never materialized in practice. Although the government lowered its involvement in direct service provision, it could not also transfer responsibility, i.e., taxpayers still hold the government, not private firms, accountable for the services they receive. In consequence, whereas the number of civil servants involved in service provision dropped, the number of public employees involved in contract bargaining, monitoring, and regulation sharply increased. In the United Kingdom, for example, overall civil service numbers fell by 30 percent between 1976 and 1995, but total staffing in public sector regulators dramatically increased by 90 percent during the same period (Hood, James, Jones, Scott, & Travers, 1999, p. 31). In the language of economics, NPM supporters failed to take into account the increased transaction costs incurred by the producer/provider split (Lowery, 1998). Widespread usage of training vouchers is especially likely to increase transaction costs since the ultimate decision to purchase training service rests with the customers. Therefore, government officials have to make sure that all the choices available to customers meet some basic standard of quality, however defined. In turn, the constant government monitoring may turn out to be a disincentive for private providers to participate, preventing their involvement in sufficient numbers to guarantee an adequate level of competition between providers.

Vouchers in Occupational Training in the U.S. – A Short History

The U.S. workforce development system has seen several major rounds of transformations in the last half century, owing to shifts in the larger political, economic, social, and cultural environment. Until recently, however, vouchers were not popular as a method for delivering vocational training or other types of employment services. In this section, we present a short history of job training in the U.S., focusing on the increasing role played by vouchers and other types of alternative provision in the field of publicly supported employment services.

In the U.S., ETA has been the primary government body charged with the provision of targeted training programs. From the standpoint of voucher usage, the history of targeted training can be divided into two main parts. In the early period, most training was delivered through contracts between public workforce entities and public and private training vendors. While some contracts were well targeted towards labor-market demand and quality service providers, many were not. A common problem was that by the time a contract was completed and enrollment began, demand for the occupation may have shifted. Further, training large numbers of individuals and having them enter the labor market at the same time created an over-supply that often contributed to poor employment outcomes. However, by the 1990s, as a more customer-focused workforce system began to emerge under Title III of the Job Training Partnership Act (JTPA), service delivery areas increasingly began to offer more individual choice to potential trainees. This trend culminated in passage of the Workforce Investment Act (WIA) in 1998. Before this event, the role of vouchers in U.S. vocational training was minimal. After passage of WIA, however, a significant part of targeted job training in the U.S. has been taking place through what is essentially a job-training voucher or Individual Training Accounts (ITA), as required under legislation. Let us examine these two periods in turn.

By the mid-1990s, the JTPA framework was facing mounting criticism. The central points of contention were its alleged overreliance on expensive training programs, its restrictive eligibility criteria, and its relative lack of flexibility in accommodating customers' preferences. The shift away from the JTPA took place within two larger trends occurring at the same time. One trend, already discussed, was the "reinventing government" agenda popularized by the U.S. administration at the time, which itself was rooted in a NPM framework. The second trend, which was taking place within the larger field of welfare policy, aimed to transform the U.S. welfare state based on a "work first" approach, though this trend has less to do with vouchers than does expanding customer choice. According to this logic, which inspired the passage of the

Personal Responsibility and Work Opportunity Reconciliation Act of 1996, federal aid to the poor is inefficient and creates perverse effects that perpetuate poverty. Therefore, the federal government must cut welfare spending and give individuals more freedom to choose solutions to their problems. In the realm of workforce development, the work-first approach translated into a reluctance to offer job training to individuals who might otherwise find a job faster, thus incurring less cost for society as a whole.

WIA abandoned the strict eligibility criteria imposed under JTPA and instituted universal access, regardless of one's income status (D'Amico & Salzman, 2004)[1]. However, in keeping with the cost saving principle featured in both NPM and work-first philosophies, WIA no longer offered training as a standard option. The workforce development model that WIA established features three levels of service, with training featured only as a tool of last resort and to be used only after the first two tiers, core and intensive services, have failed to generate employment.

From the perspective of the present study, the most important changes that WIA instituted had to do with the specific form of the training. Following the principles of increased marketization and customer choice, WIA introduced vouchers, rather than contract training, as the primary vehicle for job training. The only exceptions are for on-the-job training (OJT) and customized training. Customized training is allowed when community-based organizations with proven expertise in serving special populations exist that can provide training, and when there a dearth of training providers available for the ITA approach making this approach unfeasible (Barnow, 2008). WIA gives states considerable leeway in administering ITAs, for example, letting the states and local workforce areas choose key parameters, such as the monetary cap and duration. The law requires that all training program selections occur from a state-based eligible training provider list. Approved training providers have to submit annual reports about the outcomes of their students, and for the providers to remain eligible, these results have to meet or exceed performance requirements established by states and localities. The list, with its quality requirements, is intended to help solve the quality risks, bounded rationality, and asymmetrical information of a market-based approach underlying a voucher system.

The Effectiveness of Training Vouchers

Considering the increasing popularity that training vouchers have been enjoying with policymakers during the last two decades or so, some legitimate questions arise: Have these new training instruments (training vouchers) proved more effective than traditional ones? Have training vouchers led to more competition, lower prices, and better outcomes for customers? The

---

[1] In practice, the WIA adult program generally retains some income eligibility criteria based on self-sufficiency income levels.

existing evidence on the effectiveness of training vouchers is far from conclusive. Neither the available quantitative evaluations nor the qualitative evidence found in available research are able to paint an unequivocal picture. Before moving on to the meta-analysis, is it however useful to review the existing evidence. In what follows, we organize the available information around some of the better-known training voucher programs developed since the 1970s for which we have data. For each program, we aim to combine the existing quantitative and qualitative data into a comprehensive assessment.

The Seattle/Denver Income Experiments

The Seattle/Denver Income Experiments (SIME/DIME) were two large, randomized experiments conducted in the 1970s to test the feasibility and effects of a minimum guaranteed income on low-income families. As a part of the experiment, the Counseling and Education Subsidy Program, tested the effect of three nondirective counseling and training options on participants: counseling only; counseling and a 50 percent subsidy for any educational or training program; and counseling plus 100 percent subsidy. Employment counselors were trained to abstain from making any explicit recommendations to customers, leaving it up to them to make decisions. Thus, for all intents and purposes, this was a training voucher program. The working hypothesis was that the higher proportion of subsidies received, the higher the participation in the program and consequently, the higher the subsequent earnings.

The experiment results, however, were puzzling. Although subsidies led to increased participation, the higher amount of training either led to either no change in subsequent earnings or to lower earnings (although these were often not statistically significant). The experimenters concluded that the complete freedom of choice featured in the experiment was the main cause of failure to raise earnings because it allowed participants to form unrealistic expectations about their labor market prospects and to pursue overly ambitious plans (Dickinson & West, 1983). In other words, the evidence from the SIME/DIME experiments reiterated the potential of voucher programs to create market failures via information asymmetries and bounded rationality.

Career Management Accounts

The Career Management Account (CMA) demonstration project took place from 1995 to 1997 to assess the feasibility, impact, and cost-benefit ratios of vouchers for dislocated workers compared to the traditional approach used in the Title III of the JTPA. Key goals for the CMA demonstration encouraged local experimentation to foster increased customer choice, satisfaction, and flexibility in choosing and designing a reemployment and training program for dislocated workers. The experiment was conducted in 13 competitively chosen sites. The sites differed significantly in the treatments offered, the activities and services covered by the vouchers, and in other services and activities offered to participants. The study showed slightly better results for CMA participants compared to regular JTPA Title III customers. Thus, the CMA group had a positive termination rate that was four percentage points higher than the one recorded in the comparison group, and wages for CMA participants grew four percent faster than for the comparison group (Public Policy Associates, 1998). However, CMA programs ended up spending 74 percent more funds per participant compared to JTPA Title III participants, thus raising concerns regarding the program's eventual feasibility (Barnow, 2009). Many proponents of vouchers, as we have seen, emphasize the cost reduction feature of such programs. By contrast, the CMA demonstration seemed to suggest that voucher programs may not necessarily result in lower costs.

Individual Training Accounts (under WIA)

Several quasi-experimental studies of WIA have been recently published (Hollenbeck, Schroeder, King, & Huang, 2005), (Heinrich, Mueser, Troske, Jeon, & Kahvecioglu, 2009). These studies assess the effect of WIA as a whole on both adults and dislocated workers, but they also measure the distinct effect of WIA training. In addition, a substantial amount of qualitative research is available on WIA training. These enable us to offer a detailed picture of the program's operation and impacts.

Both quantitative studies cited above reach the conclusion that WIA training programs have a positive impact on both employment and earnings, although the size of this impact is larger in the Hollenbeck *et al* study compared to the one reported by Heinrich *et al*. One reason that could explain this difference is that the Hollenbeck *et al* study measures impacts for program exiters whereas Heinrich *et al* measures impacts beginning with program entry (Decker, 2010). However, neither study was a randomized experiment, and the experimental and control groups might have differed in ways unknown to the study designers, thus biasing the results. Another caveat in interpreting these results is that they do not capture only ITA effects but instead capture the total effect of WIA training programs. Although ITAs are the preferred method, they are not the only method to administer training under WIA. Thus, if the non-ITA training programs under WIA are much more effective than the ITAs, they could contribute a disproportionately large share of the positive effect, making it look like ITAs are effective when in fact they are not. But since the overwhelming bulk of training is carried out through ITAs, it is likely that these studies' results are close to voucher impacts.

Another important observation is that the way in which ITAs operate throughout the U.S. is highly variable, depending as it does on state and local policies. Although all local areas cap the amount that customers can spend using ITAs, the thresholds vary widely, ranging from under $2,000 to more than $7,500 (D'Amico & Salzman, 2004). Also variable is the percentage of customers who receive training, which ranges from 50 percent in some cases to almost zero in others. This variability makes it more difficult to estimate the overall impact of ITAs since, presumably, the general pattern observed nationwide may not describe accurately what is happening in various locales.

On the matter of program operation, most analysts agree that ITAs have enjoyed popularity and support from both customers and workforce system staff. Customers appreciated the greater flexibility that ITAs offered, whereas in most cases front-line staff managed to exert some guidance in customer's decision-making process. In other words, in most cases ITAs have not been "pure" vouchers but have orbited around a "guided choice" model (D'Amico & Salzman, 2004). In the literature, this model seems to be favored as the most likely one to avoid some of the potential market failures associated with vouchers (Ellis, 2001). However, it appears that not all of the potential downsides of vouchers were averted. One of the thorniest issues has turned out to be the selection and management of training providers. Often, local boards were unable to obtain reliable data about potential training providers. In some rural areas, local boards faced a different problem, namely the scarcity of providers (GAO, 2005). What is more, despite the inherent increase in competition, ITAs did not have a significant effect on training prices (O'Leary, Straits, & Wandner, 2004).

Recent Demonstration Projects: ITA, PRA, and CAA

The recent years have seen a rapid increase in the number of programs offering training vouchers. Together with the above-discussed ITAs, which are a permanent feature of the WIA framework, during the 2000s, ETA launched several other voucher programs, all of which were offered through demonstration efforts.

Thus, the ITA Experiment tested three different approaches to managing customer choice of training programs, ranging from a structured choice where customers had little flexibility and employment counselors could veto customers' choices to a "maximum choice" option (which was closest to a true voucher) where counseling services were optional and customers could select any providers on the Eligible Training Provider List (ETPL) (McConnell, et al., 2006). The experiment found that the three approaches differed little in their effects on employment rates, weeks worked, and earnings.

Another demonstration program titled, Personal Reemployment Accounts (PRAs), offered lump sum accounts of $3,000 that were fully managed by the customer and were valid for one year. The program allowed recipients to choose how and when to spend funds from their account and attempted to motivate the account holders by stipulating that they could receive 60 percent of any remaining balance in their PRA if they started working full-time by the end of the 13th week of unemployment insurance (UI) benefit receipt (Kirby, et. al., 2008). Lacking an experimental design, the PRA evaluation study could not compare the outcomes of PRA users with that of other training options. However, the study found that bonus-focused users entered employment quickly and had employment rates in each of the three follow-up quarters that were significantly higher than all other groups of PRA recipients.

A non-experimental[2] Career Advancement Account (CAA) demonstration project was launched in 2006 to examine a new training-voucher concept whereby workers were given an account of up to $3,000 per year for up to two years, which they could use to procure education or training services of their choice. Unlike ITA recipients, CAA recipients did not have to fulfill any pre-training prerequisite activities. In addition, the program did not require states and local areas to restrict access to those training providers on the WIA ETPL (Salzman, Wiegand, Leufgen, & Moazed, 2010). The CAA evaluation study found that CAA recipients were able to get into training promptly and with fewer requirements than their ITA counterparts.

Some of these demonstration projects, however, have been criticized for a number of reasons. Some analysts noted the "unconstrained" nature of the PRA vouchers and alleged that preference error on the customers' part, meaning customers' inability to make choices that reflect their true preferences, would be more likely under this type of arrangement (Hipp & Warner, 2008). Other analysts took issue with the cash incentive aspect featured in PRA, stating that the central reason behind offering a cash-out incentive—that jobless workers need such incentives to be convinced to return to work—is both demeaning and inutile since unemployed workers have already plenty of reasons to go back to work (Stettner & Chasanov, 2005).

The existing literature on the effectiveness of training vouchers offers a partial and often conflicted assessment. As expected, some of the studies find a positive relationship between vouchers and individual outcomes, while others either do not find a strong association or, as in the case of the SIME/DIME, the relationship seems to carry a negative sign. In addition, there is no way of knowing whether the variability of impacts found by the existing studies is produced by the vouchers or whether it has to do with some of the characteristics of the studies themselves, for example, whether the study was experimental or not, or whether it was part of a demonstration project or an already existing program. And even without these variations, we still have to account for the fact that the voucher programs that the studies evaluated were not identical. Some, like SIME/DIME and PRA, featured a "maximum choice" model whereby case managers did not attempt to advise customers; by contrast, the ITAs currently offered under WIA operate based on a "guided choice" model. To account for these various types of particularities, in the remainder of this report we will conduct a meta-analysis of the existing studies. The meta-analysis will attempt to estimate, based on the existing studies, whether the true population effect of training vouchers on employment and earnings is different from zero.

---

2 An experimental test of the CAA model, with an evaluation, was planned, but it never was implemented.

In addition, we will be able to estimate the moderating effect of study characteristics and of the various features of the voucher programs on voucher effectiveness.

III. METHODOLOGY AND DATA

Meta-analysis is a widely accepted method of summarizing the results of quantitative empirical studies within the behavioral, social, and medical sciences. The main reason for the existence of meta-analysis is that research reports studying a certain phenomenon frequently yield different results. Whereas traditionally researchers took this variability as a sign of immaturity of their sciences, meta-analysts treat existing studies as point estimators and try to use them to generate unbiased estimates of population parameters (Rosenthal, 1991). According to meta-analysis, there are two sources of study variability. The first source is sampling. Two studies may be identical in every way and still yield different results simply because of chance. On the other hand, results from research studies can vary because of study-related factors, such as the way in which the data were collected. Meta-analysis can be viewed as a technique akin to a standard survey analysis, but one where research reports, rather than individuals or organizations, are being surveyed. For each relevant research report, the researcher codes key information into a database and checks the data for errors and outliers. Finally, specific statistical techniques are applied to the data in order to describe the pattern of findings from the selected studies.

Calculating Effect Sizes

The main goal of meta-analysis is to estimate, based on the existing research studies, the "true" size of a phenomenon of interest. A central concept in meta-analysis is that of effect size. Because various studies on the same topic frequently use different ways of measuring their variables of interest, the researcher has to standardize the measurements so as to make them comparable across studies. The effect size is the statistic that allows such comparisons. The way in which the effect size is calculated depends on the kinds of data that the original studies reported. In general, however, research reports tend to present four kinds of effects:

- *Central tendency descriptions* present a characteristic of interest measured on a single sample of respondents. The distribution of the variable of interest is represented using some measure of central tendency, such as the mean, median, or mode.
- *Pre-Post contrasts* compare two measures of central tendency measured at two different time intervals for the same respondents.

- *Group contrasts* involve one or more variables measured on two or more groups of research participants; experiments frequently use this type of reporting.
- *Association between variables* examines the covariance of two variables of interest over the same group or sample.

All of the studies available for this meta-analysis, which is assessing the true impact of training voucher programs on customers' employment and earnings, are based on either experiments or group contrasts broadly understood. Research of this type is frequently meta-analyzed (Lipsey & Wilson, 2001) and fortunately, there are formulas readily available that can be used to compute effect sizes. Employment is usually measured as a dichotomous variable (employed/not employed) and reported as a proportion or percentage. In this case, the effect size is typically calculated as the difference in logged odds-ratios between groups. This measure was chosen because odds-ratios are a standard way of calculating effect sizes for dichotomous variables. However, odds-ratios are somewhat difficult to interpret directly because they are centered around one not zero. To circumvent this problem, logged odds-ratios are used, which have an approximately normal distribution with a mean of zero.

In addition to the effect size, the inverse of its variance (*w*) is calculated because it will be later used to weight the effect size for each study. The theory behind this technique is that the larger the sample used in a certain study, the more precise its measurements will tend to be. Similarly, the lower the standard deviation of a certain study, the more precise the measurement will be. In other words, the meta-analytic aggregate effect size is a weighted average of study-level effect sizes. The formulas are

$$ES_{OR} = \frac{p_a(1-p_c)}{p_c(1-p_a)}$$

$$ES_{LOR} = log_e(ES_{OR})$$

$$SE_{LOR} = \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$$

$$w = \frac{1}{SE_{LOR}^{2}},$$

where ES_{OR} is the effect size based on odds-ratios, ES_{LOR} is the effect size based on logged odds-ratios, *a* is the number of successful cases in the treatment group, *b* is the number of failure cases in the treatment group, *c* is the number of successful cases in the control group, *d* is the number of failure cases in the control group, p_a is the proportion of success in the treatment group, and p_c is the proportion of success in the control group.

For earnings, the same technique cannot be used to calculate effect sizes. Instead, the effect size is calculated as the standardized mean difference between groups. When the mean group difference is divided by a measure of pooled standard deviation, a statistic is obtained that can be used to compare results across all the studies:

$$ES = \frac{\overline{X}_t - \overline{X}_c}{s_p},$$

$$s_p = \sqrt{\frac{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}{(n_t - 1) + (n_c - 1)}}$$

where $\overline{X}_t$ is the mean in the treatment group, $\overline{X}_c$ is the mean in the control group, s_p is the pooled standard deviation, n_t is the size of the treatment group, and n_c is the size of the control group. However, this effect size formula has been shown to lead to upwardly biased estimates (Lipsey & Wilson, 2001). To correct for this, the following formula is used:

$$ES' = \left[1 - \frac{3}{4N - 9}\right] ES,$$

Consequently,

$$SE = \sqrt{\frac{n_t + n_c}{n_t n_c} + \frac{(ES')^2}{2(n_t + n_c)}},$$

$$w = \frac{1}{SE^2},$$

Where *N* is the total sample size (the sum of treatment and control group sizes), and *ES'* is the corrected effect size.

Several of the studies selected for analysis did not report means and standard deviations for earnings. Despite this impediment, it was still possible to compute effect sizes for some of these studies. If the study reported the results of a *t* test of independence between the two means, the following formula can transform the *t* statistic into the desired effect size:

$$ES = t\sqrt{\frac{n_t + n_c}{n_t n_c}},$$

Other studies in this meta-analysis do not report *t* values directly, but they make available information that allows this transformation. For example, the Rodgers and Hebbar (2004) study reports unadjusted coefficients and their standard errors from a regression of wages on a dummy variable that equaled 1 if the study participant was in the treatment group and 0 if they were in the comparison group. In this case, dividing the coefficient by its standard error yielded the desired *t*-value.

Many of the studies used for the current meta-analysis report a series of data that could be used to compute multiple effect sizes per study. Several studies, for example, report series of quarterly data, while others offer data broken down by state, demonstration site, or some other form of geographic location. Using several effect sizes per study would be incorrect because meta-analysis requires all the effect sizes to be statistically independent. It is possible to use serially correlated data, but this would require a researcher to know or estimate the covariance matrix between all of the statistically dependent measures. For this meta-analysis, the extra effort and time required by collecting this information, assuming it was possible to do so, would strain considerably the resources allocated for this report. In addition, using state-level data will increase the number of effect sizes; however, much of the resulting variation might have to do with mechanisms the present study is not adequately equipped to explain. Therefore, the main strategy for this report was to average serial data and aggregate geographically dispersed data to arrive at just one effect size per study.

The Data

For the present meta-analysis, seven studies were selected generating eight effect sizes for employment and nine effect sizes for earnings. This imbalance is the result of the SIME/DIME study only analyzing effects on earnings. Among the seven studies, only two–SIME/DIME and the ITA experiment (not to be confused with the regular ITA under WIA) were conducted along the lines of a classic randomized experiment. That only a third of the studies present experimental data constitutes a challenge to the conclusions that could be drawn based on the existing data. In evaluation research, randomized experiments are considered the "gold standard", meaning that estimates obtained using such methods are relatively bias-free. Not the same thing can be said about non-experimental research. Even careful quasi-experimental studies based on matching can provide biased estimators if the treatment and control groups differ in some characteristic unaccounted for in the matching design. This does not mean that conclusions obtained in the present meta-analysis are invalid, but it does mean that any and all generalizations have to be approached with considerable care. For this meta-analysis, an attempt will be made to detect bias induced by the research method by comparing mean effect sizes for experimental and non-experimental data. If the mean effect sizes were radically different, this would provide strong evidence of method-induced bias.

The number of studies selected for meta-analysis could be larger. A number of studies, however, had to be excluded because they did not contain sufficient information that would enable the calculation of effect sizes. The CAA study had to be excluded because it reports pre-post data, meaning data collected before and after the demonstration project for the same individuals, rather than group contrast data. In meta-analysis, it is good practice not to include both pre-post effect sizes and group comparisons in the same analysis because the former tend to be, on average, higher than the latter. Another interesting project, the evaluation of the Lifelong Learning Accounts (Public Policy Associates, 2006), could not be used because the study investigators have not yet released the impact evaluation data. In other cases, data could not be located for a number of smaller, state- and county-level demonstration projects, for example, the JTPA cases studied by Trutko and Barnow (1999). Finally, there was one instance, namely the ITAs under WIA, where there were two studies available. In accordance to the principle of not including more than one effect size per program to be evaluated, the Heinrich *et al* (2009) study was selected for this analysis rather than the Hollenbeck *et al* (2005) study. The reasoning behind this choice was that the Heinrich study reports impact data for the entire duration of the program, unlike the Hollenbeck which only analyzed program exiters. By taking advantage of having the entire duration of the program reflected in the data, the analysis should present a more adequate picture of the opportunity costs incurred by participants in the early phases of programs when they have to give up working so they can undergo training (Decker, 2010).

Some of the key technical details of each study are summarized in the table below. Most table headings do not require special explanations, but a few do. Under “Control Group”, the table lists, for each program, the composition of the control group. The major distinction here is that between control groups that were comprised of non-recipients of a certain program (for example, people who receive PRAs versus people who do not) and control groups that received a different sort of training than the one administered in the experimental group (say, JTPA Title III recipients versus CMA users). This distinction is potentially important because the counterfactual of the experiment is of a different kind, thus possibly leading to different results. Purely hypothetically, the effect size recorded in the recipient/non-recipient formats is expected to be larger because in the other scenario, control group participants receive at least some training. On the other hand, all the control groups that were denied participation in a certain training voucher program could access other federal or state level training programs. In any event, the distinction is important and deserves closer examination during the analysis.

Lastly, under the “Guided Choice” label, attempts were made to assess whether the program being evaluated was closer to a “pure” voucher model where customers have almost complete control over their choices or whether the program was closer to a “guided choice” model that assumed a close partnership between employment counselors and customers in making decisions. The existing theory tends to suggest that a guided choice model should fare better, but this is an assumption that must be tested against data[3].

[3] Follow-up survey data will be used to determine the longer-term impacts on employment outcomes for these participants under the study, *Individual Training Account Experiment Extension*.

Exhibit III-1:
Key Characteristics of Studies Included in Meta-Analysis

	Title	Method Used	Program Analyzed	Population	Control Group	Guided choice?	Source
SIME/DIME	The Seattle/Denver Experiments	Experimental	Defined by experiment; 2 states	Adults	Non-recipients	No	(Dickinson & West, 1983)
CMA	Career Management Accounts	Non-experimental	Demonstration Project; 13 states	Dislocated workers	Title III recipients	Yes	(Public Policy Associates, 1998) and personal correspondence
PRA	Personal Reemployment Accounts	Non-experimental	Demonstration Project; 8 states	Adults	Decliners (those who declined to participate)	No	(Kirby, Sullivan, Potamites, Kauff, Clary, & McGlew, 2008) and personal correspondence
ITG	Individual Training Grants	Quasi-experimental	Existing Program: 1 state	Dislocated workers	Non-recipients of training	Yes	(Rodgers & Hebbar, 2004)
NEG	Military Base National Emergency Grants	Quasi-experimental	Existing program: 3 sites	Adults	Non-recipients of training	Yes	(Needels, Bellotti, Dadgar, & Nicholson, 2006)
ITA	Individual Training Account Experiment	Experimental	Defined by experiment	Adults	Recipients of highly structured training programs	Yes for one experimental group, and no for another	(McConnell, et al., 2006) and personal correspondence
WIA	Individual Training Accounts	Quasi-experimental	Existing program: nationwide	Adults and dislocated workers	Non-recipients	Yes	(Heinrich, Mueser, Troske, Jeon, & Kahvecioglu, 2009)

IV. ANALYSIS AND FINDINGS

In this section, the analytic approach is described and results from the analyses are examined. When conducting meta-analyses, the first step is to examine the raw data to identify the variation in means or other indicators across the studies included, and this is discussed in the first section of this chapter. From these raw data, the effect sizes for each of the studies can be calculated, as well as the weight that should be given to each study based on its relative sample size. Once the effect sizes are computed, an examination must be conducted of whether the varying effects are drawn from a single sample or whether they appear to be taken from multiple independent samples. This test of homogeneity, described in the third section of this chapter, is critical for determining the appropriate statistical tests for assessing the results of the meta-analysis. Thus, the fourth section of the chapter identifies the distinction between fixed- and random-effects models and discusses why a random-effects model is appropriate for the present analysis. Finally, the chapter concludes with an analysis of whether certain features of the studies included in this analysis, or the programs on which the studies were conducted, yield greater or lesser estimates of the effects of voucher-based programs.

Descriptive Indicators

Exhibits IV-1 and IV-2 offer a general description of the raw data, including the source of the data and the main types of information used to compute effect sizes[4].

[4] For the ITA experiment, two sets of effect sizes are computed, labeled as "ITA Guided" and "ITA Max". These are the differences between the control group of those who received a structured training approach and two experimental groups: those who received a "guided choice" approach and those who received a "maximum choice" approach. Measures for adults and dislocated workers in the WIA evaluation are also reported separately, which was deemed feasible because the degree of co-enrollment in the two programs is low, allowing these two measures to be treated as statistically independent.

Exhibit IV-1:
Descriptive Indicators-Employment

	Treatment Group Size	Control Group Size	Number of measurements	Treatment Group Value (average proportion)	Control Group Value (average proportion)	Difference Between Treatment and Control Groups
CMA	3,808	17,244	1	0.78	0.71	0.07
PRA	2,173	348	3	0.51	0.55	-0.04
ITG	2459*	2480*	12	0.58	0.60	-0.02
NEG	1,216	982	1	0.57	0.50	0.07
ITA Guided	1309	1322	5	0.54	0.56	-0.02
ITA Max	1302	1322	5	0.54	0.56	-0.02
WIA Adult	17,651	17,651	1	n/a**	n/a**	0.03
WIA Dislocated	13,331	13,331	1	n/a**	n/a**	0.02

*During the last quarter of measurement, sample sizes were 2424 and 2180, respectively

** Heinrich *et al* do not report the proportions within the treatment and control groups but the difference between them and the standard error of the difference, separately for men and women. The average difference was 0.03 for the adult sample and 0.02 in the dislocated worker sample (calculated by averaging indicators between men and women).

A quick visual inspection of the table above reveals that there is a wide discrepancy between various studies regarding both the size and difference between the experimental and control groups employment rates. A similar pattern is observed in the case of earnings, as well, as shown in Exhibit IV-2.

Exhibit IV-2:
Descriptive Indicators-Earnings

	Treatment Group Size	Control Group Size	Treatment Group Average*	Control Group Average*	T-value	Regression Coefficient****
CMA**	3,808	17,244	11.33 (11.0[5])	11.65 (11.00)		
PRA	2173	348	6,070 (6,571.65)	4813 (5,563.71)		

[5] The standard deviation of the CMA data for earnings was not available at the time of the analysis so it had to be guessed.

	Treatment Group Size	Control Group Size	Treatment Group Average*	Control Group Average*	T-value	Regression Coefficient****
ITG	2459*	2480***				-0.05 (0.02)
NEG	805	615			2.23	
ITA Guided	1309	1322	3291.50 (4562.63)	3444.08 (4667.79)		
ITA Max			3107.00 (4294.89)	3444.08 (4667.79)		
WIA Adult	17,651	17,651			1.04	
WIA Dislocated	13,331	13,331			2.95	
SIME/DIME	391	1041			-0.96	

* Standard deviation in parentheses
** Hourly wages
*** During the last quarter of measurement, sample sizes were 2424 and 2180, respectively
*** Standard error in parentheses

An important observation is that the SIME/DIME data used for this meta-analysis are for men only. This was a particularly difficult choice to make, but the way in which the SIME/DIME experiment was designed and reported (the treatment groups were husbands, wives, and single female household heads) made it very difficult to average the data. Since training outcomes of husbands and wives are not statistically independent, averaging the standard errors of their regression coefficients would have been a serious potential source of bias. Therefore, the data on men was chosen because at the time of the experiment, men had a much higher probability to be hired than did women.

The raw data described above can be used to compute the effect size associated with each study, along with the inverse variance which can be used to weight the data. These are shown in Exhibit IV-3.

Exhibit IV-3:
Effect Sizes and Weights for Employment

Study	Effect Size (ES)	Inverse Variance of ES (w)
CMA	0.35	558.40
PRA	-0.16	73.85
ITG	-0.12	277.39

	Effect Size (ES)	Inverse Variance of ES (w)
NEG	0.31	134.54
ITA Guided Choice	-0.07	149.08
ITA Maximum Choice	-0.08	144.90
WIA Adult	0.02	4412.62
WIA Dislocated	0.05	3331.60

As expected, there is a wide discrepancy among studies in terms of both the sign of the relationship between training vouchers and employment and its size. The CMA and NEG studies show the largest positive impact on employment, whereas PRA and ITG indicate the largest negative impact. Overall, the WIA-ITA indicators have the largest weight, given the large sample sizes for those studies; and given the effect size is positive, one might expect the overall mean effect size to be positive.

Exhibit IV-4 displays similar calculations for earnings in each of the studies.

Exhibit IV-4:
Effect Sizes and Weights for Earnings

	Effect Size (ES)	Inverse Variance of ES (w)
CMA	-0.03	3118.99
PRA	0.21	299.05
ITG	-0.06	1226.34
NEG	0.12	348.04
ITA Guided Choice	-0.03	657.64
ITA Maximum Choice	-0.07	677.24
WIA Adult	0.02	4412.61
WIA Dislocated	0.05	3331.64
SIME/DIME	-0.07	390.51

As can be seen in this exhibit, data for earnings displays virtually the same amount of variation between studies, but the overall mean effect size is harder to discern. Whereas the WIA effect sizes continue to exert a powerful pull in the positive direction, many other studies report a negative relationship, some of them with significant pull effect.

Test of Homogeneity

An important question in meta-analysis is whether the variation observed in the set of effect sizes is caused simply by sampling error or if there are systematic sources of error that combine with sampling error to produce the variation (Hedges & Olkin, 1985). The distribution of the effect sizes is considered homogenous, and thus caused by sampling error alone, if the dispersion of effect sizes around their mean is no greater than that expected from sampling error alone. In such cases, the weighted mean effect size is a good indicator of population mean. If, however, the dispersion of effect sizes is larger than what would be expected from standard error, it follows that effect sizes cannot be used to estimate a common population mean. The test of homogeneity commonly used in meta-analysis is based on the Q-statistic (Cochran, 1954), which is distributed as a chi-square test with *k*-1 degrees of freedom, where k represents the number of effect sizes, calculated using the following formula:

$$Q = \sum w_i \left(ES_i - \overline{ES}\right)^2$$

Results using this formula applied to the voucher data are shown in Exhibit IV-5.

Exhibit IV-5:
Results of the Q-Test of Homogeneity

	Test value	Degrees of Freedom (df)	Expected Value at α=0.05
Employment	79.32	7	14.07
Earnings	41.03	8	15.51

Because the test value is considerably higher than the expected value, in both the case of employment and earnings the Q-test provides convincing evidence that the set of effect sizes observed for these voucher studies is not internally homogenous. This, therefore, implies that the variation observed in these effect sizes is caused by sources other than solely sampling error.

Random-Effects Models

Since the available evidence indicates that there is significant heterogeneity in the set of effect sizes for both employment and earnings, a random-effects model, rather than a fixed-effects model, is more appropriate (Hedges & Vevea, 1998) to estimate the overall impact of voucher training on employment and earnings in the population. This is because a fixed-effects model assumes that each effect size is attempting to approximate a fixed population mean with a calculable sampling error. A random-effects model, by contrast, assumes that there are multiple effect sizes in the population, each of which has its own sampling error. A random-effects analysis can estimate the range of all possible values of the multiple population-level effect sizes, and is thus most appropriate when there is significant heterogeneity in the set of effect sizes in the sample.[6]

A random effects model weights each study by the inverse of the sampling variance plus a constant that represents the variability across the population effects, as shown in the following:

$$w_i = \frac{1}{se_i^2 + \hat{v}_\theta}$$

$$\hat{v}_\theta = \frac{Q_t - k - 1}{\sum w - \left(\frac{\sum w^2}{\sum w}\right)}$$

where w_i are the recalculated weights, $\hat{v}_\theta$ is the random variance component, Q_t represents the Q-value calculated earlier, *k* represents the number of effect sizes, and *w* are the initial weights (Lipsey & Wilson, 2001).

Results derived from a random-effects model of the impact of voucher training on employment are shown in Exhibit IV-6.

Exhibit IV-6:
Meta-Analytic Results-Employment

Mean ES	RE Variance Component	95% Credibility Interval
1.05	0.013	[0.829, 1.272]

Given the computed random-effects variance of approximately 0.013, the 95% credibility interval (the equivalent of the confidence interval for a random-effects analysis (Hunter & Schmidt, 2004)[7] is thus [0.829, 1.272]. Because this confidence interval contains 1, which signifies that employment is equally likely in both treatment and control groups, it cannot be

[6] Precisely for this reason, however, the confidence interval calculated using a random-effects model is usually considerably larger than for fixed-effects estimations, which is why fixed-effects models are preferable when the effect sizes are homogenous.

[7] The credibility interval is calculated as the point estimator of the mean effect size (1.0504) plus and minus the standard deviation of the infinite-sample effect size (the square root of v multiplied by 1.96).

determined that there is a positive or a negative relationship between voucher training programs and employment in the population of studies included in this analysis.[8]

Results using earnings as the outcome of interest mirror those obtained for the relationship between vouchers and employment, and are shown in Exhibit IV-7.

Exhibit IV-7:
Meta-Analytic Results-Earnings

Mean ES	RE Variance Component	Credibility Interval
0.006	0.003	[-0.099, 0.111]

This variance component tells us that the standard deviation of the population of effect size values, which is calculated as the square root of the variance component, is about .054. With a random effects (RE) mean of 0.01, the entire population of effect sizes runs from about -0.09 to 0.11 (assuming normality of the population of effects). As can be seen in this exhibit, the random-effects mean does not differ significantly from 0, as the credibility interval includes 0. Thus, although there is a positive mean effect size, the limited statistical power of the model may limit it from showing a significant relationship if one exists at all.

Categorical Analysis

The final step in the analysis is to estimate the importance of study-level factors, or otherwise known as mediator variables, in generating the outcomes observed. This should be particularly instructive in the present case given the strong evidence of heterogeneity in the data described above. Also, from a policy-making perspective, it would be useful to know if specific features of the programs being evaluated generate a larger increase in outcomes. The ideal way to conduct this analysis would be using a weighted least squares regression in order to study the combined effect of moderator variables on employment and earnings. Unfortunately, however, the limited number of effect sizes makes this approach unfeasible. As a result, instead of testing the effect of all moderator variables at once, a suitable approach is to study the effect of these variables one

8 The fact that the point estimator is greater than 1 suggests that the limited number of studies available for the meta-analysis limits the power of the model. As noted above, a random effects model yields a larger confidence interval. Thus, this larger interval, in combination with the small number of studies available and included in the analyses, makes it impossible to find an effect of vouchers on employment, even though one exists. Under this hypothesis, a larger sample of studies would increase the power of the model and perhaps enable one to conclude that an effect does exist.

by one using a modified form of ANOVA[9]. For each moderator variable, two types of modified ANOVA were conducted, a fixed-effects model and a mixed-effects model.

The fixed effects model works from the assumption that some or all of the excess variability in effect sizes is systematic and can be modeled with intervening (moderator) variables. This procedure partitions the total variance into a portion which is said to be explained by the categorical variable (Q_B) and a residual portion (Q_W). Similar to ANOVA, Q_B is an index of variability between the group means and Q_W is an index of variability between groups. If Q_B is statistically significant, then it can be inferred that the categorical variable explains more variation than what would be expected from sampling error alone (Lipsey & Wilson, 2001).

The mixed-effects model assumes that the variability between effect sizes is due to sampling error plus variability in the population of effects. Therefore, part of the observed variance in effect sizes is explained by fixed-effects, but part of the variance remaining is unexplained. The unexplained part (not accounted for by sampling error or other artifacts) is remaining random-effects variance. The analysis proceeds by computing the random effects variance component and then re-estimating the between and within variance (Lipsey & Wilson, 2001).

Given the high level of heterogeneity in the sample of effect sizes among voucher-based programs, mixed-effects ANOVA models are the best mechanism for examining mediator variables (Hunter & Schmidt, 2004). However, since the small sample of effect sizes also means the models have low statistical power, the analysis first examined fixed-effects models to check for potential moderator effects, even though these effects may prove insignificant in the mixed-effects models. In the table below only the mediator variables that showed potential for statistical significance in the fixed-effects models are displayed. However, none of them also proved significant in the mixed-effects models. The effect of four categorical variables was examined: *experimental* (1 if the study was a randomized experiment, 0 if it was not); *guided* (1 if the voucher was close to the "guided choice" model, 0 if it was not); *adult* (1 if the universe of the study was low-income adults, 0 if the universe was dislocated workers); and *demonstration* (1 if the study was an evaluation of a demonstration project, 0 if not). The results are displayed in Exhibit IV-8.

---

9 This modified form uses the same algorithm to calculate the statistics but utilizes a modified algorithm for computing standard errors (Lipsey & Wilson, 2001).

Exhibit IV-8: Categorical Analysis-Employment

	Experimental	Guided	Adult
Fixed-Effects ANOVA			
Between Group Variance	4.414	5.142	8.797
Significance	0.036	0.023	0.003
Group Means	1=-0.076	1=0.049	1=0.016
	0=0.049	0=-0.107	0=0.079
Mixed Effects ANOVA			
Between Group Variance	2.032	2.951	0.478
Significance	0.154	0.086	0.489

In the fixed-effects models, three categorical variables appeared to have a potential effect on the relationship between training vouchers and employment. Specifically, studies employing quasi-experimental methods tended to show greater effects on employment than those employing purely experimental methods. Similarly, studies of vouchers providing guidance from employment counselors to the customer tended to show greater effects on employment, and studies of dislocated workers tended to show greater effects than studies involving other adults.

Examining results using the mixed-effects models, however, only one of these, i.e., the degree of conformity to the guided choice model, seems to be close to having a significant impact. The results seem to indicate that the voucher programs that are run using a guided choice model, in which customers benefit from significant input from employment counselors, are associated with higher employment. This finding supports the theoretical notion that voucher programs run on purely free choice basis tend be less efficient than those that compensate for potential market failures, such as customers' lack of complete information or access to all potential training programs (Ellis, 2001). Also, although not significant, there is a gap in the observed effects on employment between experimental and quasi-experimental evaluations. Specifically, results from experimental studies were less likely to show improvements in employment compared to other types of methods. Given that experimental studies are considered to be the gold standard and should produce unbiased estimates of effects, this suggests, albeit in a limited way given there is no statistical difference, that non-experimental methods generate some upward bias when considering the effects on employment of voucher-based programs.

Similar analyses for earnings are displayed in Exhibit IV-9.

Exhibit IV-9:
Categorical Analysis-Earnings

	Experimental	Demonstration
Fixed-Effects ANOVA		
Between Group Variance	7.252	6.985
Significance	0.007	0.008
Group Means	1=-0.055	1=-0.024
	0=0.014	0=0.022
Mixed Effects ANOVA		
Between Group Variance	3.698	0.565
Significance	0.055	0.453

Although the fixed-effects analysis indicates that both the experimental and demonstration mediator variables appear to have a significant impact, it is in the mixed-effects models, only the effects of the method (i.e., experimental vs. quasi-experimental) approach significance. As with the employment outcomes, the group mean on earnings for the studies conducted as experiments is significantly lower than that for non-experimental studies, which reinforces the notion that the type of method used in the evaluation of voucher programs seems to play a significant role and provides further evidence that non-experimental studies of voucher-based programs may yield an upward bias in the estimate of the programs' effects on earnings.

V. CONCLUSION

The conclusions that can be drawn based on the available data are tentative. The main challenge that this project faced was the relative small number of available studies, which translates into a lack of statistical precision in detecting meta-analytic effects. Simply put, there are not enough studies which evaluate the effect of vouchers on employment and earnings that would enable the researchers to predict population data with a high degree of accuracy. Along with data insufficiency, the project also faced problems insofar as many of the available studies did not contain enough information to allow the correct computation of effect sizes. Although in most cases the study's original authors were contacted, and they graciously offered to provide some of the required data, in some other cases the data were simply too dated (or, on the other hand, too new such that they had not yet been released for use by the general public), and thus the present study had to work with the available information. Thus, when coding and transforming the data, a number of assumptions may have altered the final result. As an example, the effect sizes for earnings were not measured in the same way across each of the studies. Some studies published mean quarterly earnings, while others only made available the mean hourly wages, and still others published t-values or regression coefficients. This difference in measurement may lead to important measurement errors. For instance, high hourly wages may not translate into high quarterly wages if study subjects work on a part-time schedule.

Nonetheless, the present report did uncover some potentially important patterns. First, an analysis of homogeneity confirmed that the studies included in this meta-analysis did not come from a unique sample. This confirms that the relationship between training vouchers and individual outcomes is not a simple one that applies across all samples and types of studies. Using categorical analysis (both fixed-effects and mixed-model effects) demonstrated that this connection seems to be mediated by both the way in which evaluation studies are conducted, e.g., randomized experiments seem to be less likely to show an effect compared to non-experimental ones, and also by some features of the various voucher programs. The most salient among the program features was whether the programs were run based on the guided choice model, a feature that seemed to boost the chances for customers to find employment. However, the role of the guided choice model seemed to be insignificant as a predictor of higher earnings. Finally, dislocated workers may have a better chance of finding employment while enrolled in a voucher training program compared to adult, low-income workers, but again this relationship does not hold for earnings.

Overall, combining the studies included in this analysis suggests that using vouchers for employment training programs may have a positive, albeit quite small effect on both employment and earnings. However, results using random-effects models, which are more appropriate given the variation among the studies included, also yield larger confidence intervals that leave open the possibility that the actual effects of these voucher-based programs may be zero (or even negative). It is clear that the small number of effect sizes in our sample (due primarily to the limited number of voucher-based programs that could be included in this study) substantially weakened the statistical power of the tests, thus artificially inflating the credibility intervals for the mean effect sizes which results in far less precise conclusions. As more evaluation studies become available, the precision of these meta-analytic results may significantly improve, and subsequent examinations may well be able to more precisely identify the specific effects of voucher-based programs on employment and earnings.

VI. BIBLIOGRAPHY

Barnow, B. S. (2008). Vouchers in U.S. Vocational Training Programs: An Overview of What We Have Learned. *Consumer Choice in the Provision of Public Services: Vouchers and Related Delivery Mechanisms* (pp. 1-29). Nuremberg, Germany: Institute for Employment Research.

Barnow, B. S. (2009). Vouchers in U.S. vocational training programs: an overview of what we have learned. *Zeitschrift für ArbeitsmarktForschung , 42* (1), 71-84.

Burwick, A., & Kirby, G. (2007). *Using Vouchers to Deliver Social Services: Learning from the Goals, Uses, and Key Elements of Existing Federal Voucher Programs.* Washington, DC: Mathematica Policy Research.

Carey, M. (2008). Everything Must Go? The Privatization of State Social Work. *British Journal of Social Work , 38* (5), 918-935.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics , 10*, 101-129.

D'Amico, R., & Salzman, J. (2004). Implementation Issues in Delivering Training Services to Adults Under WIA. In C. J. O'Leary, R. A. Straits, & S. A. Wandner, *Job Training Policy in the United States* (pp. 101-134). Kalamazoo, MI: W.E. Upjohn Institute.

Decker, P. T. (2010). Ten Years of the Workforce Investment Act (WIA): Interpreting the Research on WIA and Related Programs. *What the European Social Fund Can Learn from the WIA Experience* (pp. 1-38). European Commission Directorate-General for Employment, Social Affairs and Equal Opportunity; University of Maryland School of Public Policy.

Dickinson, K. P., & West, R. W. (1983). *Final Report of the Seattle-Denver Income Maintenance Experiment.* Menlo Park, CA: SRI International.

Ellis, N. (2001). Individual Training Accounts under the Workforce Investment Act of 1998: Is Choice a Good Thing? *Gerorgetown Journal on Poverty Law & Policy , 8* (1), 235-256.

Freedland, M. (2001). The Marketization of Public Services. In C. Crouch, K. Eder, & D. Tambini, *Citizenship, markets, and the State* (pp. 90-110). Oxford, UK: Oxford University Press.

GAO. (2005). *Substantial Funds Are Used for Training, but Little Is Known Nationally about Training Outcomes.* Washington, DC: Government Accountability Office.

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

Hedges, L., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods , 3* (4), 486-504.

Heinrich, C. J., Mueser, P. R., Troske, K. R., Jeon, K.-S., & Kahvecioglu, D. C. (2009). *New Estimates of Public Employment and Training Program Net Impacts: A Nonexperimental Evaluation of the Workforce Investment Act Program.* Columbia, MO: University of Missouri.

Hipp, L., & Warner, M. E. (2008). Market Forces for the Unemployed? Training Vouchers in Germany and the USA. *Social Policy & Administration , 42* (1), 77-101.

Hollenbeck, K., Schroeder, D., King, C. T., & Huang, W.-J. (2005). *Net Impact Estimates for Services Provided through the Workforce Investment Act.* Washington, DC: U.S. Department of Labor, Employment and Training Administration.

Hood, C. (1995). The “New Public Management” in the 1980s: Variations on a Theme. *Accounting, Organizations and Society , 20* (2/3), 93-109.

Hood, C., James, O., Jones, G., Scott, C., & Travers, T. (1999). *Regulation Inside Government Waste-Watchers, Quality Police, and Sleazebusters.* Oxford, UK: Oxford University Press.

Hunter, J., & Schmidt, F. (2004). *Methods of meta-analysis: Correcting error and bias in research findings.* London: Sage.

Kettl, D. J. (2002). *The transformation of governance. Public administration for twenty-first century America.* Baltimore: Johns Hopkins University Press.

Kirby, G., Sullivan, M., Potamites, E., Kauff, J., Clary, E., & McGlew, C. (2008). *Responses to Personal Reemployment Accounts (PRAs): Findings From the Demonstration States.* Washington, DC: Mathematica Policy Research.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-Analysis.* Thousand Oaks, CA: Sage.

Lowery, D. (1998). Consumer Sovereignty and Quasi-Market Failure. *Journal of Public Administration Research and Theory , 8* (2), 137-172.

McConnell, S., Stuart, E., Fortson, K., Decker, P., Perez-Johnson, I., Harris, B., et al. (2006). *Managing Customers' Training Choices: Findings From the Individual Training Account Experiment.* Washington, DC: Mathematica Policy Research.

Needels, K., Bellotti, J., Dadgar, M., & Nicholson, W. (2006). *Evaluation of the Military Base National Emergency Grants - Final report.* Princeton, NJ: Mathematica Policy research.

Niskanen, W. (2007). *Bureaucracy and Representative Government.* Chicago: Aldine Transaction.

O'Leary, C. J., Straits, R. A., & Wandner, S. A. (2004). Public Job Training: Experience and prospects. In C. J. O'Leary, R. A. Straits, & S. A. Wandner, *Job Training Policy in the United States* (pp. 289-310). kalamazoo, MI: W.E. Upjohn Institute.

Osborne, D., & Gaebler, T. (1993). *Reinventing government: how the entrepreneurial spirit is transforming the public sector.* New York: Plume.

Ostrom, V. (1987). *The Political Theory of a Compound Republic: Designing the American Experiment.* Lincoln: University of Nebraska Press.

Pollitt, C., Bathgate, K., Caulfield, J., Smullen, A., & Talbot, C. (2001). Agency Fever? Analysis of an International Policy Fashion. *Journal of Comparative Policy Analysis , 3* (3), 271-290.

Public Policy Associates. (1998). *Career Management Account Demonstration Project Evaluation Final Report.* Lansing, MI: Public Policy Associates.

Public Policy Associates. (2006). *Evaluation of the Lifelong Learning Accounts Demonstration: Second Interim Report.* Lansing, MI: Public Policy Associates.

Rodgers, W. M., & Hebbar, l. (2004). The Impact of Training Vouchers on the Re-employment and Earnings of Dislocated Workers. *Annual Research Conference of the Association for Public Policy and Management*, (pp. 1-35). Atlanta, GA.

Rosenthal, R. (1991). *Meta-Analytic Procedures for Social research.* Thousand Oaks, CA: Sage.

Salzman, J., Wiegand, A., Leufgen, J., & Moazed, T. (2010). *Evaluation of the Career Advancement Account Demonstration.* Oakland, CA: Social Policy Research Associates.

Savoie, D. J. (1994). *Thatcher, Reagan, Mulroney: In Search of a New Bureaucracy.* Pittsburgh, PA and London: University of Pittsburgh Press.

Shaw, K. M., & Rab, S. (2003). Market Rhetoric Versus Reality in Policy and Practice: The Workforce Investment Act and Access to Community College Education and Training. *The ANNALS of the American Academy of Political and Social Science , 586* (1), 172-193.

Simon, H. (1997). *Administrative Behavior* (4th ed.). New York: The Free Press.

Somers, M. R., & Block, F. (2005). From Poverty to Perversity: Ideas, Markets, and Institutions over 200 Years of Welfare Debate. *American Sociological Review , 70* (2), 260-287.

Stettner, A., & Chasanov, A. (2005). *Setting the Wrong Course: Personal Reemployment Accounts fail to address the real needs of jobless workers.* New York: National Employment Law Project.

Suleiman, E. (2003). *Dismantling Democratic States.* Princeton, NJ: Princeton University Press.

Trutko, J., & Barnow, B. S. (1999). *Vouchers under JTPA: Lessons for Implementation of The Workforce Investment Act.* Arlington, VA: James Bell Associates.

Valkama, P., & Bailey, S. J. (2001). Vouchers As an Alternative Public Sector Funding System. *Public Policy and Administration , 16* (1), 32-58.