Supporting Material is available for this work. For more information, follow the link from the Table of Contents to "Accessing Supporting Material".

Exploring Change in Organizational Rule Systems:
Learning Dynamics in Performance Measurement

Ignacio J. Martínez-Moyano [*]
im7797@albany.edu

Gerald Marschke
marschke@albany.edu

Rockefeller College of Public Affairs and Policy
University at Albany
Milne Hall 318, 135 Western Av. Albany, NY 12222, USA
Tel: (518) 442 3859 Fax: (518) 442 5298

Abstract

A model of how systems of rules in organizations are used and, over time, changed by learning processes and rule-following preferences of their actors is presented. The paper uses the case of the performance measurement system of the Job Training Partnership Act (JTPA) to articulate an endogenous theory to explore the impact that changes in performance measurement systems have in the way in which these systems evolve over time. In the model, the principal presents a system that the agent learns how to use (and possibly game) over time. The mutual learning (agent's learning about opportunities present in the system and the principal's learning about the problems generated by the agents' activities) creates pressures to change the system and modify the existing rules. Implications of the model results are presented.

Keywords: Performance Incentive, Performance Measurement, Gaming, Rules, Rule Dynamics, Learning, Organizational Learning, System Dynamics

[*] Corresponding author

> "All human social action is interaction—with others, ourselves, our natural and created physical world—within culturally defined contexts that largely determine, not only action, but its meaning."

Barbara **Frankel** in '*Metatheory in Social Science*' (1986, p. 360)

> "…objective understanding is an illusion created by 'inter-subjective agreement' on what is objectivity..."

Felipe **Fernandez-Armesto** in '*Truth: A History and a Guide for the Perplexed*' (1997)

> "validation is an inherently social process. It depends on the cultural context and background of the model builders and model users. It depends on whether one is an 'observer' (e.g., an academic researcher) or an 'operator', (e.g., a decision maker who must act without waiting for more data of further analysis)"

John D. **Sterman** in '*Appropriate Summary Statistics for Evaluating the Historical Fit of System Dynamics Models*" (1984, p. 51)

Table of Contents

List of Figures

1. Introduction to the Study of Rules

This work investigates mechanisms that influence rule evolution in organizations over time as a function of endogenously generated pressures. Human action is organized around rules and these rules fit together to create and maintain social systems.

Scholars in several fields have recognized that rules are the enablers of action in organizations and everyday life, that possibly there is no action that is not related to a rule, either written or unwritten. In addition, it has been recognized that rules and rule-based action are central features of all human society and human behavior. Furthermore, more than a century ago, scholars (in sociology) have been interested in the characterization of change in social structures and functions (Ward 1883; Small 1895; Ward 1895) that will provide a better framework to study the stability and transformation of social processes. Social norms or rules are a powerful form of control, fundamental to human behavior. However, despite the powerful role and possible learning, rule formation and its dynamics have been studied only rarely, less understood.

As portrayed in Figure 1, rules, through a process of implementation, influence action. Over time, and through a process of interpretation, action becomes history that conditions what is understood as effective and useful in organizations. Once a set of insights have been recorded as history, through a process of adaptation, lessons from history are incorporated to the rule repertoire of organizations closing the evolutionary cycle of rule formation and change. Organizational rule systems become then, at least potentially, organizational repositories of lessons from history: organizational memory in action.

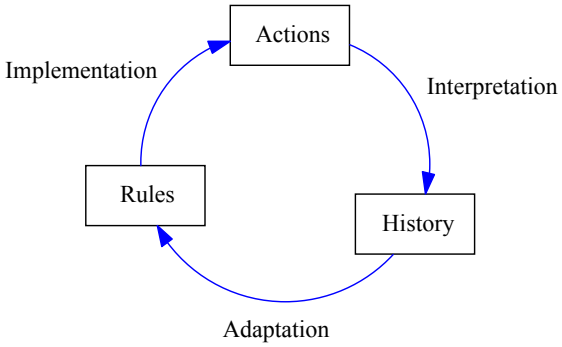


Figure 1.—Cycle of Rule Evolution (adapted from March, Schulz , and Zhou 2000)

The view of rules in organizations presented in Figure 1 makes important assumptions about the origin, nature, and process of creation and maintenance of systems of rules in organizations, especially about its origin based on history. This paper presents a view that explores this link and investigates what part of history plays a stronger role in transforming experience into rules.

2. Measuring Performance in Organizations

The task of measuring performance in organizations has been mainly addressed as a static endeavor (Courty and Marschke 2003a). The main assumption underlying this way of studying

measurement in organizations is that, once identified, an optimal scheme will endure over time. However, some scholars and practitioners have started paying attention to the dynamic nature of the establishment of performance measurement systems in organizations and their implications. Particularly with respect to improvement, goal attainment, operational implementation of changes, and stability in organizations. Consider, for example, what Courty and Marschke (2003a) have identified in their research of a federally funded training program in the US, or what Pauley and Ormerod (1998) explain in their investigation of performance measurement in the mining industry.

Pauley and Ormerod (1998) explain that "a key issue on the agenda of the mining panel was the corporate [e.g. the principal] interest in comparing actual performance between operations [of their mines]. Mine managers [e.g. the agents] wanted to judge [measure] performance in their own terms, specific to the conditions of their particular mining operations." Principal and agents in this case, through the mining panel formed, agreed that they needed to motivate higher performance by means of measurement mechanisms specially in elements associated with the capacity of the mines. Pauley and Ormerod (1998), in their conclusions, acknowledge as a key insight that "performance [measurement and] comparisons can cause conflict between central and divisional managers [principals and agents] in many organizations, both public and private." Furthermore, they identify a specific measure that allows the conflict level to be low between actors—capacity potential—and briefly discuss the implications for the future. They say:

> "…it is still too early to say whether the mine has changed its performance capability permanently or whether it (or the other sites we discussed) will settle into the new method of operation [performance measurement] only to find that it is inappropriate at some point in the future. The mine must seek to identify and implement continuous improvements and adjust its operations as conditions change." (Pauley and Ormerod 1998, p. 116)

Pauley and Ormerod (1998) present a very compelling case for the presence of 'dynamic behavior' in performance management systems that calls for adjustment over time as a necessary condition for stability in it's functionality. Interestingly enough, it seems to be that change provides stability in this context.

Normally, performance measurement in organizations is considered a very simple activity. Austin (1996), in his investigation of performance in organizations, states that "few management tools appear as simple as obviously useful as measurement." They describe it by saying that all one has to do is "establish numeric goals, take actions, and measure how the actions affect progress towards goals. Based on what the measures reveal, you adjust your actions. You continue in this way. Simple." (p. xv) Simply said, not simple at all in practice. Several problems arise when these processes are to be used in organizations. Establishing numerical goals is the firs step towards successful measurement, this activity is a tricky one because the definition of these goals will determine the definition of success and, further action towards improvement. Poorly defined goals can, and will, inhibit action and divert resources in the organization. Measuring results emerged from action is also key in this theory, however, in order to achieve this in an adequate way, identification of how the action is linked to the result is critical, thinking about how to establish the link is rarely reported. Lastly, for now, to be able to adjust the action based on the difference of the actual results and the goals implies a deep understanding of the 'true' drivers of the organizational results which, if truly known, would have been used in the first place to achieve the goals. In reality, this process is one of learning and discovery of the drivers of performance in organizations. Managers and directors create theories, some formally and others mentally, about what drives the true performance of their organizations and 'test'

them continuously by applying performance measures and management systems that adjust actions to the perceived results. Some of these hypothesis testing mechanisms are not systematic nor scientific causing unexpected consequences (in most cases, bad consequences) that trigger new changes based on new hypothesis that only, at best, maintain performance over time. A more rigorous testing mechanism could bring stability to improvement efforts and certainty in the sense of 'doing what is sensible' to maintain and increase performance over time. This way of thinking switches the traditional view of 'finding the best system to implement' and once you have it, protect it at all costs to 'maintain the focus on the evolutionary pattern that benefits you the most.' Ironically, both versions can be identified as quests for stability and control. The former focuses on static equilibrium design and the latter in a dynamic equilibrium design.

Some scholars have attempted to take into consideration complications inherent to the types of factors being measured in performance-related efforts, Stiefel, Rubenstein , and Schwartz (1999) put forward what they call 'risk adjusted output performance measures' that, according to their research, "allow the use of metrics closer to the goals of programs and can still afford reasonable accountability of managers by controlling for factors outside the organization's influence." (Stiefel, Rubenstein , and Schwartz 1999, p. 83)

Despite of the natural appeal towards performance measurement as means for improvement, very little insights have been generated with respect to the multiple dysfunctions that measuring performance in organizations can have. Literature in economics, contracts, and organizational behavior and theory explore some of the implications, however, a dynamic view of the endogenous nature of the changes in performance measures over time has not been articulated. This work attempts to generate a detailed articulation of an endogenous feedback theory of learning dynamics in performance measurement evolution pulling evidence from the literature and the case of a federally funded job-training program in the United States.

3. Case Study

This investigation of change and adaptation of systems of rules over time centers around the case of a Federally funded training-on-the-job initiative: the Job and Training Partnership Act of 1982. A description of the case is presented.

3.1. The Job Training Partnership Act—JTPA

The Job Training Partnership Act—JTPA—of 1982 created one of the largest federal employment and training programs in the country (Courty and Marschke 2002). The JTPA replaced its precursor Comprehensive Employment and Training Act—CETA—as the major job-training program for the poor. Its main innovation over the previous program in place—CETA—was the use of a performance-contingent incentive system (Cragg 1995). JTPA had a budget of nearly $4.0 billion dollars and served a constituency of almost one million people annually (for details of the numbers presented see Courty and Marschke 2002, p. 5 footnote 3).

JTPA programs spanned from 1982 to 1998 serving millions of economically disadvantaged individuals. Figure 2 presents data on counts of JTPA terminations from 1990 to 1997. In this graph one can appreciate the tendencies of the number of people served under the different titles of the act. This investigation centers on the behavior of terminations of individuals in Title II-A Adults. Other titles of the Act had different performance measurement systems with somewhat

different dynamics. In general, a characteristic cycle of change-reaction-assessment-new change can be observed.

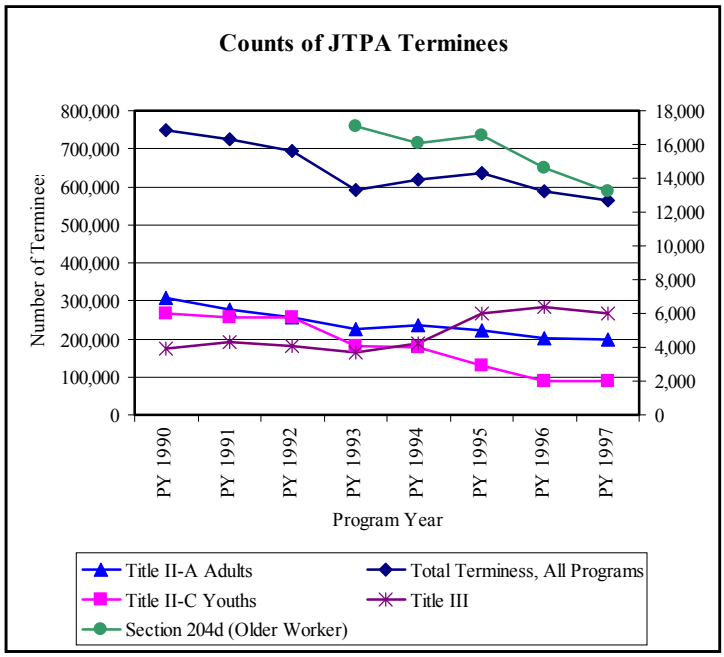


Figure 2.—JTPA Terminees

Changes to JTPA performance measurement system is used as case study in this investigation because of some important characteristics that the program had. JTPA was:

(1) Highly decentralized
(2) A program in which training agencies had significant decision discretion, and
(3) A program in which the Federal government used financially-backed performance incentive systems.

In addition, "between 1984 and 1999, JTPA was one of the largest federally funded job-training programmes for the economically disadvantaged. From 2000 through the present, the programme has continued, but was modified under the Workforce Investment Act of 1998" (Courty and Marschke 2003a, p. 278) providing a fertile field for studying change in performance measurement systems.

JTPA's mission has been defined as to "raise the earnings ability and lower the welfare dependency of the poor." (Courty and Marschke 2003a, p. 275) Alternatively, it has been identified as to increase the long-term human capital of program enrollees. JTPA's mission, in other words, was to help the economically disadvantaged do better in the long run via improving their skills and capacity to become employed and be self-sufficient, a very noble mission indeed[1].

The original titles of the JTPA legislation established four different programs. Title IIB authorized a summer youth program, Title III funded a program for dislocated workers, and Title

[1] Very difficult to measure its effectiveness though, but noble indeed.

IV governed various federally administered programs. Title IIA authorized the largest of these programs to serve economically disadvantaged youth and adults, accounting for the majority of JTPA client enrollments and training expenditures. In the early 1990s, Title IIA was split and a new Title IIC was created specifically for economically disadvantaged youth. Title IIA was re-authorized to serve adults only.

Congress intended JTPA activities to influence participants' human capital by helping participants become more capable and efficient workers by means of training. In order to provide incentives for the service providers to align their efforts with the intended results, performance measurement changed over time. Performance measurement in JTPA changed from cost-based performance measurement, to termination-based performance measurement, to follow-up-indicators-based performance measurement. Under any of these mechanisms, the service agency graduates enrollees as part of their normal process. An enrollee that has finalized the training, under ideal circumstances, would be graduated and reported to the state. However, because the graduation date is the date the training agency officially closes an enrollee's case and removes him from its rolls; and that date has a possible impact on the agency performance evaluation—especially under the termination-based mechanism—agencies tended to report graduation dates differently than the actual date in which the enrollees finished training.

3.2. Performance Measurement Systems in JTPA

Performance Measurement Systems in JTPA generated unintended consequences that influenced the performance of the providers of services and the way in which they provided the services to enrollees. Another important impact had to do with who was served because service providers had incentives to choose the type of clients[2] that maximized the likelihood of successful performance measurement depending on the performance measure in place at any point in time. It was recognized that "states believe performance measures are too high. Faced with this pressure to achieve difficult goals, there is an incentive to offer services only to those job seekers who can benefit the most and progress enough to meet performance standards." (Coalition_on_Human_Needs 2003)

According to Courty and Marschke (2002), in JTPA two different performance measurement systems were used over time[3]:

- Cost-Termination-Date-Based Performance Measurement
- Follow-up-Based Performance Measurement

In JTPA, the first two program years (1982 and 1983) represented a transition period from the system of rules that was active under CETA to the new system under JTPA. During this time, JTPA officials decided to investigate and gather data with respect to possible performance measures that could be used in the future, while doing this, the performance measurement continued to be carried out as it was under CETA. The continued use of the previous performance measurement system generated problems that were not evident to JTPA officials

[2] The service providers had the ultimate decision on whom to enroll in services. Allegedly, JTPA defined who was eligible, but did not impose any sanction if clients were denied services. This was modified later in time when JTPA rules changed making more difficult the selection bias by imposing 'minimum' percentages to be served of different types of clients as defines in the JTPA description of clients.

[3] For a list of the performance measures active during the program years of 1987 to 1989 and during the initial phase of the Work Investment Act of 2000.

immediately because of their focus on identifying new performance measures, not on identifying possible problems with current ones. This 'shift of focus' needed to acknowledge problems with the performance measurement system seem to be key to the dynamics identified.

For the purpose of investigating the dynamics of the performance measurement system in JTPA, one can identify three main rule regimes: (1) Post-CETA Rule Regime, (2) 90-Day Rule Regime, and (3) Follow-Up Rule Regime. Two important changes occurred In JTPA's performance measurement system that allowed for these three regimes. The changes can be identified to:

1- The introduction of the 90-Day Rule
2- The introduction of the Follow-up-based Measures

These two main changes in the configuration of the performance measurement system are the ones that will be modeled using system dynamics modeling to explore the elements that conditioned their initiation, development, and implementation.

The first significant rule change[4] under JTPA was the introduction of 'the 90-day' rule. After CETA was replaced by JTPA in 1982, performance measurement of the providers of services was conducted in the exact same way as during CETA, however, both public officers and managers of service agencies were trying to identify new performance measures that 'made more sense' than the previous ones. As in any other change effort, during the transition time, a more relaxed mechanism of measurement was allowed for the agents. Public officials, recognizing that the providers of services had more information about the status of the trainees and had a better understanding with respect to their needs gave them certain latitude to determine the exact termination date of the trainee. Normally, the termination date would be the date that the trainee finished his training, independently of their employment status. Under the post-CETA rule regime, managers of agencies in charge of training economically disadvantaged individuals would decide when to terminate the individual after he was done with his training depending on his employment status. This is a very simple rational choice for the managers, they were measured mainly by the metric *'employment rate at termination',* or the percent of individuals that after training had a job with respect to the population of terminated individuals (with jobs or without jobs). Under this regime, the higher the employment rate at termination (ERT), the better. Managers were given latitude on when to terminate enrollees and the definition of success was tied to ERT, it was a straightforward choice. Managers would terminate individuals only if they had jobs, not before. This created incredible high ERTs (100% if managers held on inventory enrollees that would not have jobs) that were not necessarily related, only, to the quality of the training received. Actually, in a way, the post-CETA rule regime recognized the status of the economy in which the training agency was operating. With enough time, all of the terminees would have jobs (eventually) and would have been taken out of the training agency rolls; it was just a matter of time. Perseverance and patience was key in order to achieve great performance measures under this rule regime.

ERT was a key measure during the post-CETA rule regime, not the only one though. Other measures included salary and costs. In terms of the dynamics generated, ERT was key because of

[4] In performance measurement system related rules.

the implications for managerial action related to termination procedures and their impact. Actually, the way in which managers handled the termination process made a lot of sense within that performance measurement regime. While providers of training services were dealing with the day-to-day difficulties and decisions related to certification, enrollment, training, and termination, public officials in charge of improving performance measures for JTPA were busy gathering information for the determination of the new measures. All the information gathering and analysis was aligned with this purpose: define new measures. It was not until other agencies and sources of information, like the General Accounting Office, started reporting anomalies in the way training agencies managers were handling the termination process that the Department of Labor (US_Department_of_Labor 1993) recognized that some problems had arisen that needed their attention and solution. The department of labor recognized that it had been determined "by monitors and auditors" that managers in training centers had allowed "some participants continued to be carried in an 'active' or 'inactive' status for two or three years after last contact" (US_Department_of_Labor 1993, p. 4) with their training programs. Their solution was the implementation of the 90-day rule.

With the inclusion of the 90-day rule, the DOL required training agencies to measure performance using the ERT measure but also required agencies to terminate individuals that had not received any services after a maximum of 90 days. This change, de facto changed the performance measure used by DOL. The main performance measure was still ERT but, because the method of computing it changed in a way that significantly changed the outcome measured, one can think of it as a completely different regime: The 90-day-rule regime.

The 90-day rule was designed to provide agencies "some latitude in securing jobs for their customers" (US_Department_of_Labor 1997, p. 1) [and improving their performance measure], and at the same time avoiding the possibility of hiding bad performance by not terminating unsuccessful candidates. After the implementation of the 90-day rule, many states voiced their concerns about the impact of the rule in the performance measure of their agencies asking for its removal. The States identified, quite adequately, that introducing the new rule regime had consequences. The main consequence was the providers of services now had to report a more 'realistic' outcome and therefore, their performance measurement suffered. It is interesting to realize that the introduction of a new rule regime, the 90-day-rule regime in this case, generated complaints both in the side of the providers and in the side of the agencies (states) that were supposed to be monitoring their activities. In a principal-agent framework, the providers were the agents and the states were the principals. It seems that introducing a 'better' and 'tighter' rule for the agents would be welcomed by the principals because it would create a higher value added all together for the principal, and the customer. However, in a nested hierarchical mechanism, like the one in which JTPA existed, principals at some level are agents at another one. States became the agents of the DOL and therefore needed, and wanted, to comply with certain performance measure too. Having to implement and use a better system of rules created a paradoxical problem for the states because after the implementation they were not in a position to 'help' their agents do better and, consequently, improve their own (the states') performance. In this case, and in any other in which multiple nested principal-agent mechanisms are in place, contrary to common wisdom, principals would have as much an incentive to have a 'not so perfect' system of rules as the agents would. Why? Because the principals need to have latitude to be able to play around with the system if needed and aligned with their own objectives and performance measurement

requirements for their principals (at one higher level). The principal becomes the agent and is subject to, and has incentives to, game the system as much as its agent does. Is this in a way like a fox in the hen house?

Additionally, "in addition to employment [ERT] and hourly wage measures, in the early years JTPA training centers faced a cost-based measure that judged the program's managers by how much they spent to produce an employment at termination" (Courty and Marschke Forthcoming 2004b). The cost measures were identified as critical in the first part of the life on JTPA because, it was believed that "cost-based measures judged JTPA's managers by how much they spent to produce a job placement" (Courty and Marschke 2003a, p. 278) and was a great efficiency measure. After JTPA officials monitored the activities for some time, gathered information about the cost measures, and analyzed it adequately, they realized that they were measuring the same things as with the ERT measure.

Augmented knowledge about the cost measures in the performance measurement system in JTPA led in 1992, "8 years after the cost measures were first introduced", JTPA officials to phase out "these measures because 'research and experience have shown that the use of cost standards in the awarding of incentives has had the unintended effect of constraining the provision of longer-term training programs'" (Courty and Marschke 2003a, p. 279). The cost measures implementation and removal, as the implementation of the 90-day rule, suggest that, in the construction of performance measures, understanding the dynamics of it is important.

After the introduction of the 90-Day-Rule Regime, "another important change in the measurement system was to move to 'follow-up' measures." (Courty and Marschke 2003a, p. 279). The introduction of the follow-up-rule regime became, in JTPA, the way to solve different problems identified over time in JTPA-related provision of training services.

The follow-up-rule regime was the response of JTPA officials to the problems identified after the introduction of the 90-day rule regime. The most important problem identified was the possibility of influencing the performance measure by means that were not aligned with the 'true spirit' of the program[5]. JTPA officials identified, at least, three mechanisms[6] that providers of services decided to use after the implementation of the 90-day rule regime to influence their performance outcomes. These mechanisms, to the principal, were gamming mechanisms. The mechanisms that providers decided to use to improve their performance, to the providers, were strategic mechanisms. In this case, and in almost all of the principal-agent relationships, the definition of an action as 'gamming' or 'strategic' lays on the eye of the beholder. What the principal evaluates as gamming (because of the perception of it not being aligned with the true goal) is a rational and strategic behavior for the agent (because he is trying to improve the way in which he is evaluated). As long as the agent is not 'breaking' the rule system presented to him, any action that he chooses to engage in to improve his performance will be rational and strategic. The responsibility of designing and presenting adequate performance systems is of the principal. If the system presented to the agent promotes the 'wrong' kinds of behavior, it is the principal the one to blame. Understanding the possible consequences, in terms of agents' behavior, of the

[5] One has to remember that the mission of JTPA was defined as increasing the long-term human capital of participants to help them do better in long-term employment.
[6] Described latter in the document.

systems of rules presented to agents should be a key task of the principal. Principals should be designers of adequate systems (Senge 1990) for their agents. Greatly designed systems improves significantly the likelihood of average agents excelling at their assigned tasks (Sterman 1994). Excellence in design of systems of rules should be a priority for principals, not a mere random occurrence.

The follow-up rule regime recognized problems that government investigators (GAO 1996, 1998, 2002), academic researchers (Courty and Marschke 1997, 2003a, Forthcoming 2004a, b), and JTPA officials (US_Department_of_Labor 1990, 1993, 1997), over time, identified. What they identified were several types of questionable behavior in the part of the providers of training services. Three types of activities that were identified as possible gamming mechanisms are:

1—Cream-Skimming Activities
Providers of services selected 'better' candidates that would improve their chances of having successful performance outcomes.
2—Strategic Termination Activities
Providers of services timed the termination of their idle inventory close to the end of fiscal year when it created the opportunity of improved performance.
3—Post-Training-Engagement Activities
Providers engaged in short-term solutions to increase likelihood of employment for their terminees.

The follow-up rule regime was, in essence, the same as the previous one in place but measured the outcomes three months after the termination date instead of at termination date. This change was the solution presented by JTPA officials to the problems identified with the use of the 90-day rule regime. This solution tried to address the short-term focus of the activities of the service providers. Once again, the change in the performance measurement system was the mechanism that officials in JTPA used to 'solve' problems identified and analyzed. The fact that they chose this particular solution is interesting and irrelevant at the same time. It is interesting because it is the result of very hard work of many people involved in the administration of this program. It is the result of hard work and deep thought process that tries to take into consideration all ramifications of the problems observed and come up wit a true solution that will solve every malfunction identified to date. However, it is irrelevant in a way because it fails the address possible problems for the future and possible reactions of agents to these changes. Conceivably, changing the system will solve problems, but it will bring new ones that will create pressures to change the system again in the future (see Figure 3). It seems to be that the only possible path towards a 'stable' system after a change is introduced is through the implementation of a 'perfect' solution that takes care of the recognized problems without creating new ones (Figure 3's 'B1' loop) and through the interaction with either unperceiving or unresponsive agents that will not react to the changes introduced. Either cycle, in reality, seems impossible to achieve.

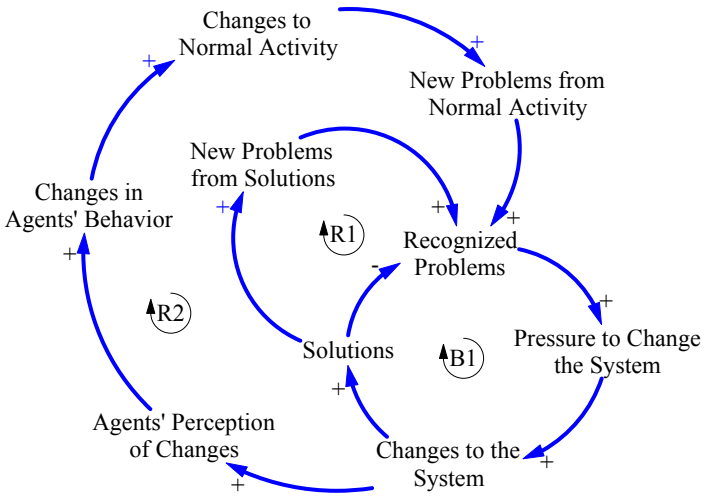


Figure 3.—The Consequences of Change

The two main rule regimes can be identified as the key changes from a dynamic perspective, however, these were not the only changes present in this time period. Two changes were introduced in 1994 (SPR 1999, p. I-7). First, individuals that did not receive services beyond objective assessment were excluded from the computation of performance standards. Second, employment outcome measures (adult follow-up employment rate, welfare follow-up employment rate, and youth entered employment rate) counted only for employment of at least 20 hrs per week[7].

Analyzing changes to the JTPA performance measurement system allows one to identify a tendency in it to continue changing over time, however, it is also interesting to note that "between 1992 and 2000, the year the Workforce Investment Act (WIA) supplanted JTPA, performance measures remained largely unchanged" (Courty and Marschke 2003a, p. 279) and then, "it further evolved under the WIA programme" (Courty and Marschke 2003a, p. 279). This particular behavior is interesting to identify why. Changes to the performance measurement system stopped for a while and then they started again.

In summary, three distinct stages can be identified in the profile of the performance measurement system in JTPA. These stages are:

1—Post-CETA Rule Regime
2—90-Day-Rule Regime
3—Follow-Up-Rule Regime

During the Post-CETA rule regime, no changes to the performance measurement rules occurred until the implementation of the 90-Day-Rule Regime. The 90-Day-Rule regime generated reactions from the agents that created pressures to change the system again and the

[7] This to conform to the requirement in Section 106(k) that states "…'employment' means employment for 20 or more hours per week." This change is an indication of a lower level of tolerance to differences with respect to how the rules are set up.

Follow-up rule regime was designed and implemented to solve the problems generated by the 90-day rule regime and after its introduction. After the implementation of the follow-up rule regime, unintended consequences were identified and, the processes of monitoring the activity was heightened, however, no changes were finalized nor implemented.

4. Modeling the Endogenous Evolution of Performance Measurement Systems

I am modeling a set of changes in the way performance measurement systems changed in JTPA over a 20-year period. Because "an important limitation of the multi-tasking literature is that it provides only a static view of the process of designing performance-measurement systems," (Courty and Marschke 2003a, p. 269) I adopt a dynamic view of the phenomena. I am using a dynamic modeling technique that has a clear purpose to link microlevel evidence with macrolevel behavior. Essentially, the model that I am set out to build represents a detailed micro-level causal theory of a macro-level observed behavior of the change of JTPA performance measurement systems.

In this modeling work, I am trying to capture "the mutual dependence of individual choices and actions on the one hand, and macrovariables or parameters of choice on the other" (Hernes 1976, p. 517-518) in a comprehensive causal theory.

The modeling effort includes the modeling of the two main changes identified in the life of the performance measurement system in JTPA: the 90-Day-Rule Regime and the Follow-Up-Rule Regime. The simulated period is 14 years in month intervals accounting for 156 months of the life of JTPA. The initial conditions of the model are set out to replicate the Post-CETA Rule regime. Then the model simulates in an endogenous way the emergence of the 90-day rule regime and subsequently captures the strategic (gamming) reaction of the agents in the system. Three reactions are modeled, the creaming gamming path, the strategic termination-gamming path, and the post-termination activities gamming path. The model also generates endogenously the emergence of the follow-up rule as a response to the dysfunctions observed in the systems and, finally reaches a stable condition of no change. This last phase represents the period of time in which the JTPA officials did not change the system after the introduction of the follow-up rule until the end of the life of JTPA.

4.1. A Simple Model of the Evolution of Performance Measurement

Building on Courty and Marschke's (1997, see figure 1 in page 385) theory of change in performance measurement systems, I propose that JTPA dynamics can be understood using an endogenous view based on the structure of the performance measurement system and the responses of the individuals subject to it. This theory says that the observed behavior influences the performance measurement system that, in turn, influences the responses of the individuals subject to that performance measurement system. The responses of the individuals create changes in the observed behavior and foster further changes in the performance measurement system.

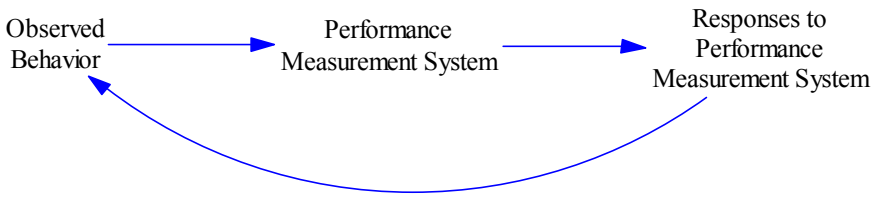


Figure 4.—Initial Proposition

4.1.1. True Objective of the System

The true objective of JTPA was defined in the text of the Act (1982) and can be defined as increasing the long-term employability of individuals via the development of useful skills through training.

A concept of *human capital* was identified in the case data and developed in the model variables. In the JTPA model, *human capital per trainee average* represents the true objective of the system, and changes to that variable are linked to value added and gamming mechanisms in the model.

4.1.2. Performance Measures

The JTPA model simplifies the performance evaluation function used by DOL in order to be able to create a simpler representation. This simplification is also consistent with evidence from the case in the sense that the measure chosen for the modeling effort, *employment rate at termination*, is recognized as the single most important performance measure active during the life of the JTPA. "In the first decade of JTPA, the employment rate at termination was the most important measure in determining a training center's award. A training agency's employment rate at termination for the fiscal year was computed as the fraction of enrollees who terminated during that fiscal year who were employed on the date of their termination. At the beginning of the next fiscal year, the slate was wiped clean, and performance measurement began anew. Training centers that exceeded their assigned standards for the year received higher awards. Thus, if in an arbitrary fiscal year a training center's standard for the employment rate at termination was 64 percent and its overall employment rate for that fiscal year was 65 percent, the training center would receive a budgetary award to begin the following fiscal year." (Courty and Marschke Forthcoming 2004b)

4.1.3. Gaming Mechanisms

From the three gamming mechanisms explored, one[8] is presented in this paper. These mechanisms are not the only gamming mechanisms identified to date in the literature but can be identified as representative of the effects that any gamming mechanism can have on the design and change of performance measurement systems. Each gamming path poses a causal theory of how gamming was developed in JTPA activities. The three mechanisms are:

[8] The cream-skimming mechanism

1- Cream-Skimming Activities
Providers of services selected 'better' candidates that would improve their chances of having successful performance outcomes.
2- Strategic-Termination Activities
Providers of services timed the termination of their idle inventory close to the end of fiscal year when it created the opportunity of improved performance.
3- Post-Training-Engagement Activities
Providers engaged in short-term solutions to increase likelihood of employment for their terminees.

4.1.4. Rule Changes to be Modeled

Out of the two rule changes modeled in this investigation, one is presented in this paper, the introduction of the 90-day rule. The full process of rule change can be described starting with an existing rule regime, then identifying problems related to is use, then knowing what to do about those problems, culminating in a process of implementation of a new rule regime that can take care of those problems. After the implementation of the changes, a new full cycle begins when new problems, that arises as a product of rule creation (endogenously) and as a product of other variables in the system (exogenously to the process of rule creation), are discovered and acted upon. The two rule changes modeled are:

1- The introduction of the 90-Day Rule
2- The introduction of the Follow-up-based Measures

4.2. Modeling the First Rule Change (The 90-Day-Rule Regime)

In order to explore the model developed in the previous section, a detailed model of the changes of JTPA performance measurement system is developed. The model will be presented in the next order. First, the first rule change is explored using JTPA specific modeling variables and some results are presented, then insights about change in performance measurement systems are extracted from the modeling process and presented. Then, the second rule change is modeled and new insights are derived that lead to the possibility of refining the theory developed and try to synthesize it in a generic model of learning in rule evolution and change.

In the case of JTPA, the true objective of the system (or the principals' goal) was to increase the human capital of the enrollees via training. In the ideal case, the providers of the training services would dedicate their effort just to provide training to the enrollees in order to maximize this goal. However, the performance measurement system (the post-CETA rule regime) used employment rate at termination as a proxy for human capital and created the possibility of distortion.

4.2.1. Basic Material Flow Structure of the 90-Day-Rule Regime

The JTPA modeling effort started by identifying the basic structure that conditioned the material flow of trainees throughout the system. The first conceptualization included the flow from individuals in need to JTPA enrollees, and then to JTPA trained individuals, and finally

JTPA terminees. The only flow out from *JTPA Trained Individuals* to *JTPA Terminees (employed)* is through *employment rate*. This formulation captures the reality of the JTPA systm operating under the Post-CETA rule regime in which the providers of training services, having the latitude to terminate individuals when it was more convenient for them, only terminated individuals when they reached employment status.

4.2.2. Origins of the 90-Day-Rule Regime

The emergence of the 90-day-rule regime was modeled identifying the specific variables that the DOL used to become aware of the problem as described in the training and employment information notice 2-92 of the Department of Labor (1993). The department of labor stresses the fact that " some participants continued to be carried in an 'active' or 'inactive' status for two or three years after the last contact with these programs." (US_Department_of_Labor 1993) This is captured in the model (see Figure 5) through the *average time waiting*[9] variable. The *average time waiting* is computed by dividing the number of people in the stock *JTPA Trained Individuals* by the sum of its outflows: *termination rate* and *employment rate*. This *average time waiting* is compared with a maximum allowed (with a value of 1 month), the *max time waiting* to identify the ratio of the measurement to the maximum as a measure of deviation from the standard.

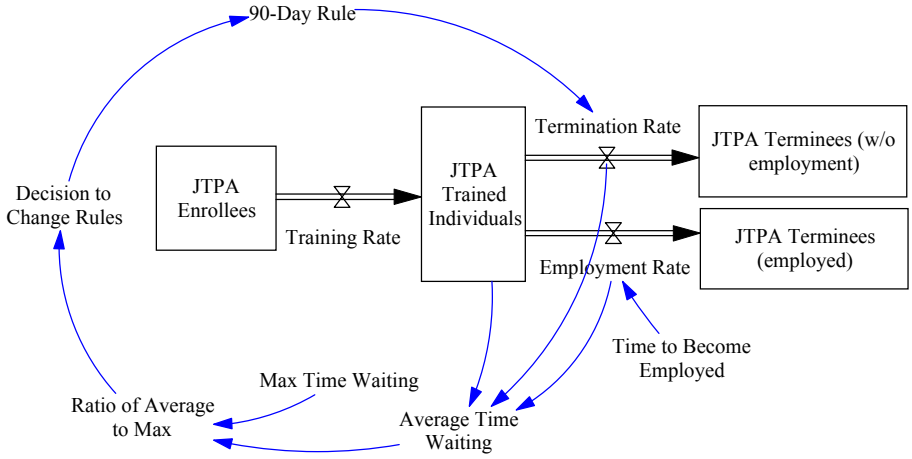


Figure 5.—Origin of the 90-Day Rule

The *ratio of average to max* is then used as a triggering variable for the decision made to change the performance measurement system to solve the 'waiting time' problem identified by the auditors and monitors (US_Department_of_Labor 1993). *Decision to change the rules* determines the creation of the *90-Day Rule* that in turn modifies the *termination rate* flow. What is actually happening is that, when the *90-Day Rule* is activated, the structure of the model is changing by 'creating' a new flow that did not exist before the emergence of the 90-Day rule. This 'creation' is possible by enabling the flow (*termination rate*) that was not active before the *90-Day Rule* changes from 0 to 1. In a way, the creation of the 90-Day Rule changes the system and modifies the structure of the model. It is a 'pre-existing' structure in the model, just waiting

[9] As a way to increase the clarity of the description of the model, I will use italics to denote variables of the system dynamics model developed.

to be enabled, but the way in which is turned on is through the endogenous creation of the 90-day rule.

This view presented in Figure 5 changes now by 'adding' the new structure created by the new rule in place. This new structure allows for the creation of an alternative way out of the stock *JTPA Trained Individuals* which now has two ways to terminate individuals from agency rolls: through the *employment rate* or through the *termination rate.*

Modeling the change in rules in performance measurement systems in JTPA has allowed me to identify the process of endogenous structural modification in system dynamics models. It is important to recognize that this endogenous modification is a special kind in which a dormant (and preexisting structure) is awakened or activated in an endogenous way. System dynamics models are described as structurally fixed, this means that their structure does not (and cannot) change during the course of the simulation, this model shows how this notion can be changed and expanded to allow the endogenous modification of (previously defined) structure of the model. The case can be made that in the real world that is the case all the time. It can be said that 'new' structures are nothing but 'pre-existing' and not-yet-identified structures that are discovered and awakened after interventions are designed and implemented.

Once the details of the model started to come together, the boundary of the system was redefined to capture the relevant dynamics to the change in performance measurement system in JTPA. The processes associated with meeting the eligibility criteria for JTPA and filling out applications, etc lay outside of the purpose of the model, therefore the basic stock and flow structure was modified to include the graduation rate (set at a constant rate of 300 people/month) as the originator of the flow.

The emergence of the 90-day rule was modeled by recognizing seven distinct processes that the DOL had to go through when thinking about the changes needed to solve the 'waiting time' problem. The DOL (the principal in this case) had to:

1—Gather information about potential problems.
The principal ha to have information.
2—Assess the value of the information gathered and identify problematic behaviors in it.
The principal has to know what the information means.

The principal has to identify the specific variable that created the problematic behavior (in this case the *average time waiting*) out form several variables that were available to them at that point in time. Actually, the fact that this particular variable was identified and that the specific 90-Day rule was implemented is more a function of the available information and not the urgency or importance of the problem. It is conceivable to say that there could have been several other variables 'misbehaving' in the system, maybe in more important ways, which were overlooked or not even identified at the time that the specific one addressed was identified. The first main problem that the any principal faces when identifying problematic behaviors to be solved is that of extracting meaningful signals out of the myriad of signals available to the him. It is similar to finding a needle in a haystack. The problem is that in this conceptual haystack, there are several needles and the principal has a very limited capacity to find them and evaluate the urgency to address each of them. Actually, the principal cannot know what he does not know

until he finds out about it. This means that principals are severely restricted to the information that they 'see' and this information conditions the type of information that they look for in the future causing a path dependence possibility in the way they asses and address problematic behavior in their systems.

3—Gather information about possible solutions.

The principal has to know what to do about the problematic behavior found (find solutions).

4—Create the conditions for change based on the solutions found.

The principal has to be able to create the necessary conditions to implement the solutions found. Including convincing the actors involved in the process (different constituencies affected by the rule change and the implications of the solution).

5—Change the systems.

The principal has to be able to change the system of rules as needed.

6—Monitor and evaluate the implementation and effects of the change.

The principal has to be able identify the variables that will tell him if the chosen solution is working as expected.

7—Identify problems associated with the solutions implemented (new rules).

The principal has to be able to monitor the systems that he intervenes in and identify changes in behavior (after the new rules are in place) that could be consequences of the solutions implemented.

In the model, the DOL (the principal) engages in three different types of effort in order to change the system. The three types are: monitoring, analysis, and formalization of changes. In order for a rule to be changed, the DOL has to exert effort in monitoring, analysis, and formalization of changes. The monitoring effort allows DOL to 'know' about the problem, exerting effort in 'analysis' allows DOL to know what to do about it, and effort exerted in formalizing changes is the one that allows DOL to be able to change the structure of the rule system.

Effort on analysis allows for the accumulation of knowledge about the problem creating a signal for the need of a new rule to 'solve' this anomaly in the system and it then when effort in formalizations of changes start to build up to create the capacity needed for change in the system of rules. As the formalization effort declines, continued effort on monitoring and analysis increase slightly to account for the post-implementation processes needed to insure that the rule has been implemented and that it is addressing the problem. After some time, DOL is done with the change process lowering its effort across all related dimensions.

The emergence of the 90-day-rule regime is captured in the model in three different phases. The fist one, captured by the *indicated need for the 90-day rule* variable, captures the moment in time in which DOL acknowledges that a new rule is needed to solve the problem identified. In this case, the excessive waiting time of trained individuals in the idle inventory. In the 'base' run of the model, this happens at time 10 (months) after having received the report from GAO on month 6.

The second phase of the emergence of the rule is captured by the *change in practice of the 90-day rule* variable. This variables picks up the moment in which the rule is ready to be

implemented in the population of providers of services, in the 'base' run this happens at time 16 (months) approximately. The difference in time between when the need is recognized and when the rule is ready to be implemented is endogenously generated as a function of the effort exerted by the principal in formalizing the new rule. This time is the 'cycle time' of administrative rule making (6 months in this case).

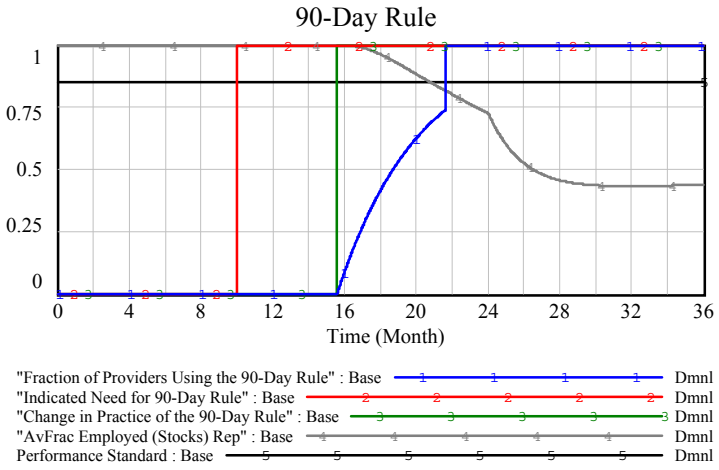


Figure 6.—90-day Rule Dynamics

The third phase of the emergence of the rule is captured by the *fraction of providers using the 90-day rule* variable that accounts for the implementation process of the rule in the population of rule subjects. Implementing the new rule takes time and can become an important source of the dynamics depending on the adoption mechanisms in the population of rule subjects[10]. The quicker the adoption mechanism, the faster that the modification of structure will take place and the modification of behavior will become evident. Changes in model structure may, or may not, change model behavior. Only if the modification created generates a leverage point, the behavior will change in a noticeable way. This creates an interesting hypothesis with which to explain the apparent incapacity of regulation to modify behavior.

According to the system dynamics tradition, the behavior of the system is a function of its structure; therefore, if you change the structure of a system, you are motivating a change in behavior. However, also according to system dynamics logic and accumulated knowledge, the bigger the structure, the less sensitive to changes it becomes leaving large systems and models virtually insensible to the vast majority of the changes that you could think of. In large-complex systems, the task of finding leverage points becomes a titanic endeavor.

H2—Only new rules that can create changes in structure in 'leverage' points of the system will generate noticeable changes in behavior. This means that the vast majority of the rules and regulations that will be created will not produce a significant change in the behavior that they try

[10] See point 6.2.3 for an explanation of the different mechanisms present in this model and their effects on model behavior. The model assumes 100% compliance with the rule once adopted by the population of rule subjects. As each provider is adopting the new rule, they comply with it and change their behavior accordingly. A very interesting analysis would be to explore the dynamics of different levels of compliance to rule changes.

to regulate. This lack of effectively would be a product of the density of the rule population under study. It seems pessimistic for regulation advocates.

H2—1—The higher the density of a rule population in a given problem domain, the lower the impact on behavior of any new rule or regulation process.

H2—2—The lower the density of a rule population in a given problem domain, the greater the impact on behavior of any new rule or regulation process.

The introduction of the 90-day rule generates consequences in the behavior of the providers of services in the system. The first, and most important, consequence of the introduction of the 90-day rule is the decline in *employment rate at termination*. ERT is modified because the implementation of the 90-day rule changes the structure by creating a new flow out of the *trained people (TP)* stock. When this new flow is created, the ERT suffers immediately and declines from its original 1 (or 100%) in which every individual terminated was employed. This is to say that now the providers need to actually increase their relative effort in order to generate successful outcomes. The second effect of the change introduced is that the *average total time waiting* drops dramatically and reaches a new equilibrium when 100% of the population of providers adopts the 90-day-rule regime.

4.2.3. Consequences of the implementation of the 90-Day-Rule Regime

The implementation of the 90-day rule in JTPA had two immediate effects: lowered the total time waiting of the individuals in the idle inventory, and lowered the effective employment rate at termination to the providers of services. The agents (providers), as a result of these changes and modified environment, changed their behavior in order to try to influence the performance measurement.

Additionally, "after DOL instituted the 90-day rule, employment rates likely better reflected training center's success at producing employment." (Courty and Marschke Forthcoming 2004b) The 90-day rule limited the latitude of the providers with respect to when to report the termination of an enrollee, however, "on the other hand, the 90-day window afforded training centers enough latitude to boost employment rates significantly beyond what they would have been had employment been measured at training end." (Courty and Marschke Forthcoming 2004b) Why did the final 90-day rule have latitude built-in? It could have been a 'successful' influence from the part of the providers during the commenting and defining parts of the administrative rule making or an honest effort in the part of the principal to 'solve' the waiting problem somehow. However, both agents and principals, in the JTPA case and in many others, have incentives to be bounded by a set of rules that provide enough latitude to be able to achieve their mutual goals. Extremely rigid systems of rules can (and most probably will) become a problem for both the principal and the agent.

4.2.3.1.Inaction Path: No Reactive Agents

The first scenario explored in the JTPA model was the behavior that was generated when the agents did not respond to the changes in performance measurement. This means that there is no change in effort allocation after the introduction of the new performance measurement rule.

4.2.3.1.1. The Behavior of the JTPA Model with no Reactive Agents

Figure 7 show the behavior of the JTPA model after the introduction of the 90-Day rule when the agents in the system do not respond to that change. This unresponsiveness of agents could be related to three things (in a theoretical perspective[11]).

Agents can become unresponsive or insensitive to changes when:

1. Agents are not capable of identifying that their performance measurement has changed (Awareness-related situation). This problem can be the result of:
    a. Not being able to see the changes (Information-acquisition problem).
2. Agents are capable of identifying the changes. Agents have the information but they are not able to make sense out of it (Information-interpretation problem). This problem could have two causes:
    a. Agents do not have enough physical capacity to analyze the information that they have (Information-overload problem)
    b. Agents do not have the adequate skills or expertise to identify 'nuggets of knowledge' out of the myriad of accessible data (Information-processing problem)
3. Agents are capable of identifying the changes. Agents have the information, they know what it means (they processed it perfectly), but they are not willing to do anything about it (Information-activation problem). The activation problem could be related to:
    a. The interpretation of the information does not represent reason enough for action (Schauer 1991b, a; Goldman 2002) (Tolerance-related situation).
    b. The interpretation of the information is reason enough for action but the agents are not sure what is the action that should be taken (Determination-related situation).
    c. The interpretation of the information is reason enough for action, the agents know exactly what could (or should) be done, but they are not willing to do it due to other considerations or pressures in the system (Quinn and Rohrbaugh 1983) (Competing-values situation).

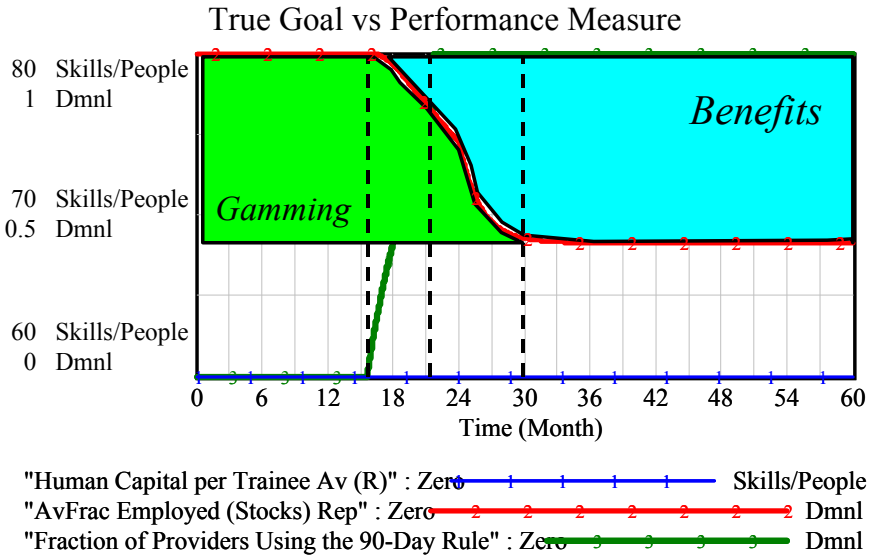
As shown in Figure 7, at time 15 the 90-day rule starts to become implemented and by time 21 the adoption of the rule is complete, at that time fraction of providers using the 90-day rule is 1 meaning that 100% of the providers are using that specific rule. At time 15, when the first providers start to adopt the rule, the *average fraction employed* or ERT, starts to decline as a result of the activation of the *termination rate* causing the performance measure used in JTPA to drop. The ERT goes from 1 (or 100% of employment rate at termination) to 0.41 or 41% of

[11] This is not what happened in the JTPA case, I am exploring this option first to see what the model predicts with respect to the overall behavior of the model.

employment rate at termination. In this case, the introduction of the rule creates a modified environment for the providers that change, in a very dramatic way, how their performance is measured. The change in the performance indicator when no reaction is allowed in the part of the providers can be interpreted as the identification of the amount of gamming that was happening before the implementation of the rule.

The introduction of the 90-Day rule reveals that, when steady state is reached, the 'true' employment rate at termination achieved by the providers is only 41% and not 100%. The difference, 59%, can be attributed to gamming. What gamming activity was that? The activity of strategically keeping the trainees in the rolls until they became employed. However, was this 'gamming' or 'strategic thinking' aligned with their goals and objectives? Was this a practice that was premeditated to cause harm to the JTPA program or was it a practice that emerged as the 'clearly logical thing to do' under that set of rules. What is the difference between 'gamming actions' and 'strategic actions'? Is this just a matter of perspective? If the principal is the one charged with the task of defining a set of rules for the agent to follow, and the principal is supposed to know what he is doing, then the agent is not gamming, he is strategically behaving. If the principal and the agent, jointly, develop the set of rules to be followed, and the agent purposefully misinforms the principal about characteristics of the technology necessary for the contract to work, and uses the set of rules to his advantage, then the agent is gamming.

The introduction of the 90-day rule uncovered the size effect of a specific type of behavior that the DOL characterized as not adequate (gamming). In the JTPA model the size of the effect can be identified and characterized. Figure 7 shows the amount of gamming before the introduction of the 90-day rule and the benefits obtained by the introduction. The green shaded area shows the amount of gamming and the blue shaded areas the potential benefits related to the introduction of the 90-day rule[12].



[12] These are 'potential' because they a dependent upon the reactions of the agents (different gamming) and the unintended consequences of the new rule itself (autonomous inefficiencies not captured in this model).

Figure 7.—Gamming-Benefits Identification

4.2.3.2.Value-Added Path: Effort on Training

Agents responded to the changes experienced after the introduction of the 90-day rule by modifying the type of activities they conducted on a day-to-day basis.

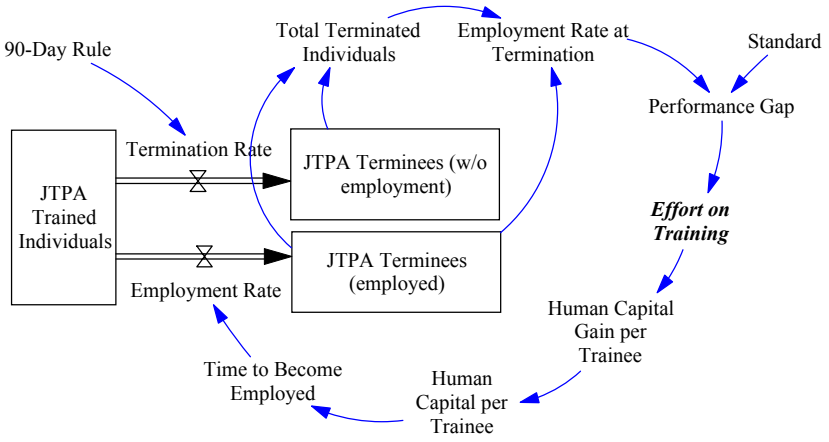


Figure 8.—Value-added Effort Mechanism

The first response from agents was to work harder to try to influence the performance outcome (see Figure 8 for a feedback description), this is to exert effort in what the principal defines as value-added activities. Additionally, agents engaged in several strategic responses related to JTPA activities. Some of the strategic responses of the providers of services in JTPA have been argued in the literature (see Courty and Marschke 1997; Prendergast 1999; Courty and Marschke 2003a, b). Out of the documented strategic responses, three were incorporated in the model: cream skimming, strategic termination of individuals, and engaging in post-training activities.

Figure 8 shows a causal model representing the mechanisms that influence changes in the amount of effort that agents exert in value-added activities. When the *90-day rule* is implemented, the *termination rate* flow is created changing the *total terminated individuals* stock. This change pressures the *employment rate at termination* down creating a *performance gap* to become evident to the agents. Once the *performance gap* is perceived, the agents increase the amount of effort exerted in value added activities (*effort on training*) influencing *human capital* and eventually the *employment rate* that will bring *the employment rate at termination* back up. Once the *employment rate at termination* increases, the *performance gap* closes and the incentives for added *effort on training* disappear. A higher *employment rate* is translated into higher performance outcomes that will close the *performance gap* that triggered the increase of effort in the first place. This balancing mechanism (for an explanation of the nature and characteristics of balancing feedback loops see Sterman 2000) acts as a control loop that conditions the amount of effort that providers exert on training actions, or value added activities.

The larger the performance gap perceived, the greater the added effort the agents will exert. After some time, added effort will influence outcomes and will close the gap. This loop is active only if three conditions are met.

(1) The providers are able to perceive the performance gap (realize that there is a gap to be closed)
(2) The actions that the providers decide to take (value added activities) have an impact on the measured outcomes.
(3) The DOL (principal) has the technology and resources to identify changes in performance measures and recognizes them.
    a. The DOL could have the technology but not the resources to use it and therefore the performance outcomes would not be recognized.
    b. The DOL could have the technology and the resources but the timing of their use could be misaligned with the changes of the performance outcomes creating delays and biases.
    c. The DOL could have the technology and the resources and adequate timing but not be willing to release information about the changes in performance creating delays and biases (basically misinforming agents).

4.2.3.2.1. The Behavior of the JTPA Model with Value-Added Activities

The scenario in which the providers of services respond to the introduction of the 90-day rule with increases value added effort is considered to be the *Base* run of the JTPA model because it captures the reactions of the agents in just one dimension: training activities. More dimensions are explored later in this document. The additional effort exerted by the providers influences the gain in *human capital per trainee Av,* or the true objective of the system (see Figure 9), above is original level captured by the *Human capital per Trainee Initial Profile*. The human capital moves from 60 skills/people to more than 70 skills per people at time 40 months. Once the human capital reaches this new level, it reaches steady-state equilibrium.

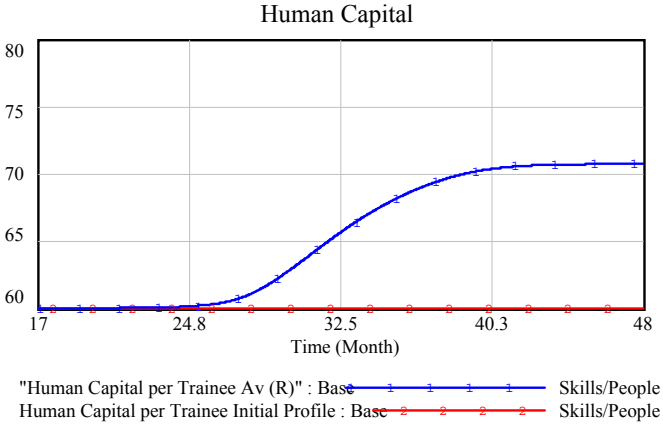


Figure 9.—Human Capital Formation

Figure 10 shows how changes in effort in value added influence, after a lag time, the behavior of human capital of the trainees. The time lag is a function of the structure that captures the relationship between these variables.

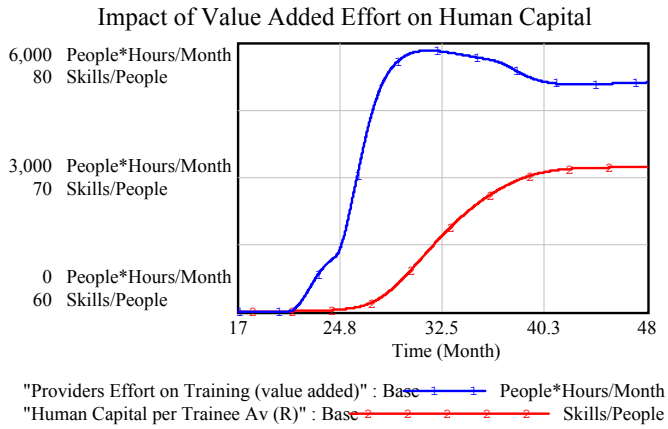


Figure 10.—Impact of Value Added Effort on Human Capital Formation

Figure 10 shows how the *providers effort on training (value added)* grows starting at time 21 influencing the growth of *human capital per trainee Av* that starts at time 24. From time 24 to 40 approximately, both variables grow (at different rates) and then a new stable equilibrium is achieved. It is interesting to denote that the allocation of effort of the providers is an endogenous process and that the value added effort is growing at a very high rate to then decline and find the stable equilibrium point. This behavior is known as 'overshoot and collapse', in this case created by the providers' search of maximization of human capital to influence ERT (the performance measure).

In the economic literature on performance-incentive design it is defined as a mechanism to find adequate performance measures those that capture the variability of the true objective of the system (Prendergast 1999), this is based on the assumption that they know what the relationship is before implementing the measures. In this literature, a measure of correlation is often used to capture this relationship. If the correlation of the proposed measure with the true objective of the system is high, then that measure is considered a good candidate to be used as performance indicator. Researchers in economics have argued that "simply selecting performance measures based on their correlation with the organization's true objective cannot suffice when measures are game-able." (Courty and Marschke 2003b, p. 27) They propose "an alternative model where the principal does not know how aligned or misaligned a performance measure is until it is used." (Courty and Marschke 2003a, p. 275)

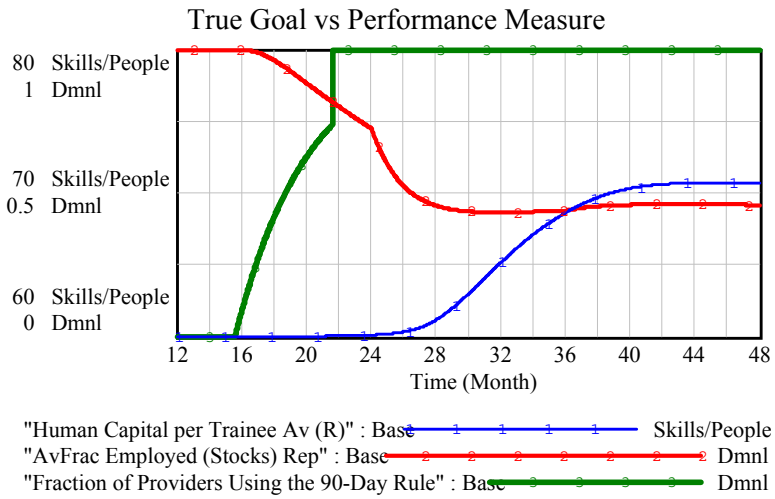


Figure 11.—True Goal vs. Performance Measure

Figure 11 shows how the relationship between the true objective of the system, *Human Capital per Trainee Av (R),* and the newly changed performance measure ERT, *AvFrac Employed (Stocks) Rep,* changes as a result of the introduction of the 90-day-rule regime captured by the *fraction of providers using the 90-day rule* variable. Before the introduction of the new rule (before time 15 approximately), both the true objective and the performance measure are constant in the model (ERT=1 and Human Capital=60 skills/people). After the introduction of the new rule, both the true objective of the system and the performance measure change because of endogenously generated pressures. The true objective of the system, *human capital per trainee av*, experiences a growth that can be characterized (due to the specifics of the model) as s-shaped growth from time 24 to 44. At the same time, the performance measure, *avfrac employed,* decreases; at first, very steeply from points in time 17 to 24 and after that in an exponential decay fashion until time 30 when it reaches steady-state equilibrium. The change in behavior of the performance measure in related to the fact that growth starts on the side of the true objective of the system that acts as a 'control mechanism' for the fall of the performance indicator until the two of them reach an equilibrium state.

The important aspect of the growth of the true objective of the system is not the specific shape that it has but the fact that the true objective of the system is growing as a result of the rule change. In other words, the rule is working creating a better world (at least for the principal, and allegedly for the customers of services), as expected. The introduction of this rule should be considered a success. Furthermore, the introduction of the rule actually 'solves' the waiting problem influencing the change in *average total time waiting* from 7.3 months to 3 months (or 90 days as requested by the rule imposed by the DOL US_Department_of_Labor 1993). However, during these successful times, the performance measure drops and continues plummeting depicting a story of failure to the agents (providers of training services) and for the principal 'evaluating' the agents' actions.

Courty and Marschke (2003a; 2003b) in their investigations, capture the clear idea that in the presence of gamming, the correlation between the true objective of the system and the

performance measure is going to change and that principals' will only be able to identify this after the measure is implemented. The use of correlation indexes as the only cue to identify good performance indicators then is doubted. Identifying the relationship between the true objective of the system and the performance indicator in this JTPA model, in the absence of gamming, allows one to see that, what Courty and Marschke (2003a; 2003b) argue that happens in the presence of gamming, might hold in the absence of it too. This will happen in the case that at least one of the conditions for change is present[13]. Even in the absence of gamming, and using a truly adequate performance measure, a change in correlation can be expected if there are delays in the system that prevent the changes to have immediate impact on agents behavior and, more importantly, on the true objective of the system.

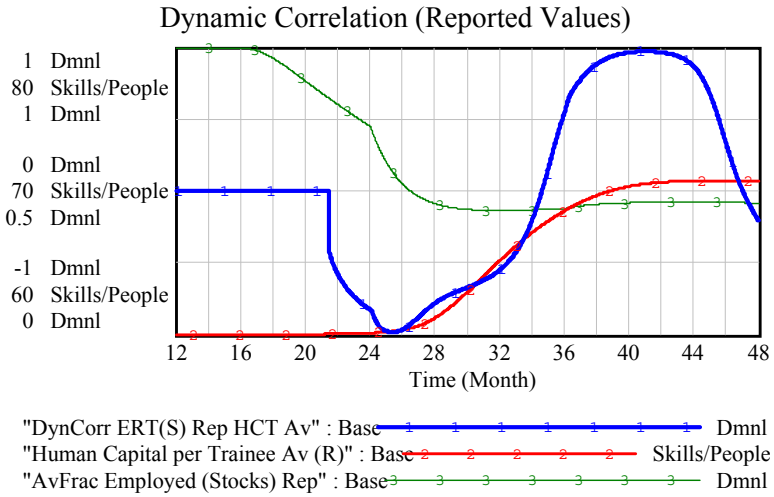


Figure 12.—Dynamic Correlation

In the case of the JTPA model (as shown in Figure 12) from point in time 21.5 to 48, the relationship (correlation) between the true objective of the system and the performance measure changes (after the introduction of the new rule and the reaction in *human capital per trainee*). Furthermore, from time 21.5 to 25, the relationship, as captured by a correlation index, becomes negative and reaches -1. This means that the better the true objective of the system becomes, the worse the performance evaluation of the agents is. From time 25 to 34.25 remains negative but on its way to become positive. After that, from time 34.25 to 41 grows in the positive side to become +1 (a perfect direct relationship). Later, from time 41 to 47, the correlation declines in the positive until it become cero (representing no relationship at all). Finally, in this simulated run, from time 47 to 48 it declines becoming negative again.

Figure 12 shows the behavior over time of the correlation between the two variables. These is a 'dynamic correlation' identification in which one can see the way in which the relationship changes as new information is added to the correlation computation over time.

Traditionally, the correlation index is thought of as a 'static' number measured at a given point in time that represents the relationship between the two variables under study. At that point

[13] See the conditions for change described above.

in time, the relationship is evaluated en determined to be as desired or not [correlation is a measure of the strength of association (usually meant to be linear) between two variables. Correlation coefficients can vary from –1 to +1 and usually are identified with the letter r. "A positive correlation coefficient indicates that as one variable increases the other increases, while a negative correlation coefficient indicates that as one variable increases the other one decreases." (Byrkit 1987, p. 761)]. In this case, we see that it changes over time as the variables change as a result of the introduction of the new performance measure. This means that, if the principal measures the relationship between the variables (true objective of the system and performance measure) at different points in time, the evaluation of the relationship will be completely different. Furthermore, if he is using this measure (the correlation) as a proxy for fit, the principal could take inadequate decisions with respect to the performance measure and with respect to the value of the behavior of the agents. In addition, once again, all of this is in the absence of gamming. Actually, in this case, the agents are doing value added activities as a result of the modification of the performance measure, and even then, the relationship between the variables change dramatically in certain periods of time.

Summarizing, Courty and Marschke (2003a) argue that "selecting performance measures on the basis of their correlation with the organization's true objective may not always be a valid approach" (p. 270) due to the emergence of gamming. According to Courty and Marschke (2003a), if gamming appears as a result of the activation of a performance measure, the correlation between the performance measure and the true objective o the system will change. Based on learning from the simulation process, it seems to be the case that, even when gamming is not present, Courty and Marschke's expected change in correlation occurs. The change in correlation is the product of the responses of the agents, not gamming responses, to influence the performance indicator. In this case, the responses of the agents are value-added oriented and beneficial to the principal. However, even then, the correlation pattern changes and can create misleading indicators of performance improvement.

In the model, the way in which human capital is accumulated represents a learning process. In this theory of how individuals learn, there is a *saturation effect* active and a *difficulty to learn effect* active. This theory implies that, at the beginning, it is difficult to learn because of the lack of knowledge, and once having a 'critical mass' of knowledge it becomes easier to learn or to accumulate human capital. Additionally, it is defined a maximum human capital that can be achieved (in this case across the entire population of trainees) that when the accumulation of knowledge approaches that point it becomes more difficult to learn anything new and eventually shots down the learning process.

If the maximum human capital is achieved, there is nothing else to be accumulated. In addition, when individuals leave the *trainees in JTPA* stock, they carry with them a proportional part of the human capital accumulated. There is no distinction made about the quality of the individual leaving training, the assumption is that human capital is evenly distributed across the population. In order to relax this assumption and to investigate the effects of that change, one would disaggregate the trainees in JTPA stock into different-by-quality-of-trainee stocks (high achievers, average, low achievers, etc). Furthermore, a *human capital loss due to forgetting* flow captures the degradation of human capital as a function of time alone as an exponential decay process governed by a constant *average time for trainees to forget*.

4.2.3.3.Gamming Path: Cream-Skimming Activities

The decision of what type of strategic response to have to changes in the environment is key to agents' success. If agents only opted for additional value-added effort, the modification to the system of rules or to the performance measurement system would be a complete success. However, this is not the case all the time, changes to the performance measurement system promotes changes that were not intended originally by the principal. JTPA providers of services facing declining performance measures (especially *employment rate at termination*) had other responses besides exerting additional value-added effort—doing what is expected of them.

In JTPA programs and services, the type of individuals that were eligible for services was determined by the act it self. Under the JTPA Title II-A (1982), providers of training services should enroll economically disadvantaged individuals that could benefit from training services. The definition itself is not 100% precise due to the nature of the individuals and to the fact that it is a 'class' of individuals being defined and not a 'single' individual. The definition of eligibility of services in JTPA is a textbook example of the nature of rules: probabilistic generalizations of behavior. Providers in JTPA had a 'clear' range when it came to eligibility criteria in which, the caseworkers had the final judgment call of who met the criteria and who did not. This, and the fact that in JTPA being eligible did not insure you services, created an interesting scenario for strategic behavior in the part of the providers of services. According to JTPA (1982) standards, those eligible could have services but did not mean that they should have them. Being eligible gave you the possibility of being served but did not include any enforcement mechanism that would make the providers serve you particularly. Furthermore, when faced with the prospect of decreasing performance outcomes after the introduction of the 90-day rule, caseworkers and managers of the training centers had incentives to select 'better' individuals that could generate better performance outcomes for them. This activity has been identified as cream skimming (for a detailed exploration of the empirical research on cream skimming see Kim 2004).

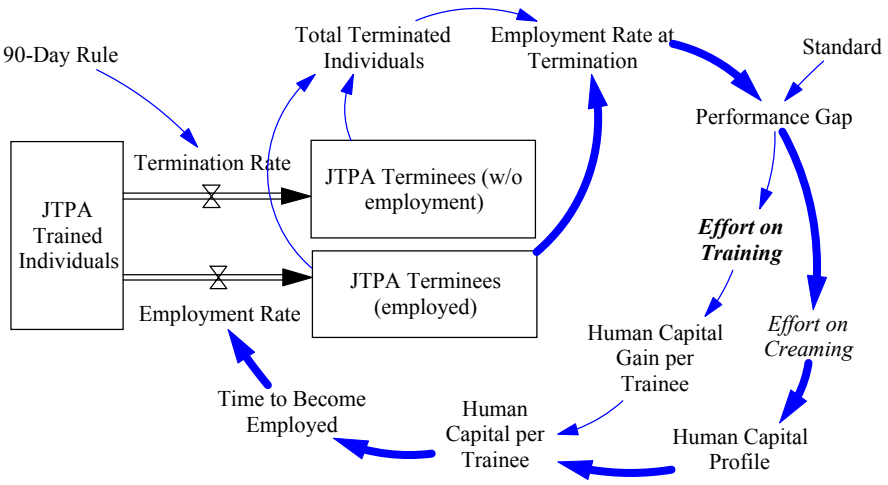


Figure 13.—Cream-Skimming Mechanism

Figure 13 show a feedback perspective on the emergence of cream skimming activities in JTPA. When the providers of services perceive a performance gap, another possible course of action besides exerting value-added effort is to engage in cream-skimming activities. In the model, this is captured by the *effort on creaming* variable that influences the *human capital profile* of the individuals being served. Human capital profile, in turn, will have an influence on the total human capital per trainee obtained that will influence the employment rate changing the performance outcome. Again, like in the case of value-added effort, once the performance outcome is affected and the performance gap closed the incentives for exerting effort on creaming activities disappear (this becomes a negative feedback mechanism that acts as a control tier with respect to the way in which providers allocate their effort). By including the creaming-activities loop in the model (see Figure 13), now we have two balancing loops acting together to control de performance gap that provides incentives for effort allocation to the providers.

Possible Scenarios		Clients	
		High-Need	Low-Need
Services	High-Need	Intended	Waste
	Low-Need	Waste	Waste

Figure 14.—Possible Scenarios in Cream Skimming

If the providers of services in JTPA chose 'better' clients by cream-skimming the pool of eligible participants, the type of services they provided might have to change too in order to adapt the services to the needs of those selected. This created four possible scenarios shown in Figure 14. The types of clients, and services provided to them, are conceptualized as high-need and low-need. A brief explanation of the scenarios follows:

Scenario 1—High-need clients and high-need services

This scenario is the one indented for JTPA service providers. This combination of services and clients is the one that would bring the adequate services to those that needed them the most. If the selection processes of JTPA service providers would include only these types of clients and these types of services, the overall value-added to the community would tend to be maximized. However, even in this 'ideal' scenario some injustices could be made to individuals depending upon the definition of the threshold that separates the high-need with the low-need clients. Under JTPA standards, all eligible are 'officially' high-need and, under that assumption, societal benefits would be maximized even when 'low-need' clients are served as long as they fulfill the eligibility requirements.

Scenario 2—High-need clients and low-need services

This scenario captures the combination of high-need clients, those who would benefit the most from the services, with what has been characterized as low-need services. This, like in the distinction of clients, is not an absolute characterization of the value of the services provided but a relative assessment of the services provided with respect to others that could have been provided instead. Low-need services are, therefore, essentially needed; but not as much as others

are. This is especially true in the case of high-need clients. This specific combination creates waste in the provision of services.

Scenario 3—Low-need clients and high-need services

This scenario captures another wasteful combination in the provision of services. The services provided are aligned with what the clients need the most, however, the clients are not those the need the most help.

Scenario 4—Low-need clients and low-need services

This last scenario can be considered 'the worst' combination possible because it depicts the situation in which the clients served are not the ones that need it the most and, additionally, these clients are not receiving the types of services that they need the most. This is a situation that would create the largest amount of effort of the four explored in this simple characterization of the relationship between the types of clients and the types of services provided at any given point in time.

JTPA eligibility criteria prevented the JTPA providers from selecting 'extremely' good individuals[14], but did not prevent providers of services from selecting 'the best' of the eligible population. In the conceptualization presented (see figure 14), low-need clients are still considered 'high-need' when compared to those that do not meet the eligibility criteria in JTPA. This means that the level of need expressed in the distinction presented in figure 14 is relative, not absolute. Both types of clients, high-need and low-need, are the same in the sense that they meet the eligibility criteria of JTPA, but they are different in the sense that they are at the two ends of the distribution of the eligible population.

4.2.3.3.1. The Behavior of the JTPA Model with Cream-Skimming Activities

In the JTPA model the *switch to activate gamming activities* variable controls for the providers' possibility to engage in gamming activities over the simulated time. Figure 15 shows the effect of allowing for gamming in the JTPA model previously shown only with value-added mechanisms activated.

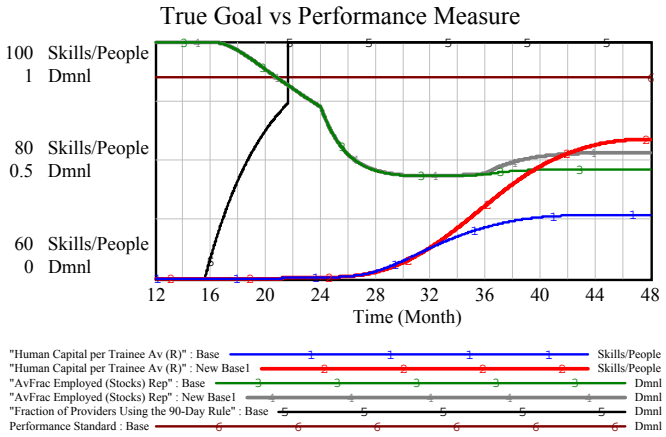


Figure 15.—Behavior with Cream-Skimming Activities

[14] Because these would fall outside the eligibility criteria. They would not need services a t all.

If gamming takes place, the human capital per trainee changes from stabilizing at time 48 at 70 skills/people to stabilizing at 82 skills/people (see lines 1 and 2 of Figure 15). The difference in human capital achieved influences the new employment rate at termination changing from 45% to 53% at time 48 (lines 3 and 4). In this case, the change in the performance measure, *employment rate at termination,* (17%) is proportional to the change in the 'true objective of the system', *human capital per trainee* (17%). The two sources of change in *human capital per trainee* are the value-added effort exerted by the agent in *training effort* and the gamming effort exerted that, in this case, is recognized in the change in *human capital per trainee initial profile*.

The initial profile is defined to be at 60 skills/people and remains constant until time 28 when the *effort in gamming* is high enough to influence a change in the profile. The profile of the trainees rises gradually reaching a maximum level at time 41 (approximately) and then declines again. The human capital per trainee changes from 60 skills/person to 82 skills/person in part because of the training effort (value-added effort), and in part because of the effort on creaming activities (gamming effort). The behavior of *human capital per trainee* is shown on Figure 16.

Figure 16-A shows that when no reaction from the agent exists, neither the *human capital per trainee* or the *human capital per trainee initial profile* change over time. If the agent responds to changes in the rule system with value-added activities only, the profile will remain constant and the human capital will increase due to the added effort on training activities.

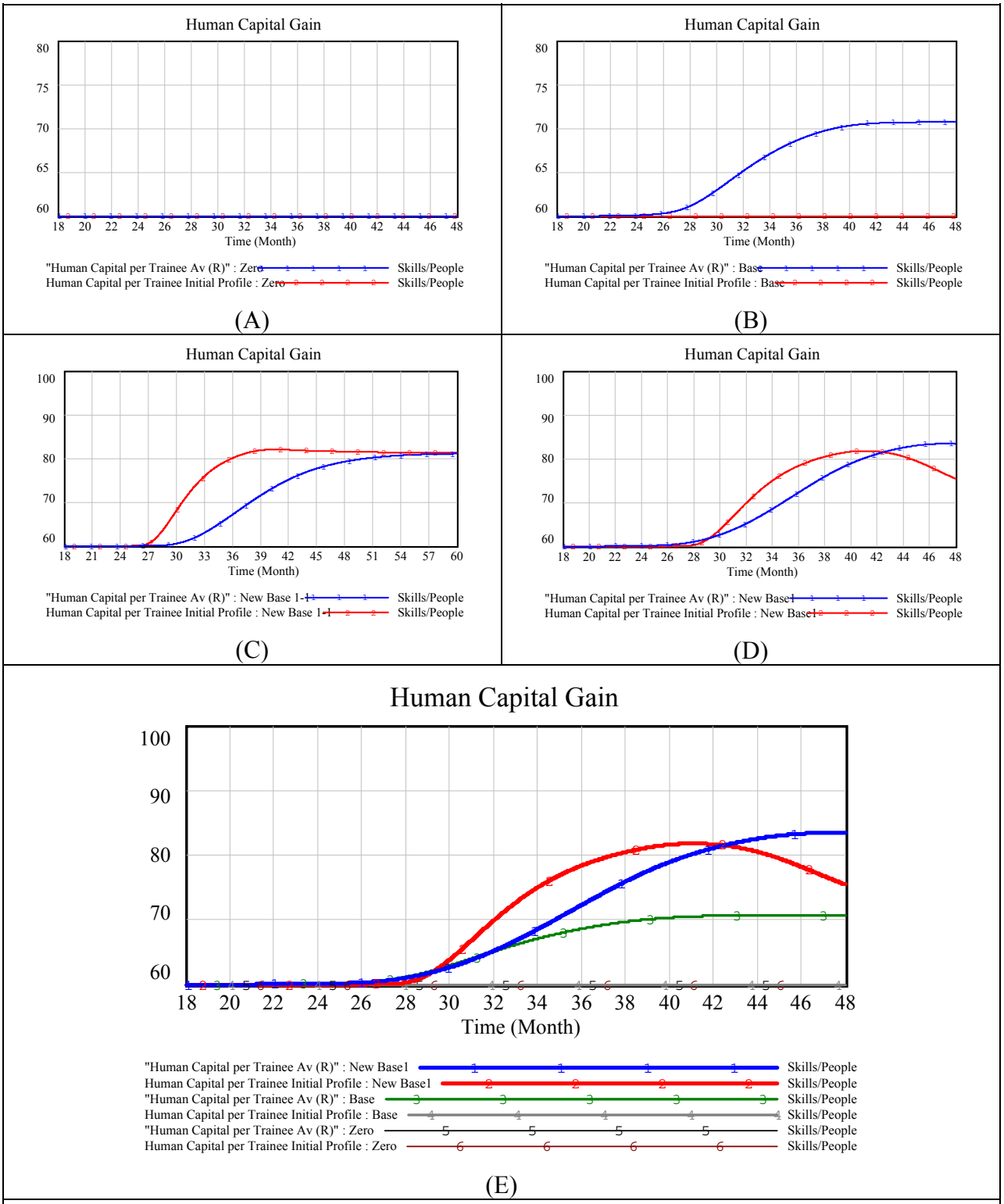


Figure 16.— Changes in Human Capital Profile and Average

In Figure 16-B a change in human capital per trainee from 60 to 71 skills/person is shown. In this case, the gain in human capital is due to value-added action only, there is no gamming activated in this run. Run *new base 1-1*, shown in Figure 16-C, show the changes in behavior due to gamming activities only. *Human capital per trainee initial profile* grows, starting at time 29, and continues to grow until time 44 when stabilizes at 81 skills/people. *Human capital per trainee* follows the growth of the profile and, after a lag period, eventually catches up at the same level at time 60. In the case of pure value-added activities (Figure 16-B), and pure one-gamming-stream activities, in this case, cream-skimming activities (Figure 16-C), is very easy to distinguish the origin and amount of change in human capital due to that specific type of effort. In reality, multi-stream effort exertion is always the case making is slightly more difficult to recognize the individual effects on the output variable of interest (or on the performance measure). Figure 16-D shows the changes on the variables when the providers of services develop both value-added activities and cream-skimming activities. In this case, the human capital per trainee grows and stabilizes at 82 skills/people while the *human capital per trainee initial profile* follows an increasing-decreasing trajectory with a maximum achieved at time 41 at the 81-skills/person mark. In order to understand what is the contribution from gamming in this case one can add to the graphical representation the base case contribution from value-added activity and subtract that from the overall gain in human capital per trainee. This process assumes that (1) information about the base case, pure value –added action exerted following a regulation implementation, is available to the agent, and (2), there is no significant interaction between the two types of activity (value-added and gamming) that will change the nature of the specific effort-type gains (gamming gain, and value-added gain). In the simulated case of JTPA, the second condition does not hold. Figure 16-E shows the 'complete' picture that allows us to identify value-added gain and separate it from gaming-related gain, however, when compared with Figure 16-C, the gamming-related behavior changes as interactions unfold over time. Actually, the gains from a pure one-gamming-stream activity are decreased in the presence of value-added activity. In order to compute an approximation of the gamming contribution to the change in human capital per trainee, the difference of the integration of the human capital (using gamming and not) can be computed.

In order to distinguish the effects of value-added activity and one-gamming-stream activity, in the specific case of human capital gain, analysis of the sources of change can be used. Identifying the flows that contribute to the change *in human capital per trainee,* in this case, two flows, the *human capital gained* flow and the *human capital built* flow, can be a clearer mechanism.

The *human capital gained* flow represents the gain in human capital obtained by changes in the initial profile of the trainees (cream-skimming activities). The *human capital built* flow represent the human capital built with training activities.

As shown in Figure 17-D and E, in the presence of value-added activities and gamming activities, the two flows change over time. Human capital difference change rate represents the amount of human capital change achieved as the difference between the value-added effort and the gamming-effort so, if the *difference change rate* is positive it represents a period of time in which the gain in human capital is larger due to value-added activities than it is because of gamming activities. If the *difference change rate* is negative, it means that the gain due to gamming activities is larger than that due to value-added activities. If the difference is cero, it

means that the gain in human capital is due to both types of activities in equal amounts or inexistent (Figure 17-A shows the latter case). In the case of pure value-added effort exerted, shown in Figure 17-B, only the *human-capital-built* flow changes, and all of the *human capital difference* is due to value-added action. The area under line 3 in Figure 17-C represent the total amount gained with the training activities. In the case of pure one-gamming-stream activity (shown in Figure 17-C), the only flow changing is human capital gained net generating a negative human capital difference from time 27 to 48. In these two cases, the pure value-added case, and the pure one-gamming-stream case, once again, it very simple to distinguish the sources of the change in human capital. However, in the case shown in Figure 17-D, in which from time 22 to 32 *human capital difference change rate* is positive, meaning that, in that time period, the change in human capital per trainee is predominately due to value-added activities. And from time 32 to 48, the *human capital difference change rate* becomes negative showing that the main driver of the rise in human capital per trainee is the gamming that the providers are engaging in. The analysis shown is able to identify the 'most influential effect' on human capital per trainee; however, it is not to be interpreted as the 'only' effect active in that period of time. As shown if Figure 17-D, from time 29 to 48, both effects are present and contribute to the rise of human capital per trainee. In order to clarify the difference, as before, one can use the base reference to distinguish the individual effects, as shown in Figure 17-E[15].

However, the most influential effect changes from value-added to gamming at time 32. Another interesting part of the analysis is to identify the trend that the human *capital difference change rate* has. This trend analysis informs about the likelihood of one effect becoming dominant to the other (or others) over time. However, clear tendencies could be reverted by the interaction of the variables in the model.

Figure 18 shows the behavior of the *provider's effort on creaming* variable and its effects on *time to become employed*[16]. Line 3 on Figure 18 shown the *zero* run in which no reaction from the agents exists; in this case, *time to become employed* remains constant at the level of 7.5 months. As the *providers' effort on creaming* goes up (see line 1 in Figure 18), the *time to become employed* changes and drops, capturing the effect that 'better' trainees have on the job market after graduating from instruction. The effect of gamming on time to become employed is captured by the difference between the *zero* run and the *new base 1-1* run.

[15] Once again, like in the previous case, interactions between the value-added action and gamming-related action will change the specific gain of gamming and value-added activities.

[16] The effort on creaming influences the time to become employed in an indirect way through the human capital per trainee. This is not a direct, or immediate, influence. The influence presented in this figure corresponds to the creaming activities only. This run, the New Base 1-1 run, controls for the value-added effort (it is cero) and other types of gamming activities (all cero).

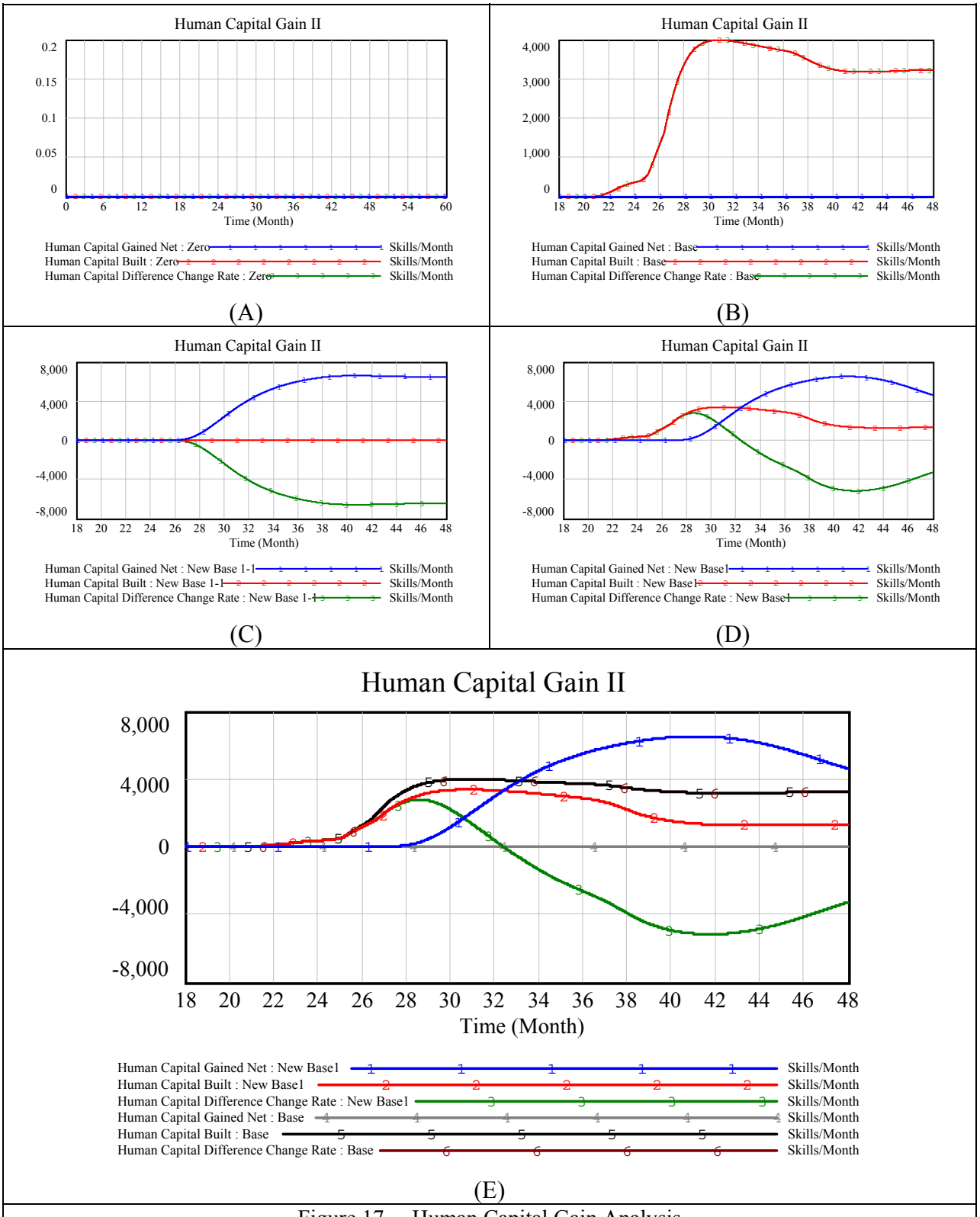


Figure 17.—Human Capital Gain Analysis

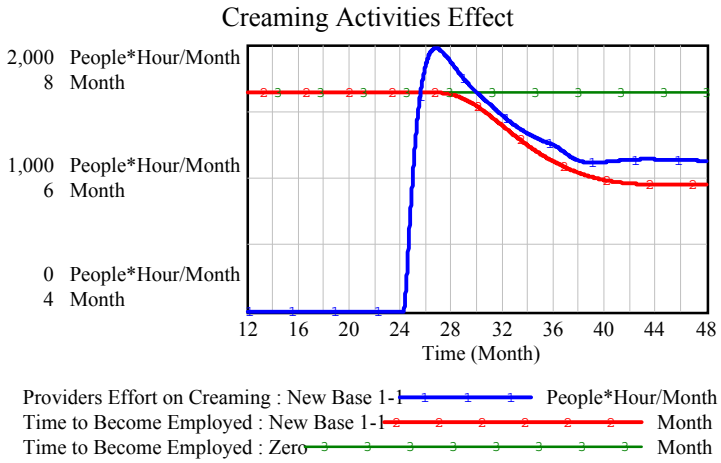


Figure 18.—Effect of Creaming on Time to Become Employed

The *providers' effort on creaming* sharply increases from time 25 to 27 when it reaches a maximum and then, from time 27 to 38, it declines in a linear fashion reaching a new minimum and then increases again. Time to become employed starts its decrease at time 28 and continues declining until time 43 (approximately), this decline is in part the reaction to the increased gamming activities that the providers exerted at times 25 to 27. From time 30 to 36, both the *time to become employed* and *providers' effort on creaming* decrease in the same fashion (see Figure 19). This is a good example of the difference between correlation and causation. If someone analyzed the relationship between *providers' effort on creaming* and *time to become employed* from time 30 to 36[17], that person would find a positive correlation (probably close to '+1') between the two. This finding could inform the researcher that, contrary to what would be expected[18], decreasing effort on creaming reduces the time to become employed, this would mean that the providers that do not engage in this gamming behavior would be better off. However, this 'temporal coincidence' is a product of the causal structure and the delays associated with it. The 'truth' is that the providers that do engage in creaming activities would be better off, *later,* in the process. Nevertheless, this insight might never be available to researchers having limited information on the behavior of both variables. These researchers might end up having 'great' statistical results, with 'very high' r^2 but without any real sense of what is driving the behavior of the variables under study (see Figure 19).

[17] Under the assumption that it is a time period that allows the researcher to gather 'enough' data to conduct statistical analysis.

[18] A negative correlation (a '–1' would be the ideal). The negative correlation is expected between a gamming effort and the time to become employed because the aim of the gamming mechanisms is to increase the performance measure that, in this case, is defined as the *employment rate at termination* that is naturally linked to the *employment rate* flow.

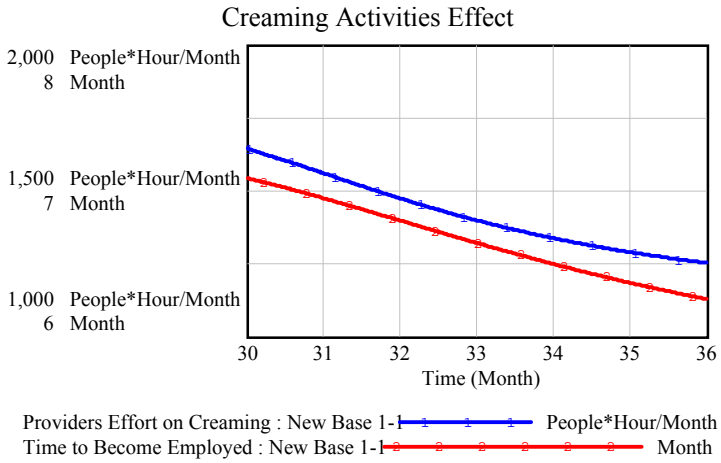


Figure 19.—Lying with Statistics

5. Conclusions and Future Research Opportunities

This simulation experiment generated useful behavioral information about the dynamics of JTPA enrollment and termination, learning processes, selection of systems of rules, and impact of rule-following preferences. Insights obtained from the study include: (1) JTPA behavior can be explored using dynamic models that capture the dynamics over time. (2) Feedback effects are crucial to determine the dynamics. (3) Other gamming paths besides the cream-skimming path should be studied further to analyze the possible impacts on system's overall performance and change. Finally, (4) Essential feedback mechanisms can be identified that are responsible for the dynamics of change in performance measurement systems. A preliminary 'generic' view of essential feedback structure extracted from the bigger JTPA model is presented in Figure 20. The formulation and exploration of that model is future research needed.

The model presented in figure 20 considerably reduces the complexity present in the JTPA model[19].

[19] The JTPA model has 826 variables organized in 70 model views.

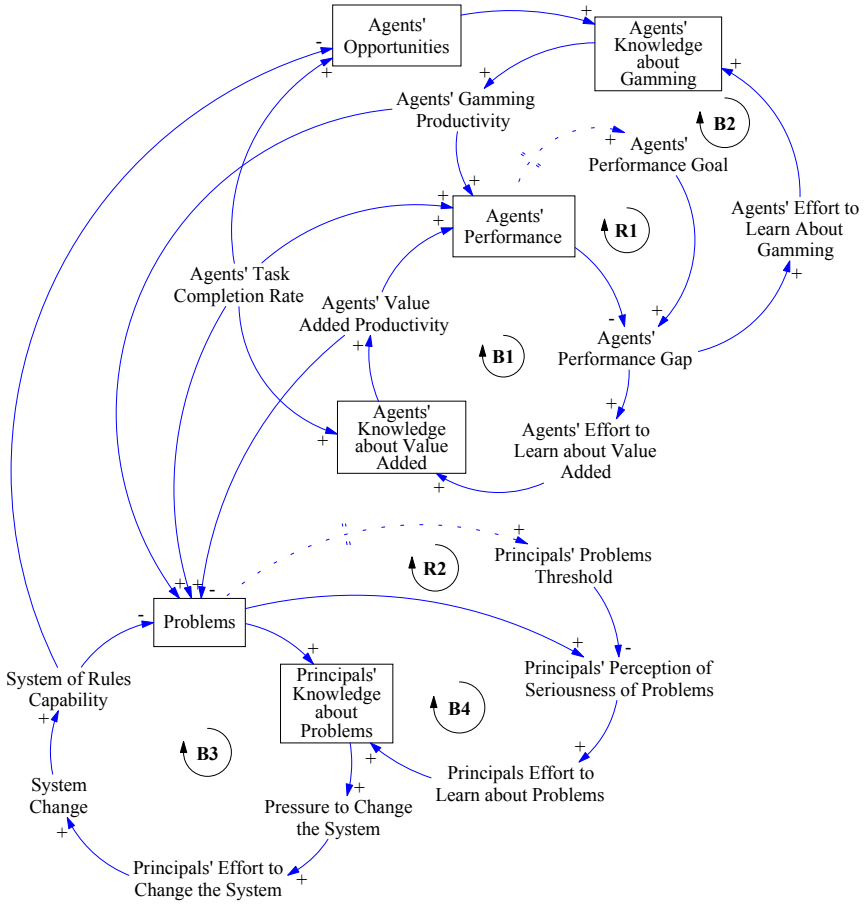


Figure 20.—A 'Generic' Model

References

1 (1982). Job Training Partnership Act. Public Law. **97-300 [S. 2036]**.

2 Austin, R. D. (1996). Measuring and Managing Performance in Organizations. New York, Dorset House Publishing.

3 Byrkit, D. R. (1987). Statistics Today: A Comprehansive Introduction. Menlo Park,California, The Benjamin/Cummins Publishing Company, Inc.

4 Coalition_on_Human_Needs (2003). Performance Measurement in JTPA**:** http://www.chn.org/issuebriefs/WorkforceInvestmentAct.asp.

5 Courty, P. and G. Marschke (1997). "Measuring Government Performance: Lessons from a Federal Job-Training Program." The American Economic Review **87**(2): 383-388.

6 Courty, P. and G. Marschke (2002). "An Empirical Investigation of Gaming Responses to Explicit Performance Incentives." Journal of Labor Economics.

7 Courty, P. and G. Marschke (2003a). "Dynamics of Performance Measurement Systems." Oxford Review of Economic Policy **19**(2): 268-84.

8 Courty, P. and G. Marschke (2003b). "Making Government Accountable: Lessons from a Federal Job Training Program." CMPO Working Paper Series No. 03/083: 1-33.

9 Courty, P. and G. Marschke (Forthcoming 2004a). "An Empirical Investigation of Gaming Responses to Explicit Performance Incentives." Journal of Labor Economics **22**(1).

10 Courty, P. and G. Marschke (Forthcoming 2004b). "Making Government Accountable: Lessons from a Federal Job Training Program." Public Administration Review.

11 Cragg, M. I. (1995). "Performance Incentives in Government Subcontracting: Evidence from the Job Training Partnership Act (JTPA)." Department of Economics. Columbia University: 1-32.

12 Frankel, B. (1986). Two extremes on the Social Science Commitment Continuum. Metatheory in Social Science. D. H. Fiske and R. A. Shweder. Chicago, The University of Chicago Press**:** 353-361.

13 GAO (1996). Job Training Partnership Act: Long-Term Earnings and Employment Outcomes. Washington, D.C., United States General Accounting Office**:** 38.

14 GAO (1998). Performance Measurement and Evaluation: Definitions and Relationships. Washington, D.C., United States General Accounting Office**:** 1-5.

15 GAO (2002). Workforce Investment Act: Improvements Needed in Performance Measures to Provide a More Accurate Picture of WIA's Effectiveness. Washington, D.C., United States General Accounting Office**:** 38.

16 Goldman, A. H. (2002). Practical Rules: When We Need Them and When We Don't? Cambridge, Cambridge University Press.

17 Hernes, G. (1976). "Structural Change in Social Procesess." The American Journal of Sociology **82**(3): 513-547.

18 Kim, D. H. (2004). Cream Skimming in the Federal Job Training Programs:Literature Review of Seven Empirical Research. Unpublished Manuscript. Albany, NY.

19 March, J. G., M. Schulz and X. Zhou (2000). The Dynamics of Rules: Change in Written Organizational Codes. Stanford, California, Stanford University Press.

20 Pauley, G. S. and R. J. Ormerod (1998). "The Evolution of a Performance Measurement Project at RTZ." Interfaces **28**(4): 94-118.

21 Prendergast, C. (1999). "The Provision of Incentives in Firms." Journal of Economic Literature **XXXVII**(1): 7-63.

22 Quinn, R. E. and J. Rohrbaugh (1983). "A Spatial Model of Effectiveness Criteria: Towards a Competing Values Approach to ORganizational Analysis." Management Science **29**(3): 363-377.

23 Schauer, F. (1991a). Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life. Oxford, Clarendon Press.

24 Schauer, F. (1991b). "Rules and The Rule of Law." Harvard Journal of Law & Public Policy **14**(3): 645-695.

25 Senge, P. M. (1990). The Fifth Discipline: the Art and Practice of the Learning Organization. New York, Doubleday/Currency.

26 Small, A. W. (1895). "Static and Dynamic Sociology." The American Journal of Sociology **1**(2): 195-209.

27 SPR (1999). Guide to JTPA Performance Standards for Program Years 1998 and 1999. Washington, DC, Social Policy Research Associates**:** 323.

28 Sterman, J. D. (1984). "Appropriate Summary Statistics for Evaluating the Historical Fit of System Dynamics Models." Dinamica.

29 Sterman, J. D. (1994). "Learning in and about Complex Systems." System Dynamics Review **10**(2/3): 291-330.

30 Sterman, J. D. (2000). Business Dynamics: Systems Thinking and Modeling for a Complex World. Boston MA, Irwin McGraw-Hill.

31 Stiefel, L., R. Rubenstein and A. E. Schwartz (1999). "Using Adjusted Performance Measures for Evaluating Resource Use." Public Budgeting & Finance **Fall 1999**: 67-87.

32 US_Department_of_Labor (1990). Training and Employment Guidance Letter No. 10-89. J. Roberts. Washington, D.C., US Department of Labor**:** 1-36.

33 US_Department_of_Labor (1993). Training and Employment Information Notice No. 3-92. R. T. J. A. S. o. Labor). Washington, D.C., US Department of Labor.

34 US_Department_of_Labor (1997). Training and Employment Information Notice No. 5-93 Change 2. B. A. Farmer. Washington, D.C., US Department of Labor.

35 Ward, L. F. (1883). Dynamic Sociology. New York, Johnson RCo.

36 Ward, L. F. (1895). "Static and Dynamic Sociology." Political Science Quarterly **10**(2): 203-220.

Back to the Top