Casework Job Design and Client Outcomes in Welfare-to-Work Offices

Carolyn J. Hill
Assistant Professor
Georgetown Public Policy Institute
3520 Prospect Street NW, 4th Floor
Washington, DC 20007
cjh34@georgetown.edu

March 16, 2005
ABSTRACT

Differences in performance across different locations of a human service program may be driven by client, managerial, organizational, policy, or environmental characteristics. While many of these factors are outside the control of local managers, other factors may be open to influence by local discretion and may have independent effects on performance. One issue facing local managers is how to divide job tasks among frontline staff, but little evidence is available regarding whether job design is related to performance. In this article, I examine the relationships between different casework task configurations and welfare-to-work office performance. Controlling for a number of client and office characteristics, I find that clients’ average earnings are higher over a two-year period in offices that primarily use unified case management, and in offices with a specialist who develops job opportunities. I find no effects on earnings in offices that use other kinds of specialists, and no effects of unified case management or specialists on welfare benefit receipt in the two-year period. Overall, the findings suggest that local managerial decisions regarding job design help explain the variation in performance across offices, and suggest a possible lever through which performance can be improved.
Casework Job Design and Client Outcomes in Welfare-to-Work Offices

INTRODUCTION

The Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) of 1996 imposed new work expectations and time-limited benefits on welfare recipients, and also gave states flexibility in administering these programs. Research has shown that welfare-to-work programs can be effective for some recipients, and that effectiveness varies across local sites that deliver program services (Friedlander and Burtless 1995; Gueron and Pauly 1991; Heckman, LaLonde, and Smith 1999). This variation in performance across sites may be attributed to differences in clients served; in program organization or management; or other economic, policy, or political factors.

To develop effective programs that raise welfare recipients’ long-term earnings, policy designers and managers need more and better information about why certain programs or offices are more successful than others, what roles or practices of frontline workers are associated with greater effectiveness, and how managers can influence program success, accounting for observable characteristics of clients and environments that vary across sites. In other words, what role does governance play in program effectiveness?

In this article, I focus on the division of tasks, or job design, among frontline staff in welfare-to-work offices. Guidance for local managers on job design has emphasized the various advantages and disadvantages of different configurations, but little evidence is available that links these configurations to performance. It is possible that job design has negligible effects on client outcomes: client characteristics and other features of the office or local environment—such as caseload size, caseworker personal characteristics, managerial actions, or service
availability—may drive variation in outcomes. Or, it is possible that casework job design emerges as an important element of service technology, even after controlling for other factors.

I develop hypotheses regarding the relationship between job design and client outcomes, then test these hypotheses using administrative and survey data collected by MDRC for evaluations of welfare-to-work programs from the late 1980s to mid 1990s. The evaluations were not designed to answer the specific questions examined in the current analysis, which attempts to glean additional insight into governance and management questions using the natural variation across offices in casework job design and in client outcomes. My findings indicate that some casework configurations—unified case management or staff specialists who develop job contacts for clients—are more effective when office performance is measured by client earnings, but no differences in effectiveness are found when performance is measured by AFDC receipt. Many characteristics of clients and offices are controlled for in the analysis, though it is possible that factors associated both with selection of casework configuration and with office performance are not included in the model.

The article proceeds as follows. First, I discuss different possible configurations of casework and develop hypotheses. Next, I describe the data, methods, and measures; and discuss results. In the conclusion, I discuss the role that this type of research can play in understanding the links between governance and performance.

CASEWORK JOB DESIGN AND HYPOTHESES

The division of casework tasks among line staff is one aspect of program design that may affect clients’ experiences in a welfare-to-work program and influence their longer-term success. Even in welfare-to-work programs with a strong state role, local program managers have discretion over how to assign staff to different casework tasks (e.g., Kemple and Haimson 1994).
Casework job design structures the interactions between caseworkers and clients, introducing boundaries within which caseworkers exercise discretion. Line staff in a welfare-to-work office are “policy makers” whose interactions with clients can determine the “range of behavioral actions from which clients may choose their responses” (Lipsky 1980, 61).

Welfare-to-work casework comprises four core tasks: (1) assessing the client’s needs for services, (2) working with the client to develop an employability plan, (3) arranging and coordinating services, and (4) monitoring the client’s progress (Hagen and Lurie 1994a; Rose and Moore 1995; Rothman 1991). Other than these core tasks, other casework activities may include developing job contacts with local employers, conducting services (e.g., job club), providing personal counseling to clients, and other activities. Caseworkers may perform these activities in addition to or instead of the core casework tasks.

One aspect of casework job design that has received much attention concerns whether welfare-to-work casework is combined with income maintenance and sanctioning activities. Some evidence about these designs is now available: A random assignment study at the Columbus, Ohio site of the National Evaluation of Welfare-to-Work Strategies examined whether having caseworkers responsible for both welfare-to-work casework and income maintenance (the “integrated model”) was more effective than separating welfare-to-work casework and income maintenance (the “traditional model”) (Scrivener and Walter 2001). Over a three-year period, clients served in offices that used the integrated model received more personalized attention, were monitored more closely, were more involved in activities such as orientation and training, and reduced their AFDC receipt compared with clients served in offices that used the traditional casework model. Overall, earnings and employment outcomes were no different across the two casework models; however, among the subgroup of clients who lacked a
high school degree or GED (at the time of random assignment), the integrated model increased earnings and employment and decreased welfare receipt, compared to the traditional model (Scrivener and Walter 2001).

An aspect of job design that has not been systematically examined is the link between welfare-to-work core casework tasks and program performance. Doolittle and Riccio (1992) identified this gap in the literature over a decade ago, and empirical evidence is still lacking. The core tasks can be divided among welfare-to-work frontline staff in one of three basic ways. With **unified case management**, a caseworker is involved in all four aspects (assessing the client’s need for services, working with the client to develop an employability plan, arranging and coordinating services, and monitoring client progress) of the client’s activities in the program.\(^4\)

Alternatively, with **sequential (or categorical) casework**, a caseworker is involved in a subset of these activities. A third model of **coordinated services** combines the unified case management and sequential casework models, where caseworkers specialize but maintain communication and work as a team to address a client’s needs (APWA 1992; Doolittle and Riccio 1992; Hagen and Lurie 1994a, 1994b).\(^5\)

These different casework models hold potential advantages as well as drawbacks (Brown 1997; Doolittle and Riccio 1992; Hagen and Lurie 1994a, 1994b). With unified case management, a caseworker is the primary contact with the client. She may obtain a holistic view of the client’s circumstances, and employ the core case management activities as an integrated whole in meeting the client’s needs, acting as a broker of services. Duplication of effort can occur, because different case managers in the same office are performing the same kinds of tasks (e.g., locating child care services or contacting service providers to monitor clients).
The advantages and disadvantages of unified case management are reversed in sequential casework, where staff specialize in particular stages of the welfare-to-work program. Clients can benefit from the expertise that a caseworker develops by specializing, but line staff who specialize in certain tasks lack direct control or knowledge of other specialties that may affect the client and thus “avoid seeing their work as a whole” (Lipsky 1980, 147).

While the effectiveness of integrating welfare-to-work casework and income maintenance has been examined using random assignment methods, the effectiveness of other casework configurations has not been analyzed empirically. Further, no clear predictions regarding the relationships between casework configurations and client outcomes emerge from prior characterizations of different casework models. The hypotheses and empirical analysis presented below aim to inform the decisions of local managers and program designers regarding how best to configure casework tasks in local offices.

Hypotheses

In the current study, I analyze existing data sources from the 1980s and 1990s (thus, it is not possible to pair the quantitative analysis with contemporaneous qualitative data collection). I employ theory as a heuristic to generate testable hypotheses. While the theoretical literature on job design and performance is not extensive, relevant frames of reference can be found in multitask principal-agent theory (Holmstrom and Milgrom 1991) and organization theory on service technology characteristics (Hasenfeld 1983; Thompson 1967).

Multitask principal-agent theory suggests that constraints may be effective for directing the work of agents toward managerial or organizational goals, whether or not performance is formally measured (Holmstrom and Milgrom 1991). Even though financial or other direct incentives may not be available to local managers to influence the performance of frontline
workers, agents have incentives to devote more effort to the aspects of their jobs that are measurable (and that supervisors can observe more easily), and to devote less effort to aspects that are unmeasured or unobserved. Thus, Holstrom and Milgrom argue, one agent or group should be assigned tasks that are relatively easier to observe and measure, and another agent or group should be assigned tasks that are relatively more difficult to observe and measure. When explicit reward incentives are not available, constraints on the tasks in a job “are substitutes for performance incentives and are extensively used when it is hard to assess the performance of the agent” (Holmstrom and Milgrom 1991, 27).

To generate predictions from multitask principal agent theory about optimal casework job design, the measurability of core casework tasks and other tasks in the welfare-to-work office thus need to be considered. For the core tasks, neither outputs nor outcomes are easily measured (characteristics of Wilson’s [1989] “coping” organizations). The outputs (interactions between caseworkers and clients) are difficult to observe and measure due to the very nature of frontline work in human services (Hasenfeld 1983; Lipsky 1980). The outcomes (the eventual results of client-caseworker interactions) are also difficult to observe and measure due to many factors, including the fact that clients themselves are inputs to the process, the number of possible results of interest, and the long time frame over which client responses may accrue (Hasenfeld 1983). Thus, every one of the four core casework tasks is difficult to measure, and multitask principal-agent theory suggests that agents’ work will best be directed toward organizational goals when these difficult-to-measure tasks are grouped together in a single job (i.e., a unified case manager).

The core tasks are not the only ones that caseworkers in a welfare-to-work office may be involved in, as noted earlier. These other tasks may be relatively easier for managers to observe
and assess: For example, the outputs of job developers may be measured by number of contacts with employers, or number of job positions actually developed. Even though outcomes remain difficult to observe for the reasons just discussed (characteristics of Wilson’s [1989] “procedural” organizations), multitask principal-agent theory suggests that agents’ work will best be directed toward organizational goals when these relatively more easily-measured tasks are separated from less-measurable tasks (i.e., a specialist conducts these other tasks).

To summarize, multitask principal agent theory applied to the measurability of various core and other casework tasks thus suggests the following hypotheses:

\[ H_1: \] Offices that use unified case management will have better client outcomes than offices that do not use unified case management.

\[ H_2: \] Offices that use specialists for job development or conducting services will have better client outcomes than offices that do not use specialists for these activities.

Insight into effective casework job design can also be obtained from a different perspective, which holds that service technology issues may be more important than incentives to understand the grouping of tasks into jobs, and their relationship to performance (Lazear 1995). Using a service technology lens, if tasks are complementary inputs into production then their grouping together in a job can improve effectiveness (and may improve efficiency). In welfare-to-work offices, the core casework tasks are likely complementary inputs to helping clients increase earnings and reduce welfare receipt: a unified case manager may gain useful information through each of the four core functions that enhances the effectiveness of the others.

In addition to task complementarity, task interdependence is a relevant characteristic of service technology. Thompson (1967) posited that rational organizations would group together positions according to their features of pooled, sequential, or reciprocal interdependence, so that
the appropriate form of coordination could be applied (coordination by standardization, by plan, and by mutual adjustment, respectively). Thompson’s discussion refers to the grouping of positions into teams or work groups; however, the same logic can apply to the grouping of tasks into jobs—a similar problem at a lower level.

The welfare-to-work core casework tasks may be viewed as sequentially interdependent, or alternatively, as reciprocally interdependent. Sequential interdependence might tend to promote a “people-processing” technology, which is focused on the definition and disposition of cases (Hasenfeld 1983). For example, a client must be assessed before an employability plan can be developed, and must receive a job or program assignment before staff can monitor her participation. In contrast, reciprocal interdependence might tend to promote a “people-changing” technology, which aims to improve clients’ wellbeing (Hasenfeld 1983). For example, assessment of client needs, development of an employability plan, assignment to program components, and monitoring may require adjustments as more information about the client is obtained and as her needs change.

In the context of welfare-to-work offices studied here, increasing client earnings and reducing AFDC receipt were (people-changing) goals of the program and the outcomes of interest, implying that bundling the core components of casework into a single job will lead to greater effectiveness. In contrast, tasks such as job development or conducting services are not complementary or interdependent (at least to the same degree as the core tasks), implying that separating these stand-alone casework tasks into different jobs will lead to greater effectiveness.

To summarize, service technology characteristics of welfare-to-work office casework point to the same hypotheses as those generated from the multitask principal-agent model.
These two hypotheses are examined using regression analysis, which holds constant a number of client, office, and local economic environment characteristics.

DATA

This analysis merges data from three random assignment studies of welfare-to-work programs conducted by MDRC: Greater Avenues for Independence (GAIN) in California, Project Independence (PI) in Florida, and the National Evaluation of Welfare-to-Work Strategies (NEWWS) in six states. The programs in GAIN, PI, and NEWWS were part of the states’ Job Opportunities and Basic Skills Training program (JOBS), which was operational under the Family Support Act of 1988, and which required certain welfare recipients to participate in work, education, or training activities. Though implemented prior to the PRWORA, the mandatory nature of and services available in JOBS programs are comparable to those that local office managers face under current program rules.

Random assignment dates ranged from March 1988 for some GAIN offices to December 1994 for some NEWWS offices; thus, the two-year follow-up periods for clients occurred over a period from the late 1980s through the mid 1990s (table 1). The MDRC evaluations provided information about program implementation, examined whether the programs decreased welfare receipt and increased client labor force participation and earnings, and measured the benefits and costs of the program. The current analysis examines additional governance questions regarding implementation and performance.

I impose two restrictions on the data. First, only females (who constitute over ninety percent of the full sample) are included. Second, only clients who were randomly assigned to the treatment group are included: information about casework design is available only for these
offices, and analyzing them will produce results comparable to other studies that do not have access to data from random assignment studies.\textsuperscript{11}

Pooling data from the three evaluations and imposing the two sample restrictions produces a data set with 47,115 female treatment group members in seventy-five offices in twenty-four counties in seven states. Administrative data provide information about each welfare-to-work client: data on labor force participation and earnings are available from state unemployment insurance systems, receipt of AFDC is available from county and state records, and client characteristics are available from administrative data intake forms.

Surveys of approximately 1,600 staff in the seventy-five offices provide information about casework practices and relationships with clients. Administered during a period that generally corresponded to the follow-up period for clients,\textsuperscript{12} the surveys had coverage rates greater than ninety percent in each office, and similarly high response rates. Because supervisors mostly had limited (if any) contact with clients (Bloom, Hill, and Riccio 2001, 87), I exclude supervisors’ responses (n=218) resulting in a sample of 1,385 caseworkers, with a range of one to eighty-three staff respondents per office, a mean of eighteen, and a median of ten.

METHODS

To maximize internal and external validity, the ideal research design would randomly assign task configurations to caseworkers within and across offices; randomly assign each client to a caseworker; validate implementation of these configurations using observation, timed journal entries, or other methods; and measure effects of different casework configurations on client outcomes. Because the current study attempts to glean insights about governance and performance using data from evaluations originally conducted for a different purpose, the ideal research design is not attainable. Instead, I use the natural variation in outcomes and casework
job designs across offices to identify the effects of different casework task configurations, controlling for many characteristics of clients, caseworkers, offices, local economic environments, and states. A drawback of this nonexperimental approach is that if certain characteristics are correlated with both the office’s casework configuration and with client outcomes—and if these characteristics are not included in the model—then the coefficients on casework will absorb those influences (i.e., they will be biased).

The analysis uses both individual- and office-level data in a hierarchical linear model (HLM) to estimate effects at each level (Raudenbush and Bryk 2002). Multilevel modeling (also known as mixed, random intercept, random coefficient, or variance component modeling) is advantageous given the research question being considered and the available data because it partitions variance components at each level, and produces proper estimates of the standard errors (Raudenbush and Bryk 2002).

The level one model uses individual client data in the following specification:

\[ Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \ldots + \beta_{nj}X_{nij} + \varepsilon_{ij} \]  \[1\]

Where

\[ Y_{ij} \]  
outcome measure \( Y \) for client \( i \) in office \( j \)

\[ \beta_{0j} \]  
mean outcome for clients in office \( j \), adjusted for observable differences in clients across offices\(^{13}\)

\[ X_{1ij} \ldots X_{nij} \]  
\( n \) individual client characteristics for client \( i \) in office \( j \)

\[ \beta_{1j} \ldots \beta_{nj} \]  
average effect of \( X_1 \ldots X_n \) on \( Y \)

\[ \varepsilon_{ij} \]  
individual-level error, assumed to be distributed Normally, with mean zero and variance \( \sigma^2 \)
The intercept, $\beta_{0j}$, and each of the coefficients, $\beta_{1j}$, $\beta_{2j}$, ... $\beta_{nj}$, may be specified as either fixed or varying across level two units (offices). In the specifications tested here, I focus on explaining variation in the adjusted average outcomes across offices, $\beta_{0j}$, and assume the effects of the other characteristics are fixed across offices, thus $\beta_{ij}=\beta_1$, $\beta_{2j}=\beta_2$ ... $\beta_{nj}=\beta_n$. One equation—to model adjusted average outcomes—then is estimated at level two using information from the seventy-five offices:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}W_{ij} + \ldots + \gamma_{0k}W_{kj} + \mu_j \quad [2]$$

Where

- $\beta_{0j} =$ adjusted mean outcome in office $j$
- $\gamma_{00} =$ overall adjusted mean outcome
- $W_{1j}...W_{kj} =$ $k$ office-level characteristics for office $j$
- $\gamma_{01}...\gamma_{0k} =$ effect of office-level characteristic $W_1...W_k$ on $\beta_{0j}$
- $\mu_j =$ office-level component of error, assumed to have mean zero and variance $\tau$

The question examined here could be estimated using ordinary least squares (OLS) at either the client or organizational level. However, in addition to producing proper estimates for the standard errors, multilevel methods are especially advantageous in governance research (when feasible), where the conceptualization of multiple levels of influence and variation can be reflected in the estimation method (Heinrich and Lynn 2001; Lynn, Heinrich, and Hill 2001; Raudenbush and Bryk 2002).

**MEASURES**

**Dependent Variables**
A number of measures of program performance or client success might be considered, including labor market measures, welfare receipt, child well-being, and other social, behavioral, and economic outcomes (see, e.g., Friedlander and Burtless 1995; or Riccio, Bloom, and Hill 2000 for discussions). In this study, effectiveness is measured by two visible and widely-acknowledged goals of these welfare-to-work programs: increasing client earnings and decreasing AFDC benefits. Specifically, the primary dependent variables are total earnings and total AFDC benefits over a two-year period after random assignment. This two-year follow-up period provides time for most persons to have completed training or education.15

As shown in table 2, average earnings (in constant 1996 dollars) for all clients over the two-year follow-up period were $6,042, with a standard deviation of $10,220. Average AFDC benefits (in constant 1996 dollars) over this same period were $7,818, with a standard deviation of $6,358. Additional outcome variables are examined in supplementary analyses; these measures and their means are listed in appendix 1.

[Table 2 about here]

**Independent Variables**

It is not possible with these data to match individual clients with individual caseworkers. Instead, measures of casework job design are specified at the office level, then clients and offices are matched at the office level. An extensive set of client- and office-level controls are included in the estimated regressions.

**Client Level**

The first panel of table 2 shows the individual client characteristics included as controls in the model for each of the 47,115 clients. These are standard in welfare-to-work and job training studies, and each is measured at the time of random assignment: age, number of children, age of
youngest child, race/ethnicity, AFDC receipt history, earnings history, and education level. Almost half of the sample was between twenty-five to thirty-four years old, and had two children on average. Forty-two percent had a child younger than six years old, fifty-six percent had a high school degree or GED, and forty-one percent had received AFDC for all twelve months prior to their random assignment. Forty-five percent of clients had worked at some point in the previous year, and average total earnings in the year prior to random assignment (including zero earnings) were $2,231.

Office Level

The first step in characterizing office casework configurations—the primary measures of interest in the current study—was to measure the activities of each caseworker in each office. Appendix 2 details this process. The next step was to characterize configurations of casework at the office level. Four different zero/one measures for each office were created (these measures were not mutually exclusive):

- whether the office primarily used unified case management;
- whether the office used at least one specialist in the core tasks of casework;
- whether the office used at least one specialist in job development;
- whether the office used at least one specialist to conduct other services.

For the first measure—unified case management—I assigned an indicator variable equal to one to offices in which at least half of all line staff were case managers ($n = twenty-seven offices). This measures whether unified case management was the primary way of structuring core casework in the office. In addition to having the intuitive interpretation that case managers constituted a majority of line staff, this cut point was at the seventy-fifth percentile of the distribution of percentage of case managers across offices. 16
Office-level measures of specialists indicate whether the office had at least one specialist in each category above (see appendix 2). Forty offices used a specialist in at least one core casework task; thirteen offices used a specialist job developer, and twenty-nine offices used specialists to conduct services such as orientation or job club.

The second panel of table 2 shows descriptive statistics for these variables. Additional tabulations (available from the author) show that the configurations of casework varied across offices and evaluations, and that no single configuration dominated. Further, to test whether the empirical results were sensitive to the measures of unified case management and specialists, I constructed alternative measures and tested them in the models. The results (available from the author) were generally robust to specification using these other measures.

Other office-level control variables were included for other aspects of service technology, structure, and environment if they were likely to affect the presence of particular job designs in an office, or if holding them constant led to a clearer interpretation of the job design coefficients.

First, caseload size may reflect the resources available to the organization, and large caseloads may prevent case managers or specialists from being able to spend time with clients (Gueron and Pauly 1991; Rose and Moore 1995). The average across offices in this study was 151 clients per caseworker, with a standard deviation of seventy-three. Second, I control for the number of caseworkers in the office: larger offices may be more likely to use specialists than smaller offices, as structural complexity increases with larger organizational size (Scott 1998). On average, offices had eighteen line staff, with a range from one to eighty-three.

Third, I include a measure of caseworker tenure to control for experience and turnover, because experience may be related to job assignments. The average tenure of caseworkers in these offices was 2.8 years, with a standard deviation of 1.6. Fourth, to control for expertise
gained on the job, I include a measure of previous staff experience in JTPA programs, in Work Incentive (WIN) programs, or as an employment counselor. On average, forty-two percent of staff in an office had previous welfare-to-work experience. Fifth, caseworker education beyond a college degree is included as a proxy for professional training. In organizations where job tasks or outcomes are not observable by managers, employing professionals is one way of addressing the complexity of the job (Scott 1998; Wilson 1989). On average, twenty-six percent of staff in these offices had education beyond a college degree.

Sixth, the emphasis that staff place on quickly moving clients into jobs (instead of building basic skills) is included as a control. This emphasis measures the formal and informal messages that caseworkers convey to clients regarding the importance of work. This feature has received much attention in the literature, and has been linked to improved performance of welfare-to-work programs (Bloom, Hill, and Riccio 2003; Hotz, Imbens, and Klerman 2000; Riccio, Friedlander, and Freedman 1994; Riccio and Orenstein 1996). Bardach (1993, 1997) has noted its importance in effective JOBS programs, referring to it as a “high expectations” approach. For example, “At all stages of the program, registrants were told that employment was central, that it should be sought expeditiously, and that opportunities to obtain low-paying jobs should not be turned down” (Riccio and Orenstein 1996, p. 17). For the current study, I use a scale constructed from four staff survey questions created by Bloom, Hill, and Riccio (2003), standardized to have a mean of zero and standard deviation of one (see appendix 1). In addition to the likely relationship with earnings and AFDC receipt, this measure also may be associated with casework task configurations in the office. For example, if offices that emphasize quick job placement are more likely to have job development specialists, then the coefficient on job
development specialists would be biased upward if overall emphasis on quick employment were not controlled for and if this emphasis is associated with higher client earnings.

Two different types of variables are included to capture local and state economic and environmental factors. The unemployment rate (constructed from county-level data) controls for local labor market conditions that vary across counties. State indicator variables control for a range of factors that reflect the economic environment, program choices or mandates, or other factors at the state level that influence client outcomes. These state indicators control for possible associations between state-level decisions and local area decisions regarding casework task configurations, but they remain black-box effects.17

Controls are not included for other aspects of interactions between clients and staff, such as trust, interaction frequency, counseling intensity, or client participation in activities. By excluding these other factors from the model, the job design variables reflect them. It is possible that these other factors are related both to job design (e.g., trust may be greater between clients and unified case managers when they interact frequently) and to outcomes. However, factors such as personalized attention or trust may be influenced by the casework task configuration in an office, but the reverse is unlikely to be true (e.g., trusting relationships between clients and caseworkers are unlikely to precede an office using unified case management).

None of the bivariate correlations between office-level measures in the model were notably strong. The strongest correlations (statistically significant at least at the 0.01 level) all involved the number of line staff: with whether an office had at least one specialist in a core casework task ($r = 0.41$); with percent of staff with education beyond a college degree ($r = -0.41$), and with the emphasis placed on moving clients into jobs ($r = -0.35$).

**RESULTS AND DISCUSSION**
Variance Components and Explanatory Power

If the mean outcomes (of earnings and AFDC benefits) do not vary across offices, or if client characteristics alone fully explain any across-office variation in mean outcomes, then office-level factors (including organizational or managerial factors) cannot logically play a further explanatory role. Thus, before estimating the full model and examining the substantive hypotheses of interest, a necessary step is to establish whether statistically-significant variation in unconditional and conditional mean outcomes exists across offices.

For the outcome measure of total earnings in two years after random assignment, ninety-eight percent of the total variation across clients, and two percent is across offices (i.e., the intraclass correlation $\rho = 0.02$). This variance composition is comparable to that found in similar studies of low-income earnings and work programs (e.g., Heinrich 2002). While two percent of total variation at the office level seems relatively small, it is both statistically significant and substantively important: unconditional mean earnings across the seventy-five offices range from $1,569 to $9,229, with a mean of $5,857 and a standard deviation of $1,588. Next, controlling for client characteristics at level one explains nineteen percent of the individual-level variation in earnings. Importantly, the variation in conditional mean earnings across offices ($\beta_0$) is statistically significant even after controlling for client characteristics, and thus remains to be explained with additional variables in the level two model. Finally, the preferred model specification (with a full set of controls at levels one and two, shown in the first two columns of tables 3 and 4) explains nineteen percent of the individual-level variation in earnings (level one) and eighty-five percent of this variation in conditional mean earnings across offices (level two). The explanatory power of these models is comparable to similar studies.
For total AFDC benefits in the two years after random assignment, sixty-three percent of the total variation is at the client level, and twenty-seven percent at the office level ($\rho = 0.27$), which is statistically significant. Next, client characteristics alone explain eighteen percent of the individual-level variation, and here too the variation in AFDC receipt across offices is statistically significant and remains to be explained through a level two model. Finally, the preferred model specification (with a full set of level one and two controls, shown in the last two columns of tables 3 and 4) explains eighteen percent of the individual-level variation in AFDC benefits (level one) and ninety-eight percent of this variation across offices in conditional mean AFDC benefits (level two).

These findings regarding variation in outcomes are notable not only as model fit statistics, but also as evidence of the potential role for public management and governance. The fact that variation in outcomes across offices remains to be explained even after controlling for client characteristics indicates that governance factors may play a role in affecting these outcomes.

**Client-Level Results (Level One)**

Table 3 shows individual-level estimates (obtained through simultaneous estimation of the level two model) from the preferred model specification, for both earnings and AFDC outcomes. For the earnings model (columns one and two), statistically significant associations, conditional on the other variables in the model, include age (compared to a baseline category of clients ages twenty-five to thirty-four); whether the client was Hispanic or Native American (compared to a baseline of white, nonhispanic); prior employment and earnings; and whether the client had a high school degree or GED. This final effect is the largest: clients who had a high school degree or GED prior to their random assignment into the program are predicted to earn
about $2,400 more on average than nongraduates over the two-year follow-up period, controlling for other individual and office characteristics. This effect is substantively moderate, representing almost one-quarter of a standard deviation increase in earnings over the two-year period.

[Table 3 about here]

In the model of total AFDC earnings (columns three and four), many individual characteristics are statistically significant predictors, all in the expected directions. These characteristics include age, Black, Hispanic, number of children, presence of young children, and prior AFDC receipt and earnings. The largest conditional effect in this model is for clients who had received AFDC all twelve months prior to random assignment: controlling for the other individual and office characteristics, these clients were predicted to receive $2,507 more in AFDC benefits over the two years compared to clients who had not continually been on AFDC in the prior twelve months.

Office-Level Results (Level Two)

Table 4 shows results from models of total earnings (columns one and two) and total AFDC benefits (columns three and four) over the first two years after random assignment. Each model contains the same set of individual covariates at level one.

[Table 4 about here]

**Earnings**

Findings show mixed support for the hypotheses regarding case management and specialization. In support of hypothesis 1, average client earnings are $343 higher over the two years following random assignment in offices where at least half of all line staff are case managers, controlling for other individual client and office characteristics. Substantively, this is a small-to-moderate effect: 0.03 of an individual-level standard deviation or 0.22 of an office-level
standard deviation in earnings. Full support of this hypothesis would be found if the coefficient on the core casework specialist variable were negative (which it is) and statistically significant (which it is not).  

As noted earlier, the positive association between unified case management and earnings may be explained through at least two channels. First, the finding is consistent with the incentive explanation from multitask principal-agent theory: that is, staff are better directed toward organizational goals when unmeasurable and measurable tasks are grouped in separate jobs. A key piece of information in interpreting this explanation is whether efforts were actually made to direct staff toward organizational goals, and whether caseworker performance was assessed. Some system of performance standards or management was in place in many of the evaluation sites (Hamilton and Brock 1994; Kemple and Haimson 1994; Riccio and Friedlander 1992), but information is not available about specific practices in each office. Thus, the finding is consistent with the incentive explanation, but it is not possible to confirm it with the available data.  

The unified case management finding is also consistent with a service technology interpretation, suggesting that the core casework tasks may be complementary inputs. This finding supports Mead’s (1997, pp. 61-62) observation that effective case management combines both “help and hassle” (but it is not possible here to confirm whether the rule enforcement role is fundamental, as Mead argues). The finding suggests that unified case management is more conducive for advancing the goals of a people-changing welfare-to-work office (measured by increasing earnings). It is possible that client-caseworker interactions in unified case management tended to be “transformational” rather than “routinized,” “particularistic,” or “instrumental” (Meyers, Glaser, and MacDonald 1998). These interpretations of the unified case management effect are suppositions, however, and should be pursued in future quantitative
and qualitative research. Further, while it is possible that unified case management leads to
greater effectiveness, it is also possible that more effective organizations tend to use unified case
management to a greater degree.

In support of hypothesis 2, average client earnings are $524 higher over the two years
following random assignment in offices with at least one job development specialist, controlling
for other variables in the model. Substantively, this too is a small-to-moderate effect,
representing 0.05 of an individual-level standard deviation and 0.33 of an office-level standard
deviation in earnings. The hypothesis is not supported for the conducting-services specialist,
which has an estimated sign in the direction predicted (positive), but is not statistically
significant.

The job development specialist finding is consistent with both the incentive and service
technology explanations, and is notable because it emerges even when controlling separately for
the informal emphasis that staff place on moving clients quickly into jobs. One possibility is that
this finding reflects a resources effect—offices with greater resources employ specialist
caseworkers and help their clients in other ways, leading to better outcomes. While this
explanation cannot be ruled out, it is unlikely for two reasons: first, caseload size serves as a
proxy for resources. Second, if a residual effect remains, one would also expect the other
specialist variables to be significant. This is not the case.22

The level two control variables are also of interest from governance and management
perspectives. First, a one standard deviation increase in the number of line staff (eighteen
caseworkers) is associated with a $238 decline in average client earnings, holding other variables
in the model constant. Stated differently, larger offices are less effective, perhaps because of less
personalized attention. Yet one would expect this feature to operate through the caseload size
variable (which is included separately in the model). Second, a one standard deviation increase in the emphasis staff place on moving clients into jobs is associated with a $441 increase in total earnings. This finding is consistent with a number of previous studies, which draw their findings from sites included in the present analysis (Bloom, Hill, and Riccio 2003; Hotz, Imbens, and Klerman 2000; Riccio, Friedlander, and Freedman 1994; Riccio and Orenstein 1996), as well as from entirely different sites (Behn [date], Mead [date], Meyers & Riccucci [date]). Finally, estimates for the unemployment rate are statistically significant and in the expected direction, indicating that a one standard deviation increase in the unemployment rate is predicted to reduce average client earnings by $285—a result similar in magnitude to the case management and job development specialist coefficients: it is more difficult for clients to increase earnings in tougher economic times.

**AFDC Benefits**

For the outcome variable of total AFDC receipt in the two years after random assignment, the job design coefficients (table 4, columns three and four) do not support the hypotheses regarding case management and specialization: Although three of the four signs are in the hypothesized direction, no estimate is statistically significant. Thus, average AFDC receipt in offices that primarily use case management, or that use specialists, is no different compared to offices that do not use these job designs.

As in the earnings model, the coefficient on the average unemployment rate is statistically significant and in the expected direction. Other office-level control variables are statistically significant and in the expected direction: holding other level one and two variables constant, emphasis on moving clients into jobs is associated with lower AFDC receipt; and caseload size and the number of line staff are associated with higher AFDC receipt.
Reconciling the Job Design Results for Earnings and AFDC

The findings reported above indicate that use of unified case management and job development specialists are positively associated with earnings, but not associated with AFDC benefits, conditional on the other individual and office characteristics in the model. To further explore this pattern of results, I examined a series of alternative outcome measures for client earnings, employment, and AFDC, including the full set of level one and level two controls in each model.

The results of these supplementary analyses (available from the author), suggest possible explanations for the presence of the positive effects of unified case management and job development specialists on average total two-year earnings while over the same period the absence of negative effects for average total AFDC benefits. If it were the case that increased earnings were produced primarily through quicker job placement (but lower-paying jobs), the effects of the job design variables would be expected to shrink over time. However, the supplementary analysis showed just the opposite: Clients in offices that primarily used unified case management had higher earnings later in the two-year follow-up period, but no differences in probability of employment. Clients in offices that used job development specialists showed higher employment rates and higher earnings later in the follow-up period. I find some evidence of corresponding AFDC benefit reductions for the job development specialist variables in the second year, suggesting that the offsetting effects of increased earnings are beginning to show.

A possible explanation for these findings is that the more employable clients are in offices with these job designs. This explanation is unlikely, however, because client characteristics that reflect employability (e.g. prior work experience, presence of children, race, age) are controlled for. A more likely explanation for these patterns is that offices using unified
case management and job development specialists produce better matches between clients and job opportunities over the longer run—they are more conducive to “people-changing” processes.

These supplementary analyses provide suggestive—but not conclusive—evidence that estimates for these job design variables may underestimate clients’ longer-term benefits from being served in offices that primarily use case management and that employ job development specialists. This premise could be tested using a longer followup period.

CONCLUSION

If program designers and managers are to plan and implement welfare-to-work programs, or any program carried out in units across dispersed locations, a need exists to explore variations in performance across offices, identifying whether and how implementation practices, management actions, and governance mechanisms are related to performance. The research reported here focuses on a specific feature of local offices—casework job design—that local managers may be able to change to increase program effectiveness. It complements existing studies that find variation in performance across offices, but that do not address the question of why performance varies, or that focus on client characteristics; and studies that examine implementation in depth, but that are not designed to systematically link these findings to outcomes.

Ideally, the question of how casework job design is related to performance would be addressed by randomly assigning casework task divisions across caseworkers and across offices, randomly assigning clients to caseworkers, verifying implementation conditions through contemporaneous data collection and observation, and measuring the value-added of different casework configurations using a number of different outcome measures. Similar ideal designs, which would maximize internal and external validity, might be pursued for a number of
governance research questions. Because such ideal designs are seldom feasible, however, alternative approaches are needed to analyze the possible links between governance and performance.

In lieu of the ideal research design, in this article I develop and test hypotheses, using a rich data set on individuals and organizations that was collected for other purposes. An advantage of this strategy is that it has produced one of the first studies to examine the link between core casework job design and office performance, offering some empirical evidence to public managers and program designers in the PRWORA era as demands increase for people-changing technologies (Brown 1997; Meyers 1998). The strategy has some disadvantages, mostly reflecting the fact that the ideal research design and data collection strategies described above could not be attained.

A recent edited volume on implementation research and welfare reform (Lennon and Corbett 2003), and recent literature on methods for governance research (Gill and Meier 2000; Lynn, Heinrich, and Hill 2001) provide roadmaps for accumulating a body of knowledge about the links between governance and performance in welfare-to-work offices and other human services programs. The analysis presented in this article aims to contribute to this body of knowledge with regard to casework job design. While this study represents one of the first to provide estimates of the links between casework job design and client outcomes, future qualitative and quantitative research should test the robustness of the findings reported here, especially operationalizing case management, specialists, and services received in other ways, using other measures, with other data, in other settings. Accumulation of such findings will help us further understand the contribution of this aspect of governance and public management to organizational effectiveness.
APPENDIX 1
Construction of Dependent and Independent Variables

Outcome Variables: Individual Client Level

Total earnings in two years following random assignment

Source: State unemployment insurance (UI) administrative data. Sum of earnings in eight calendar quarters, where the first quarter is the quarter after random assignment. All dollar amounts were converted to constant dollars (1996 base) using the CPI-U. Maximum earnings in each quarter were top-coded at $20,000.

Total AFDC benefits in two years following random assignment

Source: AFDC administrative data. Sum of AFDC benefits in eight calendar quarters, where the first quarter is the quarter following the quarter of random assignment. All dollar amounts were converted to constant dollars (1996 base) using the CPI-U.

Additional Employment, Earnings, and AFDC outcome variables Used in Supplementary Analyses (created from the above sources):

Earnings

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total for Year 1</td>
<td>$2,509</td>
</tr>
<tr>
<td>Total for Year 2</td>
<td>$3,533</td>
</tr>
<tr>
<td>In Last Quarter of Follow-up</td>
<td>$941</td>
</tr>
<tr>
<td>(eighth quarter after random assignment)</td>
<td></td>
</tr>
<tr>
<td>Earnings per quarter worked</td>
<td>$1,118</td>
</tr>
<tr>
<td>(two-year period)</td>
<td></td>
</tr>
<tr>
<td>Earnings per quarter worked</td>
<td>$832</td>
</tr>
<tr>
<td>(Year 1)</td>
<td></td>
</tr>
<tr>
<td>Earnings per quarter worked</td>
<td>$1,072</td>
</tr>
<tr>
<td>(Year 2)</td>
<td></td>
</tr>
</tbody>
</table>

Employment

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total quarters employed (two-year period)</td>
<td>2.76</td>
</tr>
<tr>
<td>Ever employed (two-year period)</td>
<td>63%</td>
</tr>
</tbody>
</table>
Ever employed (Year 1) 50%
Ever employed (Year 2) 52%
Employed in Last Quarter of Follow-up (eighth quarter after random assignment) 38%

Amount AFDC Received

- Total for Year 1 $4,552
- Total for Year 2 $3,266
- In the Last Quarter of Follow-up (eighth quarter after random assignment) $735
- AFDC Benefits per quarter received AFDC (two-year period) $1,248
- AFDC Benefits per quarter received AFDC (Year 1) $1,280
- AFDC Benefits per quarter received AFDC (Year 2) $908

Ever Received AFDC

- Total Quarters Received AFDC (two-year period) 5.46
- Ever received AFDC (two-year period) 93%
- Ever received AFDC (Year 1) 92%
- Ever received AFDC (Year 2) 69%
- Received AFDC in Last Quarter of Follow-up (eighth quarter after random assignment) 53%

Explanatory Variables: Individual Client Level

Note on missing information: To maintain as many observations as possible, I imputed the office-level mean value for an observation when a variable value was missing.

Note on transformation of variables in Level One of HLM model: I centered each individual’s value on the following variables around the overall average value (i.e., grand-mean centering). With this transformation, the Level 1 intercept for each office is interpreted as an
estimate of the office-level average outcome, adjusted for client differences across offices (Raudenbush and Bryk 2002).

Age of client at random assignment (omitted category is age from twenty-five to thirty-four years old)

Client age less than twenty-five (indicator variable)
Client age between thirty-five and forty-four (indicator variable)
Client age forty-five or older (indicator variable)

Race/Ethnicity (omitted category is White, nonHispanic)

Asian (indicator variable)
Black, nonHispanic (indicator variable)
Hispanic (indicator variable)
Native American (indicator variable)
Other race/ethnicity, nonwhite (indicator variable)

Total number of children (continuous variable)
Youngest child is less than age six (indicator variable)

High school graduate or GED recipient (indicator variable)

Received AFDC continuously for all twelve months prior to random assignment (indicator variable)

AFDC applicant (instead of ongoing recipient) at time of random assignment (indicator variable)

Ever employed in twelve months prior to random assignment (indicator variable)

Total earnings in twelve months prior to random assignment (1996 dollars)

**Explanatory Variables: Office Level**

Job Design (see appendix 2 for a full description of the creation of these variables)
Used primarily case management model

Used at least one specialist for core casework tasks

Used at least one specialist for job development

Used at least one specialist for conducting services

Staff emphasis on moving clients into jobs, versus building skill levels.

- Used staff responses to four questions on staff survey (the response scale is listed in parentheses after each question):

  1. Based on the practices in your unit, what would you say is the more important goal: to help clients get jobs as quickly as possible or to raise the education or skill levels of clients so that they can get jobs in the future? (1=skills…..7=jobs)

  2. In your opinion, which should be the more important goal of your unit: to help clients get jobs as quickly as possible or to raise the education or skill levels of clients so that they can get jobs in the future? (1=skills…7=jobs)

  3. After a short time in the program, an average welfare mother is offered a low-skill, low-paying job that would make her slightly better off financially. Assume she has two choices: either to take the job and leave welfare OR to stay on welfare and wait for a better opportunity. If you were asked, what would your personal advice to this client be? (1=welfare….7=jobs)

  4. What advice would your supervisor want you to give to a client of this type? (1=welfare….7=jobs)

- Calculated the average of available responses for each line staff person.

- Calculating the average line staff response in the office created an office-level measure, which was standardized to have a mean equal to zero and standard deviation equal to one.
Caseload

Average caseload per line staff (based on line staff survey self-reports)

Post-college education

Percentage of line staff that completed at least some graduate work, beyond college

Previous work experience in welfare-to-work programs

Percentage of line staff that have prior experience in a JTPA program, WIN program, or as an employment counselor

Tenure

Average number of years that line staff have been in their current jobs

Total number of line staff in the office

Average unemployment rate during two-year period following random assignment

First, constructed at the individual client level using monthly county-level data, covering the 24-month follow-up period after random assignment (SOURCES: U.S. Bureau of Labor Statistics, Local Area Unemployment Statistics; and California Employment Development Department). Second, computed the office-level average of individual-level values.

AFDC benefit level for a family of three during follow-up period (in hundreds of 1996 dollars)

First, constructed at the individual client level as a weighted average (based on the state and the time period covered in each client’s follow-up period) of the maximum AFDC benefit for a family of three (SOURCE: 1990, 1993, and 1996 Green Book, U.S. House of Representatives). Second, computed the office-level average of individual-level values.

State indicator variables

For California GAIN (the omitted category), California NEWWS, Florida, Georgia, Michigan, Oregon, Oklahoma, and Ohio.
APPENDIX 2
Construction of Job Design Variables

Casework job design measures for each office were created using questions on the GAIN, PI, and NEWWS staff surveys. First, I measured the job tasks of each caseworker in the office. Then, I used these to create an aggregate measure to characterize the office.

Creating Job Design Measures at the Caseworker Level

Respondents indicated the time they spent on each task in their job using the following scale: 23

<table>
<thead>
<tr>
<th>Response Value</th>
<th>Description of Amount of Time</th>
<th>Percentage of time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>None</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>Some</td>
<td>1 – 25%</td>
</tr>
<tr>
<td>3</td>
<td>Moderate</td>
<td>26 – 50%</td>
</tr>
<tr>
<td>4</td>
<td>Substantial</td>
<td>51 – 75%</td>
</tr>
<tr>
<td>5</td>
<td>Most</td>
<td>76 – 100%</td>
</tr>
</tbody>
</table>

Respondents were told that their responses need not sum to 100 percent. For the categorization of casework responsibilities, I relied on the literature about casework activities and case management, which identifies four core areas of casework: (1) assessing client needs; (2) developing client employability plans; (3) arranging and coordinating services; and (4) monitoring client progress (Hagen and Lurie 1994a; Rothman 1991). I then mapped each survey question about tasks into these general categories. 24

A caseworker was considered a case manager if she had at least some involvement (i.e., she must spend one percent or more of her time) in all of the core areas of casework:

-- interviewing and assigning clients to activities
-- arranging and coordinating services
-- monitoring

For remaining line staff (i.e., those who are not case managers), I identify specialists in various tasks.²⁵

**Specialist caseworkers** are defined as staff that spent greater than seventy-five percent of their time on any one core casework task.

**Job development specialists** are defined as staff that spent greater than seventy-five percent of their time on job development.

**Service specialists** are defined as staff that spent greater than seventy-five percent of their time conducting either orientations or job club workshops.

Furthermore, I classified a staff person as a specialist only if she reported spending greater than seventy-five percent of her time on the specified task, and not spending greater than seventy-five percent of her time on some other task. Sensitivity tests (available from the author) show that the earnings results were not sensitive to alternative definitions of “specialist,” however, the AFDC results were sensitive.

**Creating Job Design Measures at the Office Level**

To classify each office as using primarily case management, I assigned an indicator variable equal to one to offices where at least half of caseworkers were case managers. Office-level measures of the use of specialists simply indicate whether any specialist of a particular kind is employed in that office.

Sensitivity tests (available from the author) indicate that when instead the percentage of unified case managers in an office was used, the estimate was positive, but statistically
insignificant in the earnings model. Its use did not change the findings in the AFDC model. An alternative model using a two-variable spline showed a positive, statistically significant effect only on the highest category in the earnings model, but no different conclusions in the AFDC model.
NOTES

1 I am grateful for financial support from the Northwestern University/University of Chicago Joint Center for Poverty Research, the Searle Family Fund, and the Sydney Stein, Jr. Fund. For helpful feedback, I thank Howard Bloom, Carolyn Heinrich, Brian Jacob, Bob LaLonde, Helen Levy, Larry Lynn, Larry Mead, Colm O'Muircheartaigh, Jim Riccio, Robin Tepper, and Michael Wiseman; anonymous JPART reviewers; participants in the Harris School Ph.D. workshop; and seminar participants at the Ford School, the Maxwell School, and the Wagner School; and at the annual meetings of the Midwest Political Science Association, the National Association for Welfare Research and Statistics, and the Association for Public Policy Analysis and Management.

2 Public sector governance has been defined as “regimes of laws, rules, judicial decisions, and administrative practices that constrain, prescribe, and enable the provision of publicly supported goods and services” through formal and informal relationships with agents in the public and private sectors (Lynn, Heinrich, and Hill 2001, 7).

3 The findings and conclusions presented here do not necessarily represent the official positions or policies of MDRC or its funders.

4 Subsequently, I refer generally to frontline staff as “caseworkers” and to frontline staff involved in all four core functions of casework as “case managers.”

5 Elsewhere in the literature, these three models are referred to as generalist, sequential specialist, and team models, respectively.
In fact, performance measures (e.g., based on clients served, participation, inputs, outputs, and outcomes) are often used in local welfare-to-work offices (Doolittle and Riccio 1992; Mead 1997).

Two inputs $X_1$ and $X_2$ are complementary if the marginal product of $X_1$ increases when $X_2$ is increased. A classic example is the pairing of teaching and research in the position of a university professor.

Thompson (1967, pp. 57-58) motivates these ideas by arguing that coordinating reciprocally interdependent positions is the most costly, or “troublesome,” and organizations operating under technical rationality can minimize coordination costs by placing similar positions “tangent” to each other. Hasenfeld (1983) points out that human service decisionmakers may not act according to the rational model. The goal of the current inquiry is not meant to settle this question, but instead to assert that if decisionmakers are rational, then cost-minimization, and behaviors associated with it, are reasonable expectations.

Stated more generally, these hypotheses would be: (1) An organization will be more effective when it groups into a single job those tasks that are difficult to measure, that are complementary inputs, and that are reciprocally interdependent; and (2) An organization will be more effective when it separates tasks that are relatively easy to measure and that are clearly delineated as sequential elements of the organization. The hypotheses generated by the multitask principal-agent and service technology approaches are consistent in this application, but they need not be in all applications: depending on the characteristics of the tasks examined, the two approaches may suggest contradictory hypotheses.

MDRC reports describe the evaluations and findings for GAIN (Friedlander, Riccio, and Freedman 1993; Riccio and Friedlander 1992; Riccio, Friedlander, and Freedman 1994), for
PI (Kemple, Friedlander, and Fellerath 1995; Kemple and Haimson 1994), and for NEWWS (Hamilton and Brock 1994; Hamilton et al. 2001).

Bloom, Hill, and Riccio (2003), Heinrich (2002), and Hill (2004) examine the relationships between office characteristics and program performance (measured as treatment minus control differences) from random assignment studies. Datasets that support such analyses are quite rare, however.

GAIN staff surveys were administered in two waves between mid-1989 and mid-1991; PI staff surveys were administered in September and October 1991; NEWWS staff surveys were administered August through December 1993.

Each level one explanatory variable is centered on its grand mean value; thus, the intercept can be interpreted as the adjusted average outcome, controlling for differences in client mix across offices (Raudenbush and Bryk 2002). To simplify the presentation, I do not show the centering notation in equation 1.

All continuous variables in the level two model were centered on their grand means. Indicator variables were not centered.

Longer periods of follow-up data are available for GAIN and NEWWS. However, for consistency and to maintain a larger number of offices in the analysis, I examine client outcomes over the two years following random assignment using information from all three evaluations.

Possible alternative measures of unified case management include a continuous measure of the line staff who are case managers, or a series of indicator variables at various points in the distribution. While the continuous measure conveys interesting descriptive information about the extent of unified case management in an office, the interpretation of its coefficient in a regression model at the office level is not as substantively interpretable as the one
A series of indicators could reflect additional nonlinearities, but too few offices in the data set with large percentages of case managers (e.g., greater than seventy-five percent) exist to test these findings in a meaningful way.

Additional model specifications included average maximum AFDC benefit in each year and state for a family of three, which provide a measure of the relative benefits of working and which vary across states and over time. For example, the yearly maximum AFDC benefit (in nominal dollars) for a family of three in 1990 ranged from $3,528 in Florida to $8,328 in California; and in 1996 ranged from $3,360 in Georgia to $7,284 in California. Because AFDC benefits are set at the state level, model specifications that include state indicator variables did not include AFDC benefit controls. The main results did not change significantly when using these alternative measures.

An alternative specification that used the percentage of unified case managers in an office resulted in a positive, but statistically insignificant, estimate in the earnings model; and no effect in the AFDC model (same as for the measure reported in the main findings). Another alternative specification with a two-variable spline showed a positive, statistically significant effect only on the highest category in the earnings model, but no differences in the AFDC model. These alternative measures of case management were not featured in the analysis for reasons discussed in note # 15. Further, the basic findings do not change substantially when office-level controls for program participation in job search, basic education, and vocational training are included for the available subset of fifty-nine offices.

Riccucci et al. (2004) find that the caseworkers’ perceptions of the tasks their managers monitor were significant predictors of their perceptions of agency goals.
In contrast, Brodkin (1997, 20) argues that the combined roles that caseworkers must play of “helper and agency enforcer” are contradictory. In the caseworker-client interactions she observed, the combined roles “skewed caseworker incentives to favor the role of enforcer, limited their ability to help, and, ultimately, led to bureaucratic coping mechanisms that minimized the expression and pursuit of client interests.” Brodkin’s research focus was not on the type of casework job designs considered in the current paper, so it is not possible to conclude whether the results reported here are consistent with her observations.

Meyers, Glaser, and MacDonald (1998) analyzed transactions involving income eligibility workers (not welfare-to-work caseworkers), but the types of interactions they identified are applicable in this setting as well.

Alternative specifications were also estimated that used less restrictive measures of specialists than those described in appendix 2. The basic conclusions hold for the earnings model, but the AFDC model results are sensitive, in particular for coefficients on specialists in the core casework tasks and specialists in conducting services. All other estimates remain stable.

The GAIN and PI surveys accompanied each number response with description (e.g., “None”) as well as the percentage range; the NEWWS survey accompanied each number response only a description, but no percentage range.

Available from the author. A combined category “interview and assign to activities” was created to reflect both (1) and (2) because the questions did not contain enough specific information on these separate activities.

Note that it is not possible for a caseworker to be both a case manager and a specialist under these measures.
REFERENCES


Riccio, James, Howard Bloom, and Carolyn J. Hill. 2000. Management, organizational characteristics, and performance: The case of welfare-to-work programs.” In *Governance*


Secretary for Planning and Evaluation; and U.S. Department of Education, Office of the Under Secretary and Office of Vocational and Adult Education.


### Table 1: Evaluation Data Sources

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of offices</td>
<td>23&lt;sup&gt;c&lt;/sup&gt;</td>
<td>25</td>
<td>27&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Number of counties</td>
<td>6</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>States</td>
<td>California</td>
<td>Florida</td>
<td>California, Georgia, Michigan, Ohio, Oklahoma, Oregon</td>
</tr>
<tr>
<td>Number of treatment group members&lt;sup&gt;b&lt;/sup&gt;</td>
<td>14,645</td>
<td>7,111</td>
<td>25,359</td>
</tr>
</tbody>
</table>

Notes:

- a. PI random assignment occurred from July 1990 to August 1991. The sample used here imposes a sample restriction to only those clients randomly assigned in 1991.
- b. The sample is restricted to females in the group randomly assigned to the treatment.
- c. Offices in Riverside County are included in both GAIN and NEWWS
Table 2: Descriptive Statistics

<table>
<thead>
<tr>
<th><strong>Client-level variables (i=47,115 females)</strong></th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total earnings in eight quarters after random assignment (1996$)</td>
<td>6,042</td>
<td>10,220</td>
<td>0</td>
<td>160,000</td>
</tr>
<tr>
<td>Total AFDC in eight quarters after random assignment (1996$)</td>
<td>7,818</td>
<td>6,358</td>
<td>0</td>
<td>55,940</td>
</tr>
<tr>
<td><strong>Client age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age younger than twenty-five years</td>
<td>0.17</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Age twenty-five to thirty-four years (omitted category)</td>
<td>0.49</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Age thirty-five to forty-four years</td>
<td>0.27</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Age older than forty-four years</td>
<td>0.07</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White, Nonhispanic (omitted category)</td>
<td>0.42</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Black, Nonhispanic</td>
<td>0.38</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.16</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Native American</td>
<td>0.01</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Asian</td>
<td>0.02</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Other race/ethnicity, nonwhite</td>
<td>0.01</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Number of children</strong></td>
<td>1.92</td>
<td>1.05</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Youngest child less than age six</td>
<td>0.42</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>High school diploma / GED</strong></td>
<td>0.56</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Received AFDC all twelve months prior to random assignment</strong></td>
<td>0.41</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Earnings in year prior to random assignment (1996$)</strong></td>
<td>2,231</td>
<td>4,589</td>
<td>0</td>
<td>80,000</td>
</tr>
<tr>
<td><strong>Ever employed in year prior to random assignment</strong></td>
<td>0.45</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Office-level variables ( j=75 offices)</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>At least 50% of line staff were case managers</td>
<td>0.36</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Specialist: core casework tasks</td>
<td>0.53</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Specialist: job development</td>
<td>0.17</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Specialist: conducting services</td>
<td>0.39</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Staff emphasis on moving clients into jobs</strong></td>
<td>0.00</td>
<td>1.00</td>
<td>-1.91</td>
<td>2.65</td>
</tr>
<tr>
<td><strong>Caseload per caseworker</strong></td>
<td>151</td>
<td>73</td>
<td>63</td>
<td>367</td>
</tr>
<tr>
<td><strong>Staff with education beyond a college degree (%)</strong></td>
<td>26</td>
<td>21</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td><strong>Staff work experience in welfare-to-work programs (%)</strong></td>
<td>42</td>
<td>24</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td><strong>Tenure (years)</strong></td>
<td>2.76</td>
<td>1.58</td>
<td>0.49</td>
<td>7.67</td>
</tr>
<tr>
<td><strong>Number of line staff</strong></td>
<td>18</td>
<td>18</td>
<td>1</td>
<td>83</td>
</tr>
<tr>
<td><strong>Average county unemployment rate during follow-up period</strong></td>
<td>7.57</td>
<td>2.85</td>
<td>3.51</td>
<td>14.29</td>
</tr>
</tbody>
</table>

**SOURCE:** Author’s calculations from GAIN, PI, and NEWWS administrative data and staff surveys.
Table 3: Client-Level HLM Estimates for Earnings and AFDC
(values in 1996 dollars)

<table>
<thead>
<tr>
<th></th>
<th>Total earnings over two years after random assignment</th>
<th>Total AFDC benefits over two years after random assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Standard Error</td>
</tr>
<tr>
<td><strong>Level One (Client Level; (i=47,115) female treatment group clients)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent Variables (measured at point of random assignment):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age twenty-four or younger</td>
<td>24</td>
<td>109</td>
</tr>
<tr>
<td>Age thirty-five to forty-four</td>
<td>185**</td>
<td>86</td>
</tr>
<tr>
<td>Age older than forty-four</td>
<td>-987***</td>
<td>163</td>
</tr>
<tr>
<td>Black, Nonhispanic</td>
<td>142</td>
<td>160</td>
</tr>
<tr>
<td>Hispanic</td>
<td>414***</td>
<td>168</td>
</tr>
<tr>
<td>Asian</td>
<td>-324</td>
<td>403</td>
</tr>
<tr>
<td>Native American</td>
<td>-943***</td>
<td>340</td>
</tr>
<tr>
<td>Other race/ethnicity – nonwhite</td>
<td>302</td>
<td>643</td>
</tr>
<tr>
<td>Number of children</td>
<td>-45</td>
<td>39</td>
</tr>
<tr>
<td>Youngest child less than six years old</td>
<td>-157</td>
<td>115</td>
</tr>
<tr>
<td>High school graduate / GED</td>
<td>2,443***</td>
<td>142</td>
</tr>
<tr>
<td>Received AFDC all prior twelve months</td>
<td>-99</td>
<td>180</td>
</tr>
<tr>
<td>AFDC Applicant</td>
<td>-7</td>
<td>173</td>
</tr>
<tr>
<td>Ever employed in prior year</td>
<td>428***</td>
<td>160</td>
</tr>
<tr>
<td>Earnings year in prior year (1996$)</td>
<td>0.89***</td>
<td>0.04</td>
</tr>
</tbody>
</table>

SOURCE: Author’s estimates from GAIN, PI, and NEWWS administrative data and staff surveys. These estimates were obtained from models estimated simultaneously with the level two variables shown in table 4. Statistical significance is denoted by: ** \(p < 0.05\); *** \(p < 0.01\).
### Table 4: Office-Level HLM Estimates for Earnings and AFDC Benefits  
(values in 1996 dollars)

<table>
<thead>
<tr>
<th>Level Two (Office Level (j = 75) offices)</th>
<th>Total earnings over two years after random assignment</th>
<th>Total AFDC benefits over two years after random assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent Variables:</strong></td>
<td><strong>Coefficient</strong></td>
<td><strong>Standard Error</strong></td>
</tr>
<tr>
<td>Constant</td>
<td>6,365***</td>
<td>228</td>
</tr>
<tr>
<td><strong>Job Design Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least half line staff were case managers</td>
<td>343*</td>
<td>174</td>
</tr>
<tr>
<td>At least one specialist in core casework tasks</td>
<td>-172</td>
<td>141</td>
</tr>
<tr>
<td>At least one specialist for job development</td>
<td>524***</td>
<td>172</td>
</tr>
<tr>
<td>At least one specialist in conducting services</td>
<td>182</td>
<td>148</td>
</tr>
<tr>
<td><strong>Control Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staff emphasis on moving clients into jobs</td>
<td>441***</td>
<td>84</td>
</tr>
<tr>
<td>Caseload size</td>
<td>-1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>Staff with education beyond college degree</td>
<td>2.7</td>
<td>3.6</td>
</tr>
<tr>
<td>Staff work experience in welfare-to-work programs</td>
<td>1.9</td>
<td>3.5</td>
</tr>
<tr>
<td>Staff tenure</td>
<td>28.0</td>
<td>56.1</td>
</tr>
<tr>
<td>Number of line staff</td>
<td>-13.2**</td>
<td>5.8</td>
</tr>
<tr>
<td>Average unemployment rate during follow-up period</td>
<td>-100**</td>
<td>44</td>
</tr>
<tr>
<td><strong>Level One variance explained (percent)</strong></td>
<td>19%</td>
<td></td>
</tr>
<tr>
<td><strong>Level Two variance explained (percent)</strong></td>
<td>85%</td>
<td></td>
</tr>
</tbody>
</table>

**Source:** Author’s estimates from GAIN, PI, and NEWWS administrative data and staff surveys. Estimates from each column in this table were obtained from models estimated simultaneously with the level one variables listed in the respective column in table 3. State indicator variables are included in the models, but not shown in the table. Statistical significance is indicated by: * \(p < 0.10\); ** \(p < 0.05\); *** \(p < 0.01\).