December 1997

Toward A Systematic Evaluation of Large-Scale Information Systems: A Framework and Application

James Tien
Rensselaer Polytechnic Institute

Toward A Systematic Evaluation of Large-Scale Information Systems: A Framework and Application

Dr. James M. Tien
Professor and Chair
Department of Decision Sciences and Engineering Systems
Rensselaer Polytechnic Institute
110 Eighth Street
Troy, New York 12180-3590, USA
Telephone: 518-276-6486 Fax: 518-276-8227
Email: tienj@rpi.edu

Executive Summary
Despite the growing importance of information systems and the trillions of dollars expended in their continuing development and operation, organizations remain frustrated because they have yet to experience a significant productivity. What is usually missing is a valid evaluation step. It is generally recognized that the reason most evaluations fail – or are not valid – is because their research or evaluation designs are lacking. Although there is no stock evaluation design that can be taken off the shelf and implemented without revision, there should be a systematic process or framework by which valid and comprehensive evaluation designs can be developed. This paper provides such a framework, together with an actual application of the framework.

The proposed evaluation framework is based on a dynamic roll-back approach which consists of three steps leading up to a valid and comprehensive evaluation design. The "roll-back" aspect of the approach is reflected in the ordered sequence of steps. The sequence rolls back in time from (a) a projected look at the range of system characteristics (i.e., from its rationale through its operation and anticipated impacts); to (b) a prospective consideration of the threats (i.e., problems and pitfalls) to the validity of the final evaluation; and (c) a more immediate identification of the evaluation design elements. The sequence of steps can be stated in terms of two sets of links: the first set relates program characteristics to threats to validity, while the second set relates threats to validity to design elements. The "dynamic" aspect of the approach refers to its nonstationary character; that is, the approach must constantly be updated, throughout the entire development and implementation phases of the evaluation design.

The dynamic roll-back framework has been applied in several evaluation efforts, one of which is described herein. It concerns the evaluation of a U.S.-wide, $27.6 million criminal history records improvement (CHRI) program. In essence, the framework was adhered to in the development of a systematic CHRI evaluation design that was at once valid and flexible. The system characteristics included the three CHRI goals; a number of viable threats to validity were raised and carefully considered; and appropriate design elements were identified to mitigate, if not eliminate, the possible threats or rival hypotheses. In addition to providing the necessary feedback to national policymakers and State criminal records administrators, the CHRI evaluation determined that CHRI has improved the National criminal history system; improved inter-agency cooperation; and heightened awareness of the importance of criminal history records. Other critical evaluation findings include how States view data quality issues and approaches to improving data quality.

A final observation regarding the evaluation of a large-scale information system concerns the need to identify, say, a dozen primary measures that can appropriately characterize the performance of the system. In this manner, the system can be continuously monitored and, if necessary, appropriately refined.

Abstract
Despite the growing importance of information systems and the trillions of dollars expended in their continuing development and operation, organizations remain frustrated because they have yet to experience a significant productivity. What is usually missing is a valid evaluation step. It is generally recognized that the reason most evaluations fail – or are not valid – is because their research or evaluation designs are lacking. Although there is no stock evaluation design that can be taken off the shelf and implemented without revision, there should be a systematic process or framework by which valid and comprehensive evaluation designs can be developed. This paper provides such a framework, together with an actual application of the framework.

1. Introduction

It has been less than five decades since the first commercially available computer – the Universal Automatic Computer or Univac I – was built in 1951. Yet, computer technology has evolved through more than five generations, including the vacuum tube-based computers of the 1950s, the transistor-based computers of the 1960s, the integrated circuit-based computers of the 1970s, the very large-scale integrated circuit-based computers of the early 1980s, the ultra large-scale integrated circuit-based computers of the late 1980s, and the continuing trend to make computers faster, cheaper, and smaller. Advanced computer technology, together with improvements in optoelectronics, telecommunications, signal compression techniques, multimedia file servers, parallel processing, object programming, navigational aids, simulation methods, artificial intelligence algorithms, and sophisticated interface mechanisms, have transformed large-scale information systems from being the "glue" that holds the various units of an organization together to being the strategic asset that provides the organization with its competitive advantage.

Despite the growing importance of information systems and the trillions of dollars expended in their continuing development and operation, organizations remain frustrated because they have yet to experience a significant productivity improvement. Based on over a decade of assessing large-scale information systems, especially in the public sector, Tien (1986; 1988), Tien and Cahn (1986), Tien and McClure (1990), and Tien and Rich (1994) assert that what is usually missing is a valid evaluation step. Typically, an information system is planned, specified, coded, tested, implemented, operated, considered inadequate, and abandoned, followed by a new cycle of system planning, etc. A concurrently developed and implemented evaluation effort is generally inadequate, if not missing altogether; such an effort could provide a continuous feedback to the system administrator who must constantly decide whether the system should be refined, rejected and/or expanded.

The evaluation step is not only a necessary feedback mechanism, it should also provide answers to broader policy questions concerning the information system, including: Is the system working? Is the system producing the desired results? Are there better ways to achieve the same objectives? And, is the performance of the system commensurate with the resources it requires? Unfortunately, even in situations where evaluation efforts have been undertaken, they have been unable to provide valid answers to such questions. It is generally recognized that the reason most evaluations fail – or are not valid – is because their research or evaluation designs are lacking.

Although there is no stock evaluation design that can be taken off the shelf and implemented without revision, there should be a *systematic* process or framework by which valid and comprehensive evaluation designs can be developed. This paper provides such a framework in the next section, followed by an actual application of the framework, and then concludes with a key observation.

2. Framework

The evaluation framework that is detailed in this section was first advanced by Tien [1979]; it has since been successfully employed in a number of evaluation efforts – see, as examples, Colton, Brandeau and Tien (1987), Tien (1986; 1988), and Tien and Rich (1994). Indeed, this section summarizes Tien's earlier work (1979; 1990), and extends that work through its current emphasis on the evaluation of large-scale information systems. The remainder of this section addresses, respectively, the evaluation field, the evaluation approach, and the evaluation components.

Evaluation Field

In the United States, evaluation has emerged as a formal field of practice over the past three decades, beginning in 1965 when Congress passed the Great Society education legislation with an evaluation mandate. Since then the evaluation mandate has been attached to other social programs and big ticket items at all levels of government. Indeed, even some private organizations seek evaluations of their major investments. Evaluation is growing because it helps governments and organizations to make difficult choices by providing answers to such questions as identified earlier. The need for conducting evaluations becomes more critical as systems become more complex and more costly and, concomitantly, as the tax base or resources for funding these systems remains fixed or decreases.

As the evaluation of systems grows and plays a more central role in societal decisionmaking, it is critical that evaluations be conducted in an objective manner. This is indeed a challenge as opposing political, financial and environmental forces tug at the ethical values of evaluators. House (1993, p. xvi) cautions

against "clientism (taking the client's interests as the ultimate consideration), contractualism (adhering inflexibly to the contract), managerialism (placing the interests of the managers above all else), methodologicalism (believing that proper methodology solves all ethical problems), pluralism/elitism (including only the powerful stakeholders' interests in the evaluation), and relativism (taking all viewpoints as having equal merit)."

While evaluation has become a multidisciplinary field, it is not surprising that given its initial social program focus, its roots are in the social sciences, especially in the discipline of psychology. Obviously, the conduct of evaluation has extended beyond social programs and includes, as examples, technology assessments (Porter et al., 1980), evaluation of computer aids (Sage, 1981), and, of course, large-scale information systems.

The social sciences have also provided the field of evaluation with a common understanding and shared terminology. Basically, in regard to a program's or system's life cycle, it is assumed, first, that as noted earlier, the system design and its evaluation design should be developed concurrently (so that the system is indeed amenable to evaluation), and, second, that the traditional paradigm of evaluation is in effect (i.e., evaluation provides feedback to the system administrator). In regard to the evaluation process, every unit (i.e., subject, group, site or time period) can be designated as being either test or control. During the period of evaluation, pretest or pretreatment measurements are first made of both sets of units, followed by the administration of the system intervention on each test unit, and concluding with appropriate posttest or posttreatment measurements. There may, of course, be several test units, control units, system interventions, pretest measurements, and posttest measurements.

Evaluation Approach

The proposed evaluation approach is based on a dynamic roll-back approach which consists of three steps leading up to a valid and comprehensive evaluation design. The "roll-back" aspect of the approach is reflected in the ordered sequence of steps. The sequence rolls back in time from (a) a projected look at the range of system characteristics (i.e., from its rationale through its operation and anticipated impacts); to (b) a prospective consideration of the threats (i.e., problems and pitfalls) to the validity of the final evaluation; and (c) a more immediate identification of the evaluation design elements. The logic of this sequence of steps should be noted; that is, the anticipated system characteristics identify the possible threats to validity, which in turn point to the design elements that are necessary to mitigate, if not to eliminate, these threats. The sequence of steps can be stated in terms of two sets of links: the first set relates program characteristics to threats to validity, while the second set relates threats to validity to design elements.

The "dynamic" aspect of the approach refers to its nonstationary character; that is, the components of the approach must constantly be updated, throughout the entire development and implementation phases of the evaluation design. In this manner, the design elements can be refined, if necessary, to account for any new threats to validity which may be caused by previously unidentified system characteristics. In sum, the dynamic roll-back approach is systems-oriented; it represents a purposeful and systematic process by which valid and comprehensive evaluation designs can be developed.

In general, the characteristics of a system can be determined by seeking responses to the following questions: What is the system rationale? Who has system responsibility? What is the nature of system funding? What is the content of the system plan? What are the system constraints? What is the nature of system implementation? What is the nature of system operation? Are there any other concurrent systems? What are the anticipated impact findings? Again, it should be noted that the purpose of understanding the program characteristics is to identify the resultant problems or pitfalls that may arise to threaten the validity of the final evaluation.

After almost two decades, the classic monograph by Campbell and Stanley (1966) is still the basis for much of the on-going discussion on threats to validity. However, their original 12 threats have been expanded by Tien (1979) to include 8 additional threats. The 20 threats to validity can be grouped into the following five categories.

(a) Internal validity refers to the extent that the statistical association of an intervention and measured impact can reasonably be considered a causal relationship.

(b) External validity refers to the extent that the causal relationship can be generalized to different populations, settings and times.

(c) Construct validity refers to the extent that the causal relationship can be generalized to

different interventions, impact measures and measurements.

(d) Statistical conclusion validity refers to the extent that an intervention and a measured impact can be statistically associated – error could be either a false association (i.e., Type I error) or a false nonassociation (i.e., Type II error).

(e) Conduct conclusion validity refers to the extent that an intervention and its associated evaluation can be completely and successfully conducted.

In evaluation terms, the threats to validity can be regarded as plausible rival hypotheses or explanations of the observed impacts of a system. That is, the assumed causal relationships (i.e., test hypotheses) may be threatened by these rival explanations. Sometimes the threats may detract from the system's observed impacts. It is therefore the purpose of an evaluation design to minimize the threats to validity, while at the same time to suggest the causal relationships.

Evaluation Components

It is systematically convenient to describe a system evaluation design in terms of five components.

- The test hypotheses component is meant to include the range of issues leading up to the establishment of test hypotheses. In practice and as indicated in the dynamic roll-back approach, the test hypotheses should be identified only after the system characteristics and threats to validity have been ascertained. The test hypotheses are related to the rationale or objectives of the system and are defined by statements that hypothesize the causal relationships between dependent and independent measures, and it is a purpose of system evaluation to assess or test the validity of these statements. Poor definition of the test hypotheses can threaten statistical conclusion validity, since threats to validity represent plausible rival hypotheses.
- The purpose of the selection scheme component is to develop a scheme for the selection and identification of test groups and, if applicable, control groups, using appropriate sampling and randomization techniques. There are a range of selection schemes or research designs, including *experimental*, *quasi-experimental*, and *nonexperimental* designs. In general, it can be stated that nonexperimental designs do not have a control group or time period, while experimental and quasi-experimental designs do have such controls – even if it is just a before-after control. In terms of selection scheme factors which could mitigate or control for the various threats to validity, it can be stated that randomization is the key factor. In particular, most, if not all, of the internal and external threats to validity can be mitigated by the experimental designs which can only be achieved through randomization.

  There are two parts to the measures framework component. First, it is necessary to specify the set of evaluation measures which is to be the focus of the particular evaluation. Second, a model reflecting the linkages among these measures must be constructed. In terms of evaluation measures, Tien [1979] has identified four sets of measures – input, process, outcome and systemic measures. In general, the input and process measures serve to "explain" the resultant outcome measures. Input measures alone are of limited usefulness since they only indicate a system's potential – not actual – performance. On the other hand, the process measures do identify the system's performance but do not consider the impact of that performance. Finally, the outcome measures are the most meaningful observations since they reflect the ultimate results of the system. In practice and as might be expected, most of the available evaluations are fairly explicit about the input measures, less explicit about the process measures, and somewhat fragmentary about the outcome measures. Indeed, Nas (1996) argues that a key merit of cost-benefit analysis is that it focusses on outcome measures. The fourth set of evaluation measures – the systemic measures – can also be regarded as impact measures but have been overlooked to a large extent in the evaluation literature. The systemic measures allow the program's impact to be viewed from a total systems (i.e., organizational, longitudinal, programmatic and policy-oriented) perspective.
- The list of issues and elements which constitute the measurement methods component include measurement time frame, measurement scales, measurement instruments, measurement procedures, measurement samples, measurement quality, and measurement steps. Measurement methods which could mitigate or control for threats to validity include a multimeasurement focus, a long evaluation period (which, while controlling for regression artifacts, might aggravate the other threats to internal validity), large sample sizes, random sampling and pretest measurements.
- Analytic techniques are employed in evaluation for a number of reasons: to conduct statistical tests of significance; to combine, relate or derive measures; to assist in the evaluation conduct;

to provide data adjustments for nonequivalent test and control groups; and to model test and/or control situations. Next to randomization (which is usually not implementable), perhaps the single most important evaluation design element (i.e., the one which can best mitigate or control for the various threats to validity) is, as alluded to above, modeling. Unfortunately, most evaluation efforts to date have made minimal use of this simple but yet powerful tool. As an example, Tien and Cahn (1986) employed a linear statistical model to characterize a retrospective "split area" research design or selection scheme, which was then used to evaluate the program's impact.

3. Application

Since the establishment of the former Law Enforcement Assistance Administration in 1968, the U.S. Department of Justice has assisted States to improve the quality of their criminal history records. More recently, three Federal programs have funded activities in this area.

- The Criminal History Records Improvement (CHRI) program, administered by the Department's Bureau of Justice Statistics (BJS), provided $27.6 million over a three-year period (i.e., 1990-1992) to all 50 States, the District of Columbia, American Samoa, and the Northern Mariana Island.
- The Byrne 5% program, administered by the Department's Bureau of Justice Assistance, requires that as of 1992 at least 5% of the block grant monies to the States be reserved for criminal history records improvement activities – this amounted to $23 million in 1992.
- The National Criminal History Improvement Program (NCHIP), administered by the BJS, has disbursed $75.8 million to the States thus far in 1996; it is expected to disburse another $30 to $60 million in the next two years.

These Federal programs endeavor to help the States meet the mandates of a number of Federal statutes, including: the Anti-Drug Abuse Act of 1988 which requires the development of a system for the immediate and accurate identification of felons who attempt to purchase firearms; the Immigration Act of 1990 which requires States to provide conviction records of aliens to the Immigration and Naturalization Service within 30 days of conviction; the Brady Handgun Violence Prevention Act of 1993 which requires the establishment of a national instant criminal background check system; the National Child Protection Act of 1993 which requires that records of abuse against children be transmitted to the FBI's national record system; and the Violent Crime Control Act of 1994 which added an eighth firearm ineligibility category – consisting of persons who are "subject to a civil restraining order arising out of domestic or child abuse." (The other seven firearm ineligibility categories – as listed in both the Anti-Drug Abuse Act of 1988 and the Brady Act of 1993 – include (i) convicted felons, (ii) fugitives from justice, (iii) unlawful drug abusers or addicts, (iv) mentally defectives, (v) illegal aliens, (vi) dishonorably discharged, and (vii) citizenship renunciates.)

As might be expected, given the large amounts of Federal funds involved, a formal evaluation has been underway since 1992. Although the evaluation effort is still on-going, a first evaluation report was published (Tien and Rich, 1994) which summarizes the results of the initial $27.6 million CHRI program. This first evaluation effort is briefly described herein as an example application of the evaluation framework considered in Section 2. In particular, the evaluation design, the evaluation conduct, and the evaluation findings are discussed below.

Evaluation Design

In essence, the evaluation framework described in Section 2 was adhered to in the development of a systematic CHRI evaluation design that was at once valid and flexible. The system characteristics included the three CHRI goals of (i) enhancing State criminal history records in order to accurately identify convicted felons; (ii) meeting the new FBI/BJS voluntary reporting standards (1991) for identifying such individuals; and (iii) improving the quality and timeliness of criminal history records information. They also included issues related to the quality of criminal history records (SEARCH, 1985; SEARCH, 1992; BJS, 1992).

A number of viable threats to validity were raised and carefully considered, including extraneous events, design instability, instrumentation changes, multiple-intervention interference, measures sensitivity, design complexity, political infeasibility, and economic infeasibility. In each case, appropriate design elements were identified to mitigate, if not eliminate, the possible threat or rival hypothesis.

In regard to the evaluation components, it should be noted that an aggregate evaluation was undertaken over activities being implemented in 53 "States" or jurisdictions. The test hypotheses concerned the CHRI goals; the selection scheme was essentially a before-after consideration; the measures framework covered input, process, outcome and systemic measures; the measurement methods included document/plan reviews, workshops involving selected sites, telephone interviews, site visits, and assorted data collection efforts (i.e., computer printouts, surveys and on-site observations); and the analytic techniques ranged from simple analysis, to tests of statistical significance, to stochastic modeling.

As an example, Table 1 provides a summary of the various data quality measures considered; although only input, process and outcome measures are listed in the table, systemic issues were very much considered. Especially relevant were the inclusion of stakeholders in the evaluation effort; that is, police, courts, corrections, and probation personnel, as well as firearm dealers, were surveyed to obtain their insights into the accessibility and relevance of the improved records – which are obviously the ultimate reasons for improving records quality. In particular, it was ascertained whether the records impacted such day-to-day decisions as firearm licensing, employment, sentencing, release, bail, correctional classification, probation and parole.

Evaluation Conduct

The CHRI program consisted of 81 separate CHRI projects. Clearly, a consistent and uniform classification system for describing these efforts was needed. There were, of course, many ways of describing data quality improvement efforts. For example, one approach (SEARCH, 1988) defined four groups of data quality improvement strategies – administrative, data entry, data maintenance, and regulatory. The classification system developed for the CHRI evaluation effort consisted of 12 strategies; these *strategies* reflected broad, long-term objectives related to criminal history records, including: assessment of the current system, development of a data quality improvement plan, enactment of legislation, implementation of training programs, implementation of policies and procedures, automation of the central repository, automation of disposition reporting, automation of arrest reporting, automation of custodial reporting, identification of felons, improvement of the National system, and improvement of records accessibility. The most common CHRI-supported strategy was to improve automation of the central repository: 46 of 53 States used CHRI funds for this purpose. The only other strategy which more than half the States (29 of 53) addressed with CHRI funds was automation of disposition reporting.

At a more detailed level, 46 activities were identified for the 12 strategies. Each strategy contained at least one activity. Thus, for example, there were several different activities that States could undertake to carry out the strategy of automating the central repository, including developing the necessary automated systems (e.g., installing a computerized criminal history system), processing backlogs of fingerprint cards and disposition reports, and automating manual criminal history records. Similarly, to assess the current status of data quality, a State could implement four different activities – conduct a baseline audit, conduct a user needs assessment, document the current reporting system, and implement a data quality monitoring system. In the end, the 81 CHRI projects mapped into 213 total activities. The number of activities implemented by each State ranged from a low of one to a high of ten. The average number implemented was 4.1. The level of effort required to implement each activity varied widely; for this reason, the number of activities being implemented were not considered a measure of performance. As of April 1994, 31.9 percent of the 213 activities are still on-going. These on-going activities are in 22 different States.

Finally, three additional comments should be made regarding the evaluation conduct. First, it should be emphasized that the evaluation focused on CHRI-supported activities only. These activities obviously represented but a subset of all data quality improvement activities that States were implementing. Second, the fact that almost a third of all CHRI activities were still on-going meant that a complete "before/after" impact assessment could not be made. Third, by identifying all States that were carrying out a particular activity, the evaluation report (Tien and Rich, 1994) serves as a quick reference for those States which are considering the activity – a State may now contact one or more other States which have had actual experience in implementing the activity in question.

Evaluation Findings

In addition to providing the necessary feedback to national policymakers and State criminal records administrators, the CHRI evaluation determined that CHRI has:

- Improved the National criminal history system. CHRI funds have increased i) Interstate Identification Index (III) participation (five additional States had become III participants and four others hoped to become participants by the end of their CHRI project), ii) the level of automated disposition reporting to the FBI, and iii) participation in the Felon Identification in Firearm Sales Program.
- Improved inter-agency cooperation. A by-product of many of the CHRI-supported activities, particularly the planning, auditing, and electronic interface activities, had enhanced inter-agency cooperation. This cooperation was critical because coordinated multi-agency efforts are necessary to make systemic improvements to data quality. The BJA is to be applauded for requiring that States form multi-agency criminal record task forces as a part of their Byrne 5% program.
- Heightened awareness of the importance of criminal history record information. One CHRI project leader commented that the mere existence of a Federal criminal history records improvement program sent a strong message to State legislatures that "the quality of criminal history records is a priority." Increased awareness of the importance of criminal history records will inevitably lead to improved reporting.

It should be noted that much of the impact of the CHRI program had yet to be realized. First, CHRI funds were, as of April 1994, still supporting activities in 22 States. More importantly, in most States the same data quality improvement activities supported with CHRI funds are being continued with Byrne 5% and NCHIP funds. In fact, most States have viewed the Byrne 5% and NCHIP programs as merely extensions of the CHRI program. This finding is significant, since it indicates that States have appropriately used CHRI funds to address systemic quality problems and to initiate multi-year activities.

Other critical evaluation findings relate to how States view data quality issues and approaches to improving data quality. A brief, two-page questionnaire was distributed to the 53 States; 39 responded. Table 2 contains the questionnaire; the average score for each question is also tabulated. The questionnaire actually consists of two pairs of questions. Questions 1 and 2, both of which have 36 parts, address data quality issues; Question 1 asks to what degree each identified data quality issue is critical to an effective State criminal history repository, while Question 2 asks to what degree each data quality issue is a problem in the respondent's State. An interesting statistic is the computed difference between the Question 1 criticality score and the Question 2 problematic score for each issue; as indicated in Table 2(a), this difference can be considered an *alignment* score – that is, the greater the alignment score, the more the State feels that although the issue is critical, it is not a problem for the State (i.e., the State has *aligned* its priorities and activities to ensure that the critical issue is being dealt with and poses no serious problem for the State). Obviously, a low alignment score is cause for alarm; however, the two lowest alignment scores – 0.6 and 0.7 – are for Issues 23 and 22, respectively, which are not considered by the States to be very critical (i.e., their criticality scores are only 2.8 and 3.1, respectively). Clearly, then, one should be concerned with those issues that have high criticality scores and, at the same time, low alignment scores. The top five candidates which fall in this grouping are: Issues 30 ("degree to which arrests in database have a final disposition"), 14 ("degree to which final dispositions are submitted to repository"), 15 ("delays in submitting disposition reports to repository"), 31 ("degree to which each offender's felony conviction status can be determined"), and 17 ("delays in entering disposition data in criminal history database").

The Table 2(a) analysis lumps all the States together and considers only the averages across all State responses. Of course, the States are in very different situations data quality-wise. In light of this, it is not unexpected to see how differently States viewed specific data quality issues. As examples, while all States viewed the identification accuracy issue as very critical to an effective State repository, 27 of 39 States considered it "not a problem at all" (on the other hand, four States consider this most basic of all repository functions as a "very serious problem"); while 35 of 39 States viewed the submission extent of dispositions issue as very critical to an effective State repository, the degree to which it was considered a problem varied widely – at one extreme, 6 of the 35 viewed it as a "very serious problem," and at the other extreme, 2 of the 35 viewed it as "not a problem at all"; and, finally, while 34 of 39 States viewed the database completeness issue as very critical to an effective State repository, 12 considered it a "very serious problem" within their State.

Another point of interest is whether questionnaire respondents see their States as having one overriding data quality problem or several overriding problems. That is, do States see themselves as having one

weakest link in their data quality chain or several weak links? If a respondent saw his/her State as having a single weakest link, he/she could have, say, assigned the corresponding issue a 4 or a 5 and every other issue a 2 or a 3. On the other hand, if a respondent saw his/her State as having several weak links, he/she could have assigned, say, a 5 to several different issues and lower problematic scores to the remaining issues. Based on this criteria, ten of the 39 States had a single weakest link; the remaining 29 States had multiple weak links (i.e., more than one data quality issue had the highest problematic score given). Across all States, the average number of weak links was 5.8.

Questions 3 and 4 in Table 2(b) concern approaches to improving data quality. Both of these questions have 35 parts, each addressing a different approach to improving data quality; Question 3 asks the degree to which various approaches, if implemented, would improve data quality, while Question 4 asks the degree to which the activity was actually being implemented in the respondent's State. Again, an interesting statistic is the computed difference between the Question 3 importance score and the Question 4 implementation score; as indicated in Table 2(b), this difference can be considered a *need* score – that is, the greater the need score, the more the State feels that although the activity is important, it is not being implemented at a commensurate level (i.e., the State needs to further implement the activity to ensure that it is at a level commensurate with its importance to improve the State's data quality).

In terms of Question 3, the two activities that appear to be implemented most frequently across the country are: Activities 4 ("develop a long-term data quality improvement plan") and 9 ("improve inter-agency cooperation and commitment to data quality"). This is not surprising in light of the requirements of the BJA's Byrne 5% program (i.e., convening a multi-agency task force and developing a data quality improvement plan).

Are States implementing what they consider to be the activities that could yield the greatest improvement in data quality? In the case of the two activities mentioned above, clearly yes; that is, while Activities 4 and 9 have an equally high importance score (i.e., 4.1), they have at the same time low need scores (i.e., 0.3 and 0.4). On the other hand, there are several highly rated activities that, at the same time, have high need scores. The top four candidates which fall in this grouping are: Activities 22 ("implement live-scan fingerprint systems at local arresting agencies"), 24 ("upgrade/install new electronic interface between arresting agencies and prosecutors"), 26 ("upgrade/install new electronic interface between prosecutors and repository"), and 25 ("upgrade/install new electronic interface between arresting agencies and courts"). Not surprisingly, they are all in the high cost, automation-related area.
At a more detailed level, while 27 of 39 States felt that a court-repository interface would yield a major improvement in data quality, only 13 of these 27 States had major efforts going on in this area. On the other hand, 8 of these 27 States had no effort underway on court-repository interfaces – clearly, these States were going to be focusing on such projects in the future. In regard to live-scan fingerprinting systems, and as noted earlier, it is the activity with the greatest potential but one that has been implemented the least. In fact, 11 of the 20 States that saw live-scan systems as yielding major improvements in data quality had no effort underway in that area, no doubt because of the high cost involved.

4. Observation

In the continued development and operation of a large-scale information system, it is obvious that a major evaluation effort cannot be continuously carried out. Indeed, every initial evaluation effort should yield a set of, say, a dozen measures that can appropriately characterize the performance of the system.

In this vein, the information system can be diagnosed in a manner analogous to how a physician conducts a physical examination. At the start of the examination, the doctor checks a basic set of indicators (e.g., blood pressure, temperature, heart rate). If any of these indicators signals potential problems, measurements of other indicators – indicators that "dig deeper" into the body's systems – are taken (e.g., blood test, x-ray, catscan). If any of these secondary indicators suggest problems, then other tertiary indicators (e.g., colonoscopy, biopsy) may be ascertained. As the doctor digs deeper and deeper, the root cause of the problem or symptom is, hopefully, discovered. In other words, a layered approach is taken toward problem diagnosis. Similarly, for example, in assessing the "health" or quality of a criminal history records system, a layered approach could be employed, starting with broad, easy to obtain measures and continuing, if necessary, with more focused measures. In fact, three layers of measures – primary, secondary, and tertiary – would probably be sufficient. This approach should also suggest a method for combining at least the primary measures into, say, a Record Quality Index (RQI) that could be used to help assess the data quality status on an on-going, continuous basis, just as the

Dow Jones Industrial Average serves to gauge stock market performance on a continuous basis. States with low RQIs should acquire secondary and tertiary measures in order to identify appropriate strategies for improving their RQI.

Additionally, a critical component of the doctor's treatment is continued monitoring of the patient's health. Measurements of key health indicators are periodically made, and the patient's treatment is modified, if necessary. Similarly, continual, on-going monitoring of information systems could be promulgated. However, this can only occur if the system evaluation yields a pertinent set of primary, secondary, and tertiary indicators and demonstrates how the primary indicators can be combined into an overall system index. An effective information system evaluation must identify such pertinent indicators; otherwise, the evaluation would, at best, be a one-time assessment of the system's performance.

References

Bureau of Justice Statistics (BJS), *National Conference on Improving the Quality of Criminal History Records*, Washington, DC, NCJ-133532, 1992.

Colton, K.W., Brandeau, M.L. and Tien, J.M., *A National Assessment of Command, Control, and Communications Systems*, Washington, DC: National Institute of Justice, 1982.

Campbell, D.T and Stanley, J.C., *Experimental and Quasi-Experimental Designs for Research*, Chicago, IL: Rand McNally, 1966.

Federal Bureau of Investigation (FBI) and Bureau of Justice Statistics (BJS), "Voluntary Standards for Improving the Quality of Criminal History Records and Guidelines for Identifying Felony Convictions," *Federal Register*, Vol. 56, No. 30, 1991.

House, E.R., *Professional Evaluation: Social Impact and Political Consequences*, Thousand Oaks, CA: Sage Publications, Inc., 1993.

Nas, T.F., *Cost-Benefit Analysis: Theory and Application*, Thousand Oaks, CA: Sage Publications, Inc., 1996.

Porter, A.L., Rossini, F.A., Carpenter, S.R., and Roper, A.T., *A Guidebook for Technology Assessment and Impact Analysis*, New York, NY: North-Holland Press, 1980.

Sage, A.P., "A Methodological Framework for Systemic Design and Evaluation of Computer Aids for Planning and Decision Support," *Computation in Electrical Engineering*, Vol. 8, 1981, pp. 87-101.

SEARCH, *Data Quality of Criminal History Records*, U.S. Department of Justice, NCJ-98079, 1985.

SEARCH, *Strategies for Improving Data Quality*, U.S. Department of Justice, NCJ-111458, 1988.

SEARCH, *Report of the National Task Force on Criminal History Record Disposition Reporting*, U.S. Department of Justice, NCJ-135836, 1992.

Tien, J.M., "Toward a Systematic Approach to Program Evaluation Design," *IEEE Transactions on Systems, Man, and Cybernetics: Special Issue on Public Systems Methodology*, Vol. SMC-9, No. 19, pp., 1979, 494-515.

Tien, J.M., "On Automated Correctional Data Systems," *Computers, Environment and Urban Systems*, Vol. 10, No. 3/4, 1986, pp. 157-163.

Tien, J.M., "Computer Applications in U.S. Law Enforcement," *Computers, Environment and Urban Systems*, Vol. 12, No. 1, 1988, pp. 37-48.

Tien, J.M. and Cahn, M.F., "Commercial Security Field Test Program: A Systematic Evaluation of the Impact of Security Surveys," in *Preventing Crime in Residential and Commercial Areas*, Edited by D. P. Rosenbaum, Beverly Hills, CA: Sage Publications, Inc., 1986, pp. 192-206.

Tien, J.M. and McClure, J.A., "Enhancing the Effectiveness of Computers in Public Organizations Through Appropriate Use of Technology," *Public Administration Review: Special Issue on Public Management Information Systems*, Vol. 27, No. 7, 1986, pp. 553-562.

Tien, J.M. and McClure, J.A., "Towards an Operations-Oriented Approach to Information Systems Design in Public Organizations," *Information and Decision Technologies*, Special Issue, Vol. 16, No. 2, 1990, pp. 137-146.

Tien, J.M. and Rich, T.F., *Criminal History Records Improvement Program: Evaluation Report*, Washington, DC: Bureau of Justice Assistance, 1994.

Table 1
CHRI Program: Measures Framework

Category	Data Quality Measure	Definition
Input	Guidelines/Enabling Statutes	Degree to which the State has enacted legislation and appropriate agency statutes and guidelines to encourage and ensure data quality
	Policies/Procedures	Degree to which the State has implemented effective policies and procedures to enhance data quality
	Multi-Year Plan	Degree to which the State has an appropriate plan to improve data quality
	Agency Cooperation	Degree to which the various agencies within the State work together to improve data quality
	Leadership Commitment	Degree to which the pertinent leadership views data quality as a high priority issue
	Human Resources	Degree to which repository is adequately staffed, given current volume of work
	Technological Resources	Degree to which repository has adequate computer and related equipment
Process	Submission Extent	Degree to which reporting agencies submit fingerprint cards and disposition reports
	Submission Timeliness	Given records are submitted, time lapse between event occurrence and when record is received
	Submission Accuracy	Degree to which submitted reports have discrepancies with original source documents
	Update Timeliness	Given records are received, time lapse between record receipt and database update
	Update Accuracy	Degree to which received reports have discrepancies with database records
	Identification Accuracy	Degree to which offenders are accurately identified
	Identification Timeliness	Time lapse between receipt of fingerprint card and suspect identification
	Record Linkage	Degree to which repository accurately matches new arrest and disposition reports to persons already in database
	Record Completeness	Degree to which data elements in submitted reports are filled in
	Database Completeness	Degree to which repository has automated records of all recent arrests and dispositions
	Felony Flagging	Degree to which repository can determine whether offenders are convicted felons
Outcome	Record Accessibility	Degree to which criminal history information are available for use by other criminal justice agencies and other appropriate agencies, individuals, and employers
	Accessibility Timeliness	Given records are accessible, time lapse between record request and information receipt
	Record Utility	Degree to which provided information is relevant to the needs of the requestor

Table 2
CHRI Program: Data quality Questionnaire
(a) Criticality Versus Problematic Status

Q1. On a scale of 1 to 5, to what degree do you feel the following data quality issues are critical to an effective State criminal history repository?
(1 = not critical at all, 3 = fairly critical, 5 = very critical)
Q2. On a scale of 1 to 5, to what degree do you view the following data quality issues as problems in your State?
(1 = not a problem at all, 3 = fairly serious problem, 5 = very serious problem)

	Criticality Score Q1 (N=39)	Problematic Score Q2 (N=39)	Computed Alignment Q1 - Q2
Fingerprint card reporting by local arresting agencies			
1. Legibility of fingerprints	4.8	2.8	2.0
2. Degree to which all data elements on card are filled in	4.0	2.5	1.5
3. Accuracy of data elements on cards	4.6	2.4	2.3
4. Degree to which cards are submitted to repository	4.8	2.8	2.0
5. Delays in submitting cards to repository	4.1	2.9	1.2
Suspect identification by repository			
6. Accuracy of ident/non-ident decision	5.0	1.7	3.3
7. Delays in making ident/non-ident decision	4.2	2.2	1.9
8. Delays in rap sheet transmittal to arresting agency	3.5	2.3	1.2
Arrest data entry by repository			
9. Delays in entering arrest data in criminal history database	4.6	2.4	2.2
10. Accuracy of data entry of arrest data	4.9	1.9	3.0
Final disposition reporting process by courts or prosecutors			
11. Degree to which all data elements on disposition reports are filled in	4.5	2.7	1.7
12. Degree to which reports indicate whether person was convicted of a felony	4.2	2.5	1.7
13. Accuracy of data elements on report	4.9	2.5	2.3
14. Degree to which final dispositions are submitted to repository	4.8	3.1	1.7
15. Delays in submitting disposition reports to repository	4.3	3.1	1.2
Final disposition data entry by repository			
16. Degree to which dispositions can be linked to corresponding arrests	4.9	2.4	2.5
17. Delays in entering disposition data in criminal history database	4.3	3.0	1.3
18. Accuracy of data entry of disposition data	4.9	2.0	2.9
Record request processing by repository			
19. Delays in responding to requests by criminal justice agencies	4.4	1.7	2.7
20. Delays in responding to requests by non-criminal justice agencies	3.5	1.6	1.9
21. Readability of rap sheet	4.4	1.9	2.4
Alien conviction reporting to the Immigration and Naturalization Service (INS)			
22. Degree to which alien conviction information is reported to INS	3.1	2.3	0.7
23. Delays in reporting alien conviction information to INS	2.8	2.2	0.6
24. Accuracy with which offenders are identified as aliens	3.3	2.2	1.1
Reporting to the FBI			
25. Degree to which fingerprint cards are submitted to the FBI	4.1	1.8	2.3
26. Delays in submitting fingerprint cards to the FBI	3.8	1.9	1.9
27. Degree to which final dispositions are submitted to the FBI	3.8	2.7	1.1
28. Delays in submitting final dispositions to the FBI	3.7	2.6	1.0
Completeness of repository's criminal history database			
29. Degree to which database has all arrests from past five years	4.7	2.8	2.0
30. Degree to which arrests in database have a final disposition	4.8	3.7	1.1
31. Degree to which each offender's felony conviction status can be determined	4.4	3.0	1.4
32. Degree to which all records in master name index are automated	4.5	1.7	2.8
33. Degree to which all criminal history records from past five years are automated	4.4	1.9	2.4
34. Size of fingerprint card backlog	4.5	2.2	2.3
35. Size of FBI rap sheet backlog (for States that rely on FBI rap sheets)	3.4	2.1	1.3
36. Size of disposition report backlog	4.3	2.8	1.5

Table 2
(Page 2 of 2)
(b) Importance Versus Implementation Status

Q3. On a scale of 1 to 5, to what degree do you think a major effort in the following areas would improve data quality in your State?
(1 = minimal improvement, 3 = moderate improvement, 5 = major improvement)

Q4. On a scale of 1 to 5, to what degree are these approaches to improving data quality being implemented in your State?
(1 = no effort underway, 3 = moderate effort underway, 5 = major effort underway or completed)

	Importance Score Q3 (N=39)	Implementation Score Q4 (N=39)	Computed Need Q3 - Q4
Planning			
1. Conduct a baseline data quality audit	3.3	3.1	0.2
2. Conduct a repository needs assessment	3.4	3.0	0.4
3. Conduct a criminal justice information user needs assessment	3.4	3.2	0.2
4. Develop a long-term data quality improvement plan	4.1	3.8	0.3
Legislation			
5. Enact legislation requiring reporting of criminal justice data	2.5	2.0	0.5
Training			
6. Expand training programs in reporting procedures for local arresting agencies	3.8	2.8	1.0
7. Expand training programs in reporting procedures for prosecutors and courts	3.8	2.5	1.3
8. Expand training programs for repository staff	2.7	2.4	0.3
Cooperation and Commitment			
9. Improve inter-agency cooperation and commitment to data quality	4.1	3.7	0.5
Standardizing Procedures			
10. Implement standardized procedures for arrest reporting	3.4	3.0	0.3
11. Implement standardized procedures for disposition reporting	3.8	3.2	0.6
12. Implement procedures for improving fingerprint card processing at repository	3.0	2.9	0.1
13. Implement procedures for improving disposition report processing at repository	3.0	3.0	0.0
Automation			
14. Upgrade/install new computerized criminal history system	3.4	3.2	0.2
15. Upgrade/install new automated master name index system	2.5	2.5	0.0
16. Upgrade/install new AFIS system	3.7	3.1	0.6
17. Become a participant in the Interstate Identification Index (III)	3.0	3.0	0.0
18. Develop systems to monitor delinquent disposition reports	3.6	2.6	1.0
19. Upgrade/install new court information system	4.1	3.2	0.8
20. Upgrade/install new prosecutor information system	3.5	2.4	1.2
21. Upgrade/install new information systems at local arresting agencies	3.6	2.4	1.2
22. Implement live-scan fingerprint systems at local arresting agencies	4.2	2.1	2.1
Electronic Data Sharing			
23. Upgrade/install new electronic interface between arresting agencies and repository	3.6	2.3	1.3
24. Upgrade/install new electronic interface between arresting agencies and prosecutors	3.6	1.8	1.8
25. Upgrade/install new electronic interface between arresting agencies and courts	3.4	1.8	1.6
26. Upgrade/install new electronic interface between prosecutors and repository	3.6	1.9	1.7
27. Upgrade/install new electronic interface between corrections and repository	3.8	2.7	1.1
28. Upgrade/install new electronic interface between courts and repository	4.3	3.1	1.2
Data Entry			
29. Process fingerprint cards backlogged at repository	3.0	2.7	0.3
30. Process disposition reports backlogged at repository	3.0	3.0	0.0
31. Process FBI rap sheets backlogged at repository	2.0	1.7	0.3
32. Locate and process fingerprint cards not submitted to repository	3.4	2.1	1.3
33. Locate and process disposition reports not submitted to repository	4.0	3.0	0.9
34. Enter repository's manual master name index records in automated system	2.1	1.9	0.2
35. Enter repository's manual criminal history records in automated system	2.8	2.4	0.5