

Designing and Implementing an Evaluation of a National Work Support Program

Irene Y.H. Ng^a, Kong Weng Ho^b, Tharmalingam Nesamani^c, Alex Lee^d, Ngiam Tee Liang^e

^a Corresponding author. Department of Social Work, National University of Singapore, AS3 Level 4, 3

Arts Link, Singapore 117570, Singapore. E-mail: swknyhi@nus.edu.sg. ^b Division of Economics,

Nanyang Technological University, S3-1B-65, Nanyang Avenue, Singapore 639798, Singapore. E-mail:

kwho@ntu.edu.sg. ^c Department of Social Work, National University of Singapore, AS3 Level 4, 3 Arts

Link, Singapore 117570, Singapore. E-mail: swktn@nus.edu.sg. ^d Department of Social Work,

National University of Singapore, AS3 Level 4, 3 Arts Link, Singapore 117570, Singapore. E-mail:

swkleea@nus.edu.sg. ^e Department of Social Work, National University of Singapore, AS3 Level 4, 3

Arts Link, Singapore 117570, Singapore. E-mail: swkntl@nus.edu.sg.

Abstract:

Welfare reforms in the 1990s have shifted governments around the world towards financial assistance conditional on work. While large-scale rigorous research on welfare-to-work programs has demonstrated effectiveness toward employment in other countries, no such micro-level evaluation of a policy has ever been conducted in Singapore. This article describes the process of developing a large experimental evaluation of the Work Support Program, which the Ministry of Community Development, Youth and Sports started in 2006. The lessons learned from planning and implementing the research can be helpful to future researchers in negotiating long-term rigorous evaluations in an environment where collaborators lack sufficient research knowledge. Insights include ways to focus on the essentials, find alternative experimental designs, collaborate effectively, and adapt instruments across cultures.

Keywords: Singapore, Work Support, evaluation, experimental design

Designing and Implementing an Evaluation of a National Work Support Program

Welfare reform in different parts of the world has been accompanied by large scale evaluations. In the United States, randomly assigned experiments on multiple sites were able to establish that welfare-to-work programs have successfully decreased welfare numbers and pushed people out to work, although effects on incomes and well-being were less clear (Hamilton, 2002; Bloom et al. 2003; Hendra et al. 2010). Research in other countries, such as Australia, Britain, and Hong Kong, with the use of randomized controlled trials, comparison groups, administrative data, and longitudinal surveys and interviews, have also found the success of work assistance programs toward improving employment rates (Tang, 2010; Department of Employment and Workplace Relations, 2006; Department for Work and Pensions, 2008).

To our knowledge, no such rigorous evaluations of welfare policies have been published in Singapore. In this young nation of only 45 years, where its economic growth and social transformation have been so rapid that it has evolved from a third world trading post to a first world city-state with one of the highest per capita GDP in the world, policies have rolled out in quick succession. In the early years, bold social institutions such as the Central Provident Fund, the education system, and housing development were decisively and rapidly put in place to quickly build its human capital. As a small and vulnerable new nation, it was either these initiatives or suffer grave threats to its economic and political survival. Thereafter, policies and programs were incrementally added and modified with reviews, adaptations of tested models from overseas, and rejection of systems that it felt would be ill-suited for a small, young, and Asian nation. With policies rapidly rolled out, there was little time for intensive long-term evaluations.

Singapore clearly rejected the welfare state model (Peh, 2006). Learning from the huge burden of the welfare systems in the West and the disincentive problems of a steep progressive tax system that a welfare state relied on, Singapore chose a lean welfare system that invested heavily in

social development, and gave very little in direct cash to individuals. Its main social security system, the Central Provident Fund (CPF), is a defined contribution system where one pays for one's own retirement needs. Direct cash assistance is given under a few social assistance schemes, one of which is Public Assistance (PA), a stringently means-tested program for those who are not able to work and have little or no social and financial support. There is no unemployment insurance, minimum wage, or, until recently, any form of guaranteed assistance to any one with some ability to work or has a family member who can. Instead, self-reliance is emphasized. In a country lacking in natural resources, its human resource is precious and scarce. The government's message was clear: if you can work and have family who can, you take care of yourself. The government invests in building your human capital so that you can help yourself.

However, in recent years, similar to trends experienced in other matured or maturing economies, those with lower skills and educational levels are finding it harder and harder to make ends meet (Ministry of Finance, 2010; Yap, 2010). Workfare, broadly defined as government assistance for low-wage workers, was therefore introduced. In 2006, a workfare framework was introduced, which included several existing and new strategies, namely an earnings supplement, a means-tested Work Support Program, skills upgrading training programs, and job creation schemes (Ministerial Committee on Low Wage Workers, 2009).

For a country that does not espouse welfarism, embracing policies and programs to help those who work was a shift from its emphasis on individual responsibility. It is with this backdrop that the Ministry of Community Development, Youth and Sports (MCYS) set aside budget for a multi-year evaluation of the Work Support Program (WSP), one of the workfare strategies. WSP provides financial support and services to help recipients attain financial independence through employment, and provides case management to families with children aged below 18 years. In particular, MCYS commissioned the evaluation to study families which are receiving case management from the Work

Support Self-Reliance (WSSR) track of the WSP. WSSR helps families which already have a working member.

MCYS commissioned the evaluation as a longitudinal study that would follow recipients of WSSR from program entry until three years after exit. The study was a first in two ways. Its magnitude would be a first, where 800 WSSR program participants would be studied. This number might be small for other countries, but it is large for a small country with a population size of only five million (Government of Singapore, 2010). As the design of the evaluation unfolded, the study also became the first study of a government program that incorporated an experimental design. This large-scale evaluation of an early workfare program should provide important information for future policy direction in helping low-income earners.

This article describes the research development process from when the research team was first invited to tender for the project until the time of the writing of this paper, which is when the first wave of surveys and interviews were being rolled out. It brings the reader through from conceptualization to implementation of the research. The purpose is to draw insights from the process of designing a viable evaluation while bearing in mind the administrative constraints and needs on the ground. After providing a review of existing research on work support programs in the world and describing how WSP and WSSR work in Singapore, the challenges faced and how the challenges were addressed, will be discussed. The discussion will be structured according to the main design and implementation issues, namely issues of identification of key variables, timing, variability, ground realities versus rigor, translatability of instruments, and surveying low-income families. Throughout, the program being evaluated will be referred to as WSSR, but where descriptions of WSSR also apply to the overall program, WSP will be used.

Work Support Evaluations in Other Countries

Singapore's WSP was modelled after Wisconsin-Works (W-2) in the United States. It was designed after government delegations visited the United Kingdom, Hong Kong, New York, and Wisconsin to study welfare-to-work (WTW) programs. WTW programs in these countries started in the 1990s. As welfare rolls swelled in light of stagnating low-skilled wages and rising unemployment, governments began to require work in exchange for financial assistance through WTW programs. In the United States, Aid to Families with Dependent Children (AFDC), a form of welfare cash assistance, was replaced in 1996 by a flexible state-directed program called Temporary Assistance for Needy families (TANF), giving each state a block grant to spend on welfare programs and benefits (Hamilton, 2002). Welfare reform in Britain was more centralized, through a New Deal in 1998. The first of the New Deals was the New Deal for Young People (NDYP), for those aged 18 to 24 who had been claiming jobseekers' allowances for six months or more. Subsequently, other New Deals were also introduced, including a New Deal for people aged above 25, and more targeted New Deals for lone parents and disabled people (Department for Work and Pensions, 2008). In Hong Kong, a review of the Comprehensive Social Security Assistance (CSSA) resulted in the implementation of the Support for Self-Reliance Scheme (SFS) in 1997 (Tang, 2010).

In essence, the WTWs in these countries provided intensive employment assistance as well as earnings supplement in order to support work. In Britain and Hong Kong, a case manager worked with participants to target job-related assistance according to job readiness. For more work-ready individuals, job placement with or without wage subsidies are provided. For those harder to place, jobs are arranged by providers in the voluntary or public sector. More targeted help to more challenging cases also included customized services, vocational training, or basic education (Finn, 1999; White & Riley, 2002; Tang, 2010). The programs in the United States varied across states, but W-2, the program after which WSP in Singapore was modelled, also adopted the structured approach in the other two countries, from a lowest level of employment assistance without financial

aid to the highest two levels with community work and work training (Wisconsin Department of Children and Families. 2009).

The early WTW programs have since been enhanced with other workfare programs (defined as government assistance for low-wage workers) in all the three countries. In 2000, several states in America introduced Employment Retention and Advancement (ERA) programs. While the early WTW programs aimed towards gaining employment, ERAs take the additional step of working towards staying employed and advancing to higher paying jobs. The programs are slightly different at different sites, and include linkages to employers, encouragement to change to higher-paying jobs and take up training, financial incentives for retaining or moving to more highly paid jobs, counseling on job-related issues, and pre- and post-employment services (Hendra et al., 2010). In 2003, Britain also started an ERA demonstration that was offered to New Deal participants. The British version included financial incentives to retain jobs and for training, and also emergency payments to overcome short-term barriers to employment (Greenberg & Morris, 2005).

In Hong Kong, the SFS was followed by a series of other more targeted programs. For instance, intensive employment assistance projects (IEAP) were introduced in 2003 to help recent welfare recipients, as opposed to Community Work (CW), which assisted those who had been receiving welfare for a longer period. In 2002, the Ending Exclusion Project (EEP) was also launched for single parents with young children.

The evaluations of some of these WTWs were rigorous and statistically advanced. The most sophisticated were those in America, where multi-sites experiments with random assignment were conducted (Bloom et al., 2003; Hendra et al., 2010). In Britain, a randomized controlled trial (RCT) was applied to the ERA demonstration, but not the early New Deal programs (Greenberg & Morris, 2005). Hong Kong also documented an evaluation with an experimental design of the Ending Exclusion Project (EEP) (Leung et al. 2003). Besides RCT, there were also evaluations using

comparison groups (e.g. Britain: Evans et al., 2003; Hong Kong: Tang et al., 2005; W-2: Cancian & Ybarra, 2008), longitudinal surveys (e.g. Hong Kong: Tang & Cheung, 2007; USA: Johnson & Corcoran, 2003); qualitative interviews (Britain: Evans et al. 2003; W-2: Alfred, 2005), and also interviews with service staff (W-2: Martin & Alfred, 2002). In Britain, with the availability of publicly available administrative micro data, innovative techniques such as propensity score matching and regression discontinuity were employed on administrative data, exploiting natural variations arising from age cut-offs, and timing differences in different locations (e.g. Blundell et al., 2004; De Giorgi, 2005).

So far, the WTW programs have been considered successful in reducing welfare caseloads and increasing employment rates (Hamilton, 2002; Tang, 2010; Department for Work and Pensions, 2008). Reasons for the success were partially attributed to the mandates (Bloom et al, 2003), wage subsidies (Reenen, 2003), and placement assistance (Blundell et al, 2004), although some have argued against the effectiveness of welfare-to-work through community or public works (Crisp & Fletcher, 2008). While improving employment rates and decreasing welfare reliance, little gains in incomes have been found (Hamilton, 2002; Cancian & Ybarra, 2008; Michalopoulos, 2005). This is so even for the ERAs (Hendra et al., 2010; Riccio et al., 2008; Miller et al., 2008), although it might be too early for the results on securing better paying jobs to show.

Identifying effective and ineffective program features is difficult as WTW programs are complex interventions that comprise several “interconnecting parts” (Campbell et al., 2000: 694). For complex interventions, RCTs are limited because participants either receive the program or not, resulting in a “black box” problem where the impacts are measured but how the program affects the impacts is unknown (Greenberg & Morris, 2005). Three alternatives could strengthen simple program-control RCTs. First, mixed methods could be used, where quantitative methods are complemented by qualitative interviews to tease out the “hows” (Campbell, 2000). Second, “dosages” of the different components should be identified and measured (Vlaming et al., 2010).

Third, RCTs could randomize on program components, but this would multiply the administrative burden and required sample size by the number of randomized groups (Greenberg, 2005).

None of the existing WTW studies seem to have randomized on components. However, from the results of experimental groups relative to control groups at different sites with different program characteristics in the United States, successful program features could be identified. For example, the initial experiments found that job-first were more successful than education-first sites (Hamilton et al. 2002), and that greater gains in income were experienced by sites that gave more generous earnings supplements (Greenberg et al. 2009). The ERA RCTs, in addition, suggested greater effectiveness from financial incentives mixed with services and changing to jobs where participants are matched with a higher paying employer (Hendra et al. 2010). Combinations of quantitative and qualitative findings have also highlighted the challenge of helping people find and retain jobs, especially for a population where many face multiple barriers to work (Johnson & Corcoran, 2003; Finn 1999; Freud 2007; Roccio et al, 2008).

The Work Support Program in Singapore

In Singapore, WSP is a national program under the Ministry of Community Development, Youth and Sports (MCYS). It is means-tested, where applicants are interviewed and assessed before qualifying for WSP. In general, national financial assistance schemes are targeted at households in the bottom 20th income percentile. The criteria for WSP include household incomes below S\$1,500 (about US\$1,200) per month, a needs test, little or no savings, inadequate family and community support, and demonstration of willingness to take steps to become self-reliant (Ministry of Community Development, Youth and Sports, 2010).

Aimed at helping recipients find employment and achieve financial independence through interim financial support and other assistance, recipients receive monthly cash assistance which is calibrated based on the unique circumstances of the family. In many cases, clients also receive

assistance with the paying of utilities and conservancy charges. The type of assistance needed is discussed through an action plan, where any issues that might prevent recipients from working or achieving self-reliance are put into action steps to be addressed. For households with children below 18 years of age, case management is also provided. Compared to non-case managed cases, clients are regularly monitored to assess their progress in taking the action steps.

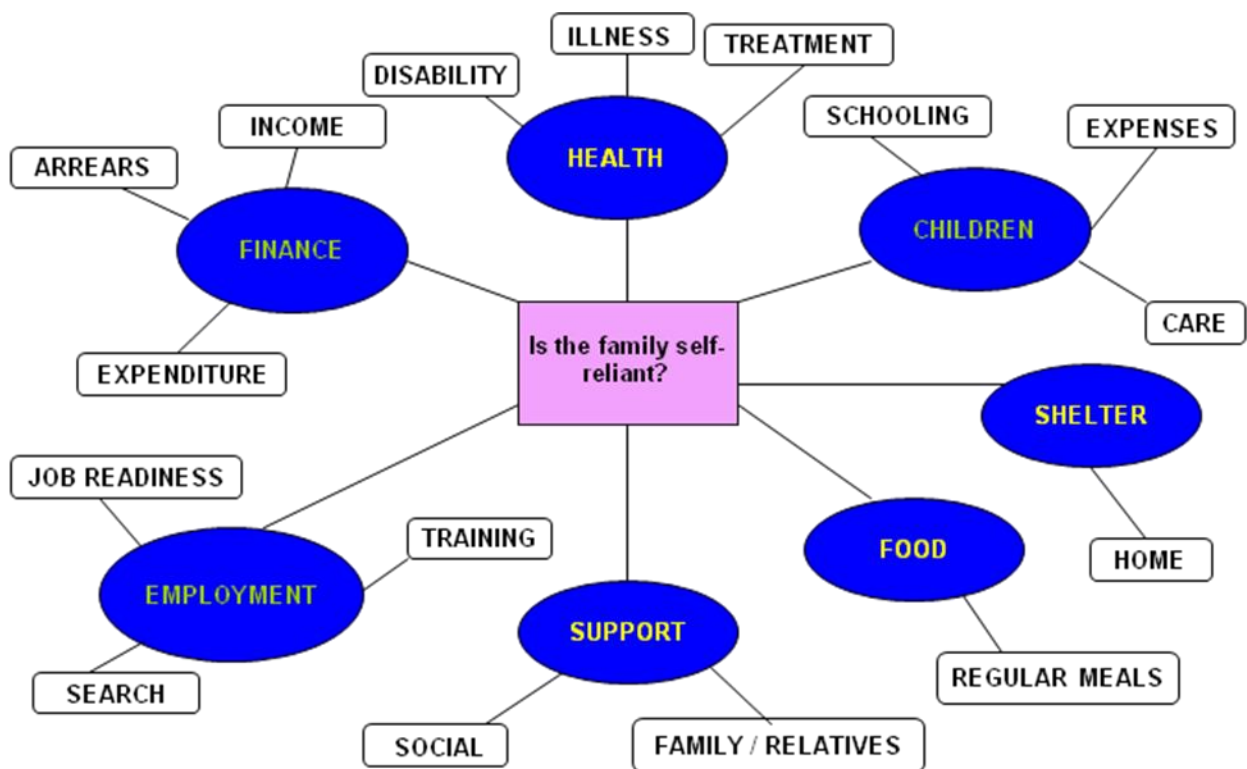
WSP has two tracks. Work Support Employment (WSE) is for households with a work-capable member who is unemployed, where assistance can be provided up to twelve months to help the person become gainfully employed. Households with a working member yet struggling to make ends meet are placed on Work Support Self-Reliance (WSSR), which is a longer track of up to 24 months. Increasing financial independence of WSSR recipients often entails encouraging a second member to work and/or improving management of household expenses.

WSP is administered by five Community Development Councils (CDCs), which are distributed according to geographical boundaries so that the whole country is covered: North-West, North-East, South-East, South-West, and Central Singapore. As the names imply, the CDCs are tasked to be the community points for residents in the respective geographical constituents. The CDCs are main hubs of community development. They are centrally located within their regions, and near subway train stations and major bus interchanges. As such, providing government assistance through them makes them accessible to the community. Besides WSP, other MCYS assistance programs such as the Public Assistance (PA) scheme and subsidies for childcare, kindergarten and student care services are also delivered through the CDCs.

South-East CDC piloted a case management framework, called the South-East Social Services Assessment Methodology, or SESAME for short. This framework was subsequently adopted by the other CDCs. Figure 1 maps the seven factors and their sub-factors that case managers are to work on with their clients within the framework, which include employment, finance, health, children, shelter, food, and social support.

Singapore’s WSP is similar to the WTW programs in other countries in terms of employment assistance through case management and financial aid. However, its case management is structured in terms of challenges faced by applicants instead of the degree of employment assistance needed. No guaranteed voluntary or public sector jobs are provided. To date, job referral and skills upgrading programs are expanding, and might become more similar to the structured employment assistance in the other countries. Future enhancements to WSP might also see it stepping up to something like the ERAs.

Figure 1. South-East Social Services Assessment Methodology (SESAME)



In the current version of WSP, the Singapore evaluation needs to be clear about the differences between WSP and the WTW programs in other countries, and also what the main aims and features of WSP are. In the first place, the starting point of WSP is different. Whereas the programs in the countries discussed are welfare-to-work, Singapore’s long-term financial assistance has been and still is strictly for elderly or disabled people who are not able to work and who do not

have family support. WSP was started to cater to the needs of a new group that are work capable but need short-term financial assistance. In this sense, reducing welfare caseloads is not an objective of WSP. With its aim being attainment of financial independence, sustained employment and earnings instead are its aims. The commissioned evaluation studies in particular WSSR families with children below 18 years of age.

Developing the Evaluation: From Conceptualization to Implementation

Table 1 marks the milestones in the development of the research. The study was proposed by MCYS in 2006, when WSP was launched. Soon after, a pre-pilot study was conducted to understand how the program works and test the quality of the administrative records. It was important to ensure that the key outcomes were being captured accurately both for operational and research purposes. In discussion with MCYS, the key measures of financial independence were identified as earnings, employment, savings, and debt. The variables should reflect correctly the financial gains as both key performance indicators and research instruments.

After the pre-pilot study, the research team conceptualized the research and submitted the proposal to MCYS. It then started process evaluation. In 2009, once the internal approvals and administrative arrangements in the university were cleared, the research design was re-conceptualized, evaluation of the program processes was updated, and survey instruments and interview guides piloted. To date, in March 2011, these have been completed and the pre-WSP wave of surveys and interviews are underway. Some post-WSSR surveys have also been piloted and started. The rest of the study follows until the last wave and analysis in 2015.

Table 1. Work Support Evaluation Milestones (as of March 2011)

Nov 2006 – Feb 2007	Pre-pilot study: <ul style="list-style-type: none"> • Initial meetings with CDCs
---------------------	---

	<ul style="list-style-type: none"> • Replication survey to test quality of administrative data
Apr 2007	Conceptualization of research design
June/July 2007	Process evaluation and job analysis
April 2008	Approval by MCYS of the budget for the WSSR evaluation Announcement by MCYS of the Request for Proposal (RFP)
August 2008	Approval by University for consultancy collaboration Request for Proposal submitted by research team
November 2008	Consultancy agreement signed by MCYS and the University
Mar 2009	Ethics approval granted by the university Institutional Review Board
April 2009	Research account in the university created Administrative issues sorted out for research to start
May-Aug 2009	Re-conceptualization of research design Design of survey questionnaires and interview guides
Sep 2009 – Jan 2010	Pilot of pre-WSSR wave
Apr 2010	Pre-WSSR wave launched

Addressing Challenges of a National, Longitudinal Program Evaluation

From the outset, MCYS wanted the evaluation to be a longitudinal study that traces WSSR recipients from the time they begin the program until several years after they have left WSSR. However, while it was open to using recipients of another program as a comparison group, it did not permit an experimental design where recipients were randomly given WSSR or not. After all, WSP as a policy had already been implemented by the time the research negotiations began.

The challenges encountered in preparing the evaluation can be categorized into six types, with three related to design issues, two related to implementation issues, and one related to both design and implementation (Table 2). First, it was important to identify the specific aims and key

factors of the research given the complexity of the program. Second, as a national program where policy-makers’ decisions were made before research design, flexibility in research design was constrained. Third, a multi-years and multi-sites research requires that issues of variations through time and across CDCs be addressed. Fourth, lack of access to administrative data constrained planning of the research design. Fifth, as a multi-agency collaboration, the research needs to sensitively handle different parties with different work scopes and aims. Sixth, that respondents have low income and low education poses particular challenges with designing and conducting surveys and interviews, and also with following up with them.

Table 2. Six Types of Challenges Related to Research Design and Implementation

Research Design	Research Implementation
Identifying key outcomes and factors	Variation through time and by site
Policy made before research design	Multi-agency collaboration
Lack of access to administrative data	
Studying respondents with low education and low earnings	

Identifying Key Outcomes and Factors

The main research aim was to evaluate the effectiveness of WSSR. However, as discussed previously about complex interventions, WSSR is a complex intervention with several related program components, making it a challenge to identify what about WSSR works. Hence, the first challenge of the research was the identification of key variables, and in turn, a questionnaire that captures the various factors and outcomes of the study succinctly yet comprehensively. First and most importantly, WSSR is about financial independence and therefore the aim of assessing socioeconomic status was given priority. The key variables were identified as sustained employment and earnings, savings, and reductions in debt. These variables must be collected and collected well.

Hence, most energy was spent on getting accurate measures and good questions that illicit the right answers for these key variables.

Next in importance were the various factors that prevent self-reliance. As there were many different possible factors, we focused on finding the most efficient measures of the factors from established surveys that studied similar populations, such as the Panel Study of Income Dynamics (PSID)¹ and the Women's Employment Study (WES)². We also used the SESAME framework (see figure 1) to guide the selection of factors. Putting SESAME together with literature from overseas provided an extensive yet lean set of instruments. All the measures of outcomes and factors were to be collected at baseline, exit from WSSR, one year, two years, and three years after WSSR.

WSSR as a factor is, of course, the main determinant being evaluated. Four main components of WSSR were identified: the monthly cash amount, the duration of assistance, the financial literacy workshop, and case management. Hence, the post-WSSR surveys and interviews were designed with specific questions on these components of WSSR.

To further enrich the data for this evaluation of a complex intervention, a mixed method was adopted, where out of the 800 survey respondents, 50 would be selected for in-depth interviews. The mixed method made the study generalized yet nuanced, and the importance of this goes beyond program evaluation. Beyond finding out the effectiveness of WSP, as a longitudinal study with rich data, this mixed-method study would also help to understand the lives of the working poor in Singapore.

Policy Made Before Research Design

¹ The PSID is the longest-running American longitudinal household survey that has been used widely in research on economic and social behaviour. It has followed respondents since 1968 (Panel Study of Income Dynamics, n.d.).

² The Women's Employment Study (WES) is a panel study of female welfare recipients who experienced welfare reform in Michigan. It followed women for five waves, from 1997 to 2003 (Gerald R. Ford School of Public Policy, 2010).

The second challenge the research team encountered was that the program had already been implemented by the time the research team was brought into the picture. WSP started in 2006, but the research was conceptualized in 2007. Research involvement after program implementation meant that applicants could not be randomly assigned for participation in the program. With the policy announced and eligibility criteria published, the service providers (CDCs) could not ethically deny service to anyone who met the criteria or accept anyone who did not. The transition could have been an opportune time to evaluate a pilot of the new program before it went to scale. An intervention group could have been placed on the new program while the control group continued to receive the old program. However, up to that point in time, national policies and programs were not conceived with experimental pilot research in mind. Singapore has had a practice of piloting programs, but evaluations were short-term and simple, mainly to show sufficient take-up and to smoothen out rough edges.

MCYS wanted the research to find out the effectiveness of WSP, but without an experimental design, it was impossible to answer this research question. Selection bias is clearly a problem, and it can be manifested in at least two ways in Singapore. First, the program is set up such that recipients agree to make efforts towards self-reliance (Ministry of Community Development, Youth and Sports, 2010). It might be the case that more motivated individuals apply for WSP, hence biasing the effectiveness of WSSR upwards. It will be difficult teasing out what is due to participants' motivation and what is due to WSSR. In the reverse direction, in a country where self-reliance is a mindset not just of policy makers but of the general population, it might be those who have exhausted their networks and resources that come forward for help. The profile of the majority of applicants having high amounts of debt signals the dire state of applicants. In our findings from a pilot survey of 50 respondents, about half the respondents indicated owing too much as a main reason for approaching a CDC for assistance, and 88% had some kind of debt. The effect of applicants approaching the government for help only when they have no where else to turn to

implies that applicants of WSSR might experience more difficult challenges than other needy families. This biases the effect of WSP downwards.

After the research officially started in May 2009, the idea of an experimental design was revisited. Unable to randomize selection into the program, randomization of the receipt of different versions of the program was explored. The different versions would differentiate on key components of WSP. During a meeting with the managers of the CDCs and MCYS in June 2009, the research team explained what randomization is and why it is important. The representatives showed initial agreement to the idea, and actively gave suggestions to improve it. The meeting was followed by individual visits to the CDCs and follow-up e-mails and telephone conversations to confirm their support and seek further inputs. None of the CDCs resisted randomization, but were in fact active in discussing how to overcome procedural problems of randomization. This turn of events that allowed for randomization by two key features of WSSR not only addressed selection bias, but also provided an additional benefit of measuring the effects of the amount and duration of assistance. The latter would not be achievable by a simple randomization of participation.

The following was finally decided on after ironing out various procedural issues. The four key components of WSSR that made the difference to WSSR participants were identified as the monthly amount of financial aid given, the duration of assistance, the financial literacy workshop, and case management. Randomization of the workshop was logistically too difficult and was dropped. Case management was the key differentiating feature of WSSR and WSE, and therefore could not be denied to WSSR participants. Besides, **as noted previously from Greenberg (2005), randomizing on many components would multiply the administrative burden and the required sample size.** The post-WSSR surveys must therefore include specific questions about the workshops and case management. The qualitative interviews must also pursue reactions of respondents to these two components of WSSR.

The research was therefore left with randomization of the amount and duration of assistance. For the duration of assistance, since participants of WSSR were reviewed every six months, a small group of 200 who were about to complete the program would be randomly extended for another six months. Although extending for six months is a fixed duration for recipients who would have been in the program for different lengths of time, recipients were similar at that point in terms of being deemed ready to graduate.

The snag was that extension countered the standard operating procedures for when a case should be closed. MCYS needed to allow for these cases to remain open despite meeting the criteria for closing. MCYS also needed to give assurance to the CDCs that any adverse results from these cases during extension would not be counted against the case officer.

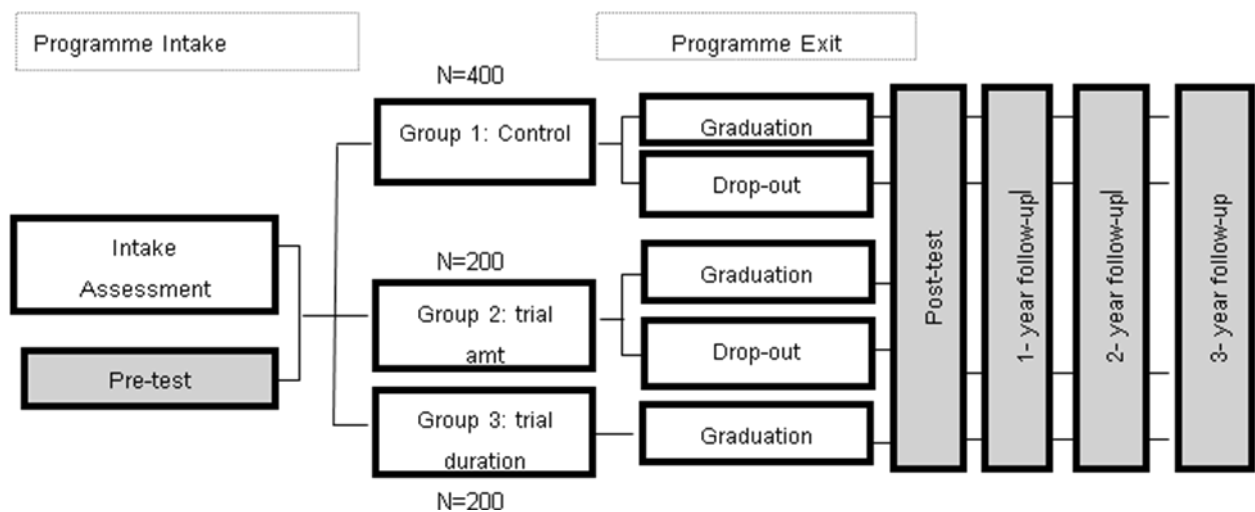
For the amount of assistance, another 200 would be randomly assigned to receive an additional amount every month. The amount had to be big enough to make a difference yet not too generous. Applying a matrix that the CDCs used to decide the monthly amount to disburse to families based on their household earnings and number of dependents, an additional monthly amount of S\$50 (about US\$40) per dependent was recommended. The S\$50 is to some extent arbitrary because at that time, the researchers had no access to data on average amounts disbursed or average family sizes. However, the S\$50 was considered a middle-ground in the matrix used by the CDCs, and it was estimated to provide for the equivalent of one additional hour of tutoring per month or an additional slice of fruit per day.

The resulting research design is made up of two treatment groups and one control group, with the hypothesis that financial independence improves with larger amounts of assistance per month and a longer duration of help (Figure 2). Being in one of the two experimental groups is akin to being on a different program, as recipients are typically sorted to several different possible programs depending on their presenting issues when they approach a CDC. The difference with

these two groups is that they are randomly sorted into the two experimental groups instead of placed by a set of criteria.

At entry to WSSR, participants are contacted for the first wave of survey and interview. At this point, 200 are randomly assigned into the trial amount group and 50 purposively selected for the qualitative interviews. . The remaining 600 continue their program with no modification until the time when they are about to complete the program. At this time, 200 are randomly assigned to the trial duration group. Upon completion of WSSR, the post-WSSR wave of survey and interview would be implemented. The next three waves follow at one year, two years, and three years after exit from WSSR. While on WSSR, subjects are also contacted every six months by telephone for quick updates in earnings, employment and contact details.

Figure 2. Work Support Research Design



Beginning in April 2010, all new WSSR recipients are being recruited into the study until 800 is reached. This is estimated to take about a year. In this way, the research is getting the annual population, and not a sampling of WSSR recipients. This means that statistical power is not as much an issue despite the small sample size that attrition will result in at the end of five surveys. The study also had budget and time constraints that prevented it from stretching out data collection to more recipients or more than a year.

Variation through Time and across CDCs

Variations through time could arise from two sources, firstly the lag time between research conception and the actual start of the research, and secondly the long duration of the research. With more than two years between the initial conceptualization and the start of the research, many changes had occurred. From how household and per capita incomes were calculated, to the types of programs offered to recipients, there were much tweaking and refinements for standardization across service providers. On one hand, such standardization and refinements were good for research because they resulted in cleaner measurements of outcomes. The effects could be more clearly attributed to policy and not because of slight differences in how the program was implemented from CDC to CDC and from one staff to another. On the other hand, programs might evolve to something rather different from onset. This could result in the initial research objectives becoming out of sync from the actual implementation. Two examples below of changes experienced so far illustrate how program objectives and elements might change. The research must keep pace with such changes.

First, financial literacy workshops became a prominent feature in WSP and case workers felt that it was important in helping participants become financially independent through better money management. It therefore became important for research to specifically find out the effects of these workshops. Second, if the research had to be modified according to changes in the past two years, it will have to take into account even more changes during the full seven years of the research. The research will need to withstand the tests of macroeconomic, social, and political trends, as well as changes in policy. These are bound to happen, as has already transpired from the pilot to the start of the longitudinal study. The pilot was conducted during the peak of the global financial crisis in 2008, when the number of applicants swelled. At that time, it was uncertain whether an upturn would be in sight soon. However, by the time the longitudinal study started, the Singapore economy was on the recovery, and application numbers had declined. Originally, we were worried that the evaluation

would be starting at a time when many WSSR recipients were only in transitory need due to temporary lay offs, and not the original targets of WSSR case management. The latter were hypothesized to experience multiple barriers to employment requiring assistance using the case framework. The longitudinal study is timely now, to follow through with a needy group that the policy had been started for, with the understanding that workfare needed to be a more holistic source of help for this group.

Besides variations through time, differences across CDCs and issues on the ground also had to be documented. First, the profile of residents and therefore applicants of WSSR are different across the CDCs. In 2009, the WSSR program in South-West was the largest, serving twice the number of clients compared to the next largest program (North-West) and five times the number in the smallest program (South East). Central Singapore had the youngest profile of clients, and Northeast served a higher proportion of Chinese than the other sites³.

Second, some work processes differed by CDC. Four years after the program was started, much of the implementation and processes have been standardized across CDCs. Regular meetings between the CDCs and MCYS have enabled information sharing and clarifications. Nevertheless, there remain some variations in procedures. For example, case officers in different CDCs had different caseloads and held different combinations of portfolios besides WSSR. South-West CDC, which had the largest number of WSSR clients in 2009, naturally had the highest case loads. Here, case managers focus on handling case-managed clients only, whereas in North-East CDC, case managers handle all the various kinds of financial assistance schemes disbursed by the CDC besides WSP. Part of the reason might be the respective CDC's organizational structure, differences in approval processes, and size of intake of applicants for the WSP and other assistance schemes.

In sum, there were differences between the CDCs in terms of size, profile of clients, and portfolios of case managers. Such variations were addressed by soliciting and documenting regular

³ MCYS does not permit publication of demographic data on recipients and caseload data of WSSR.

updates from the CDCs on processes and profiles. The variations could be used as program features to explain different outcomes by CDC. The problem of staff turnover also meant the need for constant re-orientation of ground staff and liaisons of the research project.

Access to administrative data

Lack of access to administrative data was the fourth challenge. It had been agreed that administrative data of WSSR would be extracted for the evaluation project. However, WSP's data system was a complicated system that was linked to other programs provided by the network of CDCs. Access was supposed to be granted before the actual research, as part of a process evaluation and job analysis that the team had conducted in preparation for the longitudinal study (see Table 1). However, as the administrative data was still being sorted out when the first wave of data collection started, the team decided that basic background and program information were to be collected through the longitudinal surveys and not to rely on the administrative data. For example, summary statistics of average amounts disbursed and average duration of assistance were not available. Even less is known of information by profiles such as family size or the type of barriers to employment. This posed challenges in deciding amounts and durations for randomization, and researchers had to estimate based on meetings with CDCs and guidelines in the CDCs' standard operating procedures (SOPs). Overall, extraction of the data became a complicated multi-agency affair that required many rounds of negotiations with the external vendor contracted to manage the data system.

Multi-agency collaboration

The fifth challenge, then, was the process of multi-agency collaboration. It has been challenging dealing with many different agencies, in an environment where external evaluations of this nature have never been done before. In addition, the kind of program that was being evaluated was also new, leading to a sense of uncertainty among all parties. The above factors of working across several big agencies and uncertainty from doing new research on a new project had led to the delays noted earlier. At several junctures, procedures required cross-checks between and within the

different agencies involved. However, the newness of everything also created much openness in all parties to consider options and work through them together.

Good working relations between MCYS, the CDCs, and the research team were instrumental to achieving the experimental research design discussed above. So was the willingness by all parties to be open to new ideas. The rigor and extent of the research was unheard of in Singapore prior to MCYS initiating such a research. In engaging an academic team to lead the research, they trusted the team to take it to a level that would help answer their policy questions. On our end, we made concerted efforts to reach out to the service providers. First, besides a process evaluation in the beginning of the research, continual updates with each CDC have been important to keep up with any changes within and across the CDCs. Second, the design should not be burdensome to staff. While this was not a pure action research, principles of participatory evaluation have been helpful to frame our approach to collaborative research. The research team was keenly aware that the success of the evaluation would be limited if we did not take a participatory approach, especially for a large-scale research that spanned multiple years, and which we wanted to be as rigorous as possible.

Indeed, as we collaborated with MCYS and the CDCs, we were mindful of the need to keep every staff who is involved in WSSR updated on the need for research and also the reason for the research design. We also found it helpful to inform them about findings in poverty research and from overseas evaluations. We needed to make sure every staff was adequately informed because (1) ground staff are not research-oriented and lack research knowledge; (2) staff turnover is high; and (3) it serves as reminders about what their work is for. While educating, we also sought inputs.

The CDCs have been responsive and actively give thoughtful feedback and suggestions. This process of communication is facilitated by each CDC appointing one liaison officer for the research. The liaison officer provides to the research team the case information of recipients, but all random assignment, surveys, and interviews are done by the research team. We also keep MCYS in the loop on the research progress and discussions with the CDCs. MCYS plays a pivotal role in facilitating our

communications with not only the CDCs, but also other agencies we need to deal with. This triangle of collaboration has resulted in a rigorous yet practicable evaluation that we hope also disrupts the CDCs' work processes minimally. The pains in engaging the CDCs has also yielded a common understanding of the research design and confidence in the survey instruments.

An example of the collaborative process is the ethical and implementation issues raised that might not have surfaced had we not engaged the CDCs. While all parties were quickly assured that there were no ethical concerns of giving participants in the treatment groups more help than those in the control group, since the policy allows them to give more, there was active discussion on the ethical and practical problems of whether to inform participants about the additional help. Interestingly, different people had opposite solutions even though the concerns were the same. One practical concern was that WSSR recipients who got wind of others getting more per month would also demand for more. One group's suggestion was not to inform research participants of the additional help, so that chances were that only a few people would figure out the difference. WSSR quantum was decided based on a complex matrix of household income, size and needs anyway. However, the opposite view was that recipients would demand answers if they figured out the difference and were not told of it. In addition, how would the officers explain to recipients who were supposed to have met their criteria for completion, but yet continue to be given help? There was also concern over whether it was ethical not to inform people that they are being treated differently. In the end, transparency was preferred and in the letter informing recipients about the study, participants were told which group they were in without information on how much. Although there was concern over placebo effects, it was felt that the knowledge of receiving additional help was no different from the introduction of a new policy that begins to help new groups of needy families in Singapore. This would be like extending a new form of the program to a randomized group of participants.

Studying Respondents with Low Income and Low Education

A final challenge for the evaluation resulted from the difficulties of studying subjects with low income and education. First, in multi-lingual Singapore, many people from lower income groups do not understand English and their vocabulary in the local languages is also limited. From a pilot survey conducted with 50 respondents, we found that nuanced differences in wordings sounded repetitive to some respondents (e.g. differences between nervousness, irritability, and anxiety in different questions in a depression scale adapted from America). The abstraction in some questions was also difficult for some respondents to understand, in particular scales which required subjective evaluations of frequency or the extent of agreement. Fortunately, many of these issues could be resolved after identifying them at the pilot study, and training interviewers to provide careful explanations of the problematic words and scales using descriptions and examples. Most respondents got the hang of responding to scaled instruments after a while.

In designing survey instruments, the full sets of scales from surveys of similar populations were used in the pilot. Then if necessary, the instruments were modified after factor analysis, interviewer feedback and research team discussion on their cultural relevance and meaning fidelity after translation. Modifications included dropping an instrument totally, reducing a scale to a few key items, and change in wording. An example of a reduced scale was the Lubben Social Network Scale (LSNS) (Boston College, n.d.) to measure the extent of social networks or isolation experienced by recipients. Social support is one of the factors in the SESAME framework (Figure 1). We had used the revised LSNS-12 with six questions each on family members and friends respectively. However, many respondents still found the questions repetitive, and the scale was further reduced to the shortest version, LSNS-6, with three questions each on family members and friends respectively. The LSNS-6 was adhered to, with what we felt was the barest minimum in capturing the various dimensions of social network, namely contact, confiding, and practical help. The factor loadings and internal consistency scores, although lower than the LSNS-12, were satisfactory ($\alpha=0.74$). Wordings in this set of questions were also changed to better reflect the living arrangements in Singapore. To

the original introduction “considering the people to whom you are related either by birth or marriage” was added “not living with you”. This was because many intergenerational and extended family members in Singapore live together in land-scarce Singapore. In summary, to adapt instruments from overseas, we followed standard statistical procedures but also injected our knowledge of the local context.

The second difficulty of studying low-income participants was an implementation issue. Low-income respondents are difficult to follow-up on due to their high mobility. Many live in rental flats or with relatives and do not have a permanent home. Many also cannot afford a landline or a mobile telephone subscription. They use prepaid cards and hence their contact number would change every few months. However, the study wanted to follow uncontactable as well as compliant recipients. Leaving them out would bias the sample towards more compliant recipients if those who become uncontactable do so because of difficulty keeping up with the WSSR requirements, and bias towards more dependent recipients if uncontactable subjects drop off because they managed to find other means of sustenance on their own.

The following were put in place to minimize attrition. First, contact information of three people closest to the respondent, including an address, were requested. This gave more contact points to the respondent than the CDCs had. Second, while the industry standard for surveys was to call the respondent three times before giving up, we set a higher standard of calling five times at different times of the day and different days in the week, over two weeks. This was based on an assumption that respondents might have long and unusual working hours which made them uncontactable at expected times. Third, we gave a token for participation in the form of a cash voucher to a local supermarket, as incentives were needed to keep them on the study. Fourth, respondents would also receive a phone call from the research project every six months while on WSSR, and thereafter receive greeting cards and an occasional telephone call. These serve to remind them of the survey, and update contact information, if any. A linear projection of 20 to 25 percent

attrition rate for each wave was made so that starting with 800 would yield 300 cases by wave 5, when respondents have left WS for three years.

Lessons Learned

Reflecting on our experiences since MCYS initiated the evaluation study in 2006 to when the longitudinal study began, the challenges encountered were addressed in creative ways that resulted in a fairly robust research design. The lessons learned from the experiences come with several caveats. First, one needs to bear in mind the limitations of the analysis of the research process so far. The design has been formulated using limited information, since the research team did not have full access to administrative data to make decisions based on program statistics. Another limitation is that while we are confident of the research design and survey instruments, the research is in its early stages, and we do not know if the design will work or whether the attrition rate will hold up. The next caveat is that our lessons might not apply in another context where resolve and commitment from the contracting partner is weak. In our case, the Singapore government was very committed to the research and invested heavily in it.

Despite these caveats, we believe that our experiences are relevant and helpful to others embarking on long-term rigorous evaluations in environments where research knowledge is rudimentary. The insights can be helpful to parties engaging in practitioner-led research of complex interventions. Both researchers like us who are engaged to conduct the research as well as the contracting party, such as a government unit or service provider, can learn from the challenges and resolutions we have described in this article. We summarize our lessons into four main points:

1. Maintain focus on the essentials. This point is perhaps the most important. For a large evaluation with many dimensions to consider, it is easy to get lost in details and be swarmed by

the sheer scale and complexity. It has been important to keep the big picture in mind and return to the motivation of the research, which is to understand this emerging class of the working poor in Singapore, and to learn how to better help them. Hence, be it the research aims, the experimental design, the survey instruments, or the working processes, we kept returning our focus back to these basic motivations. The randomization focused on the key components of not only WSSR, but also what would be key in any program to support work. Hence, even if WSSR is replaced with some other program, the questions of whether helping people with more money and for a longer time would still be an important research question. The survey questionnaire ensured that the main outcomes were asked in detail and accurately, and the key factors all covered efficiently. The work processes also were focused on achieving clean measures of key outcomes and factors with minimal disruption.

2. Look for alternative ways to randomize. For policy level interventions, in particular, there are often several separate programs or different main components within a program that one can randomize, even if one is not able to randomize at the policy level. There are many variations to tap on, but much investigation work needs to be done to discover these variations. One needs to also balance the number of components on which to randomize with minimizing the administrative demands and maintaining sufficient sample size for statistical power.
3. Engage in participatory research. This point relates to the challenges of documenting variations and engaging in multi-agency collaboration. In an environment that is new to research, working the ground through participatory research is essential. It informs the uninformed and befriends the unfriendly.
4. Existing instruments can be adapted. While being mindful of cultural relevance of instruments from overseas, and spending much time discussing among ourselves and getting feedback from interviewers, we have been pleasantly surprised at the universality of human conditions. The

instruments have held up well to statistical scrutiny as well as garnered sufficiently robust variations from respondents that are difficult to survey.

Acknowledgements

We gratefully acknowledge the Ministry of Community Development, Youth and Sports for their permission and comments to this article. We also thank the staff of the Community Development Councils for their inputs. The opinions expressed in this paper are, however, our own and do not reflect the positions of the Ministry or Councils.

References

- Alfred, M. V. (2005). Does Wisconsin Works work? Perspectives of participating women and their employers. *Journal of Family and Economic Issues*. 26 (3): 345-370.
- Bloom H.S., Hill, C. J., & Riccio, J.A. (2003). Linking program implementation and effectiveness: lessons from a pooled sample of welfare-to-work experiments. *Journal of Policy Analysis and Management*, 22 (4): 551-575.
- Blundell, R., Costa, D., Meghir, C. & Reenen, J.V. (2004). Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association*. 2(4): 569-606.
- Boston College. (n.d.). *The Lubben Social Network Scale Homepage*. Retrieved July 25, 2010, from <http://www2.bc.edu/%7Enorstraj/default.html>
- Campbell, M. et al. (2000). Framework for design and evaluation of complex interventions to improve health (Education and debate). *The British Medical Journal*. 321, September 16: 694-696.
- Cancian, M. & Ybarra, M. (2008). *The Earnings and Income of Wisconsin Works (W-2) Applicants*. for the State of Wisconsin, Department of Children and Families, September 2008.
- Considine, M. (1999). Markets, networks and the new welfare state: employment assistance reforms in Australia. *Journal of Social Policy*, 28(2): 183-203.
- Crisp, R. & Fletcher, D.R. (2008). A comparative review of workfare programmes in the United States, Canada and Australia. *Department for Work and Pensions Research Report No. 533*.
- Department of Employment and Workplace Relations. (2006). Customised assistance, job search training, work for the dole and mutual obligation – a net impact study. *Evaluation and Programme Performance Branch, EPPB Report 1/2006*.

Department for Work and Compensations. (2008). *Transforming Britain's labour market: ten years of the New Deal*.

De Giorgi, G. (2008). Long-term effects of a mandatory multistage program: the New Deal for Young People in the UK. *The Institute for Fiscal Studies Working Paper 05/08*.

de Vlaming, R., Haveman-Nies, A., van't Veer, Pieter, & de Groot, L. C. (2010). Evaluation design for a complex intervention program targeting loneliness in non-institutionalized elderly Dutch people. *BMC Public Health*. 10 (552): 1-9.

Evans, M., Eyre, J., Millar, J, & Sorre, S. (2003). New deal for lone parents: second synthesis report of the national evaluation. *Prepared for the Department for Work and Pensions*. University of Bath, Centre for Analysis of Social Policy.

Finn, D. (1999). Job guarantees for the unemployed: lessons from Australian welfare reform. *Journal of Social Policy*, 28(1): 53-71.

Freud, D. (2007). Reducing dependency, increasing opportunity: options for the future of welfare to work. *An Independent Report to the Department for Work and Compensations*.

Gerard R. Ford School of Public Policy (2010). The Women's Employment Study. University of Michigan. Retrieved on July 27, 2010, from <http://www.fordschool.umich.edu/research/poverty/wes/>

Greenberg, D. & Morris, S. (2005). Large-scale social experimentation in Britain: what can and cannot be learnt from the employment retention and advancement demonstration? *Evaluation*. 11(20): 223-242.

Greenberg, D., Deitch, V. & Hamilton, G. (2009). Welfare-to-work program benefits and costs: a synthesis of research. Manpower Demonstration Research Corporation, Feb. Retrieved July 20, 2010 from: <http://www.mdrc.org/publications/511/execsum.pdf>.

Hamilton, G. (2002). Moving people from welfare to work: lessons from the national evaluation of welfare-to-work strategies. Manpower Demonstration Research Corporation. Retrieved July 20, 2010 from: <http://www.mdrc.org/publications/52/full.pdf>.

Hendra, R., Dillman, K., Hamilton, G., Lundquist, E., Martinson, K. & Wavelet, M. (2010). The employment retention and advancement project: how effective are different approaches aiming to increase employment retention and advancement? MDRC, April. Retrieved February 21, 2011, from http://www.acf.hhs.gov/programs/opre/welfare_employ/employ_retention/reports/different_approaches/era_core.pdf.

Johnson, R.C. & Corcoran, M.E. (2003). The road to economic self-sufficiency: job quality and job transition patterns after welfare reform. *Journal of policy analysis and management*, 22(4): 615-639.

Leung, K., Ip, O. & Au, K. (2003). *Evaluation of the ending exclusion project for Social Welfare Department*. Hong Kong: Social Welfare Department.

- Marston, G. & McDonald, C. (2008). Feeling motivated yet? Long-term unemployed people's perspectives on the implementation of workfare in Australia. *Australian journal of social issues*, 43(2): 255-269.
- Martin, Larry G., and Mary Alfred, (2002), "What Employers and W-2 Job Experts Think about Retention and Advancement Barriers and Services", Final Report for the Wisconsin Department of Workforce Development, January 2002.
- Michalopoulos, C. (2005). *Does making work pay still pay? An update on the effects of four earnings supplement programs on employment, earnings, and income*. MDRC. August.
- Miller, C. et al. (2008). Implementation and second-year impacts for New Deal 25 Plus Customers in the UK employment retention and advancement (ERA) demonstration. *Department for Work and Pensions Research Report No. 520*.
- Ministerial Committee on Low Wage Workers. (2009). *Progress Report*. Retrieved November 10, 2010, from <http://www.moe.gov.sg/media/press/files/2009/06/progress-report-ministerial-committee-low-wage-workers.pdf>.
- Ministry of Community Development, Youth and Sports. (2010). *Help and assistance: Work support (WS) scheme*. Retrieved April 24, 2010, from http://www.mcys.gov.sg/web/serv_E_WS.html
- Ministry of Finance. (2007). *Budget speech 2007*. Retrieved July 24, 2010, from http://www.mof.gov.sg/budget_2007/index.html
- Ministry of Finance. (2010). *Budget speech 2010*. Retrieved May 25, 2010, from http://app2.mof.gov.sg/budget_2010/default.aspx
- Panel Study of Income Dynamics (n.d.). Data and documentation. Retrieved July 20, 2010, from <http://www.psidonline.isr.umich.edu/data/>.
- Peh, S. H. (Nov 14, 2006). Scandinavian-style welfare not for Singapore. *Straits Times*.
- Reenen, J.V. (2003). Active labour market policies and the British New Deal for the young unemployed in context. *National Bureau of Economic Research Working Paper 9576*. March.
- Roccio, J A. et al. (2008). Implementation and second-year impacts for lone parents in the UK employment retention and advancement (ERA) demonstration. *Department for Work and Pensions Research Report No. 489*.
- Tang, K.L. (2010). Welfare-to-work reform in Hong Kong: overview and prospects. In Midgley, J. & Tang, K.L. (eds.) *Social Policy and Poverty in East Asia: The Role of Social Security*. Routledge.
- Tang, K.L. & Cheung, J.C.K. (2007). Programme effectiveness in activating welfare recipients to work: the case of Hong Kong. *Social policy and administration*, 41(7): 747-67.
- Tang, K.L., Mok, B.H., Cheung, J.C.K., & Lou, V.W.Q. (2005). *An evaluation study of intensive employment assistance projects for CSSA and near-CSSA recipients (final report)*. Hong Kong: Department of Social Welfare.

Wisconsin Department of Children and Families. (2009). Wisconsin Works (W-2) Overview. Retrieved on July 27, 2010, from <http://dcf.wi.gov/w2/wisworks.htm>.

White, M. & Riley, R. (2002). Findings from the macro evaluation of the New Deal for Young People. *Department for Work and Pensions Research Report No. 168*.

Yap, M. T. (2010). Social assistance programmes in Singapore. In J. Midgley, & L. T. Kwong (Eds.), *Social policy and poverty in East Asia: The role of social security* (pp. 66-80). Oxford: Routledge.