

Evaluating Judicial Education and Judicial Education Organizations: A Practical Guide for Measuring Courts' Performance and Rule of Law Programs

**Thomas Nelson Langhorne, Esquire
The Langhorne Group, Inc.**

I. Introduction and Terminology

Judicial organizations in established and emerging democracies face increasing pressures to objectively demonstrate programmatic effectiveness and efficiencies. As a consequence, courts must prove that resources devoted towards judicial branch education actually yield demonstrable, desirable outcomes. With increasing incidence, courts' inability to do so subjects courts' policy makers and judicial educators to critical public scrutiny and terminated funding.

This paper describes practical and affordable methods by which courts can determine what programs should be evaluated and how to evaluate them. The footnotes contained herein were carefully selected to maximize readers' ability to practically apply principles and methods discussed.

We specifically examine evaluation models and methodologies applicable to judicial training contexts. This paper develops court administrators' and judicial educators' organizational capacity to conduct reasonably priced judicial training program evaluations, thereby reducing their dependence upon expensive, external program evaluation experts.

At the outset, several terms used in this paper will be defined. This is necessary because professionals use various evaluation terms in inconsistent fashions, thereby causing confusion to even the most experienced audience. Accordingly, for internally consistent purposes, let us agree to several important terms' definitions as used in this paper.

First, for our present purposes, "program evaluation" means the systematic assessment of a training program's results, and within reasonable means, the extent to which the training program caused those results.¹ When program evaluators speak of assessing a program's results, they typically focus on the qualitative or quantitative measurement of program outcomes or impacts resulting from the program's delivery.

"Outputs" are generally the products produced by the program activity. Examples of "outputs" include the number of judges trained in a fiscal year, the number of courses offered during the year, and/or the number of educational textbooks or materials produced or distributed to judges. "Outputs" are of secondary importance when compared to "outcomes."²

¹ For a general and pragmatic examination of program evaluation methods and relevant terms defined, see Hatry, Harry P., Kathryn E. Newcomer, Joseph S. Wholey, eds. *Handbook of Practical Program Evaluation. 1st ed.* (San Francisco, CA: Jossey-Bass Publishers, 1994), 3, 40-94.

² Many terms and definitions used in this paper are borrowed from *Measuring Program Outcomes: A Practical Approach*, United Way of America (Alexandria, VA, 1996). Copies of this excellent handbook can be ordered by calling 800.772.0008 for a cost of twenty five dollars (U.S).

“Outcomes,” on the other hand, are the changes in behavior, attitudes or skills that benefit the students as a result of their training, or perhaps even more importantly, benefit the ultimate consumers/beneficiaries of the training. Examples of “outcomes” might include:

Court clerks process intake cases faster after taking a “docket management” course.

Judges more clearly explain their technical decisions to lay citizens after enrolling in a “Courtroom Communications and Demeanor” course.

Judges begin treating litigants with more respect after taking a “valuing cultural diversity” course. Judges treating litigants with greater courtesy and respect would be an outcome in and of itself. However, if post-training event exit surveys of litigants indicate litigants’ have improved public confidence in the courts because they were treated with greater dignity and respect, this too would be an “outcome” (an arguably higher level, greater outcome!). In fact, this example illustrates how some short-term outcome can lead to longer-term, macro outcomes.

Recalling our previous definition of “outputs” (e.g., number of judges enrolled in a class, volume of educational materials distributed, etc.), we generally hope that *outputs* will result in *outcomes*.

For purpose of clarity, it should be noted that typical course evaluations (those surveys typically administered to students to measure their satisfaction with their learning experience) rarely measure true outcomes. This is because course evaluations typically do not attempt to measure whether the student’s attitudes, behaviors or skills were actually changed or improved. Nor do students’ course evaluations measure whether the course tangibly benefited the student or the larger ultimate audience. This distinction will be explored in more detail in following sections.

“Impact” evaluations differ from “outcome” evaluations in an important way. True program “impact” evaluations use scientifically sophisticated methods to answer the following single question: Did the training event, and *only* the training event (to the exclusion of all other possible events and intervening variables) cause the outcomes to occur? This methodological distinction between “outcomes” and “impacts” illustrates why, as a general rule, conducting *impact* evaluations are very expensive propositions. Accordingly, very few corporations and even fewer public sector organizations attempt to engage in true *impact* evaluations. This paper will not devote attention to impact evaluations,³ given the magnitude of resources required to execute them, coupled with their reliance on highly sophisticated scientific methodologies. Unlike “impacts,” determining program “outcomes” does not present the highest degree of empirical confidence that your program, to the exclusion of all other possible intervening events or variables, caused a particular “outcome.” Nevertheless, determining program “outcomes” provides meaningful measures of a program’s progress (or lack thereof) at a reasonable cost, with manageable effort to court organizations. Harry Hatry once proclaimed,

“Trade-offs are typically required as policy makers and managers try to obtain valid, timely, but not overly expensive evaluation information. ‘It is better to be roughly right than to be precisely ignorant’. Program evaluation is an increasingly sophisticated field,

³ For a cogent analysis of case studies that use impact evaluation methodologies, including return on investment “ROI” analysis, this author recommends Lawrence B.Mohr, *Impact Analysis For Program Evaluation*, 2nd ed. (Thousand Oaks, California: Sage Publications, Inc., 1992).

but sometimes attempts to increase rigor have discouraged smaller scale and less expensive efforts to evaluate programs- leading to major gaps in the information available to both the public and those responsible for the service programs”.⁴

The term “outcome indicator” also has important utility to program evaluators. “Outcome indicators” are not “outcomes.” These indicators are measures that help you monitor your judicial education program’s progress towards meeting your program’s stated goal(s). Outcome indicators are your selected pieces of information, however defined, that allow you to *measure* observable programmatic changes or improvements. For example, let us assume you are delivering a judicial training course that instructs judges how to expedite case resolutions and/or reduce docket backlogs. After teaching that class, one outcome indicator you might monitor would be the court’s average number of days from initial case filings to final case disposition. If the average number of days until final case disposition drops from 300 days (the pre-training event average) to 225 days (the post-training event average), you would consider that a positive “outcome indicator.”⁵

The forgoing discussion will become more vivid and applicable in the following sections. Having defined some terms used in this paper, we can now turn our attention to making a compelling case for conducting court program evaluations.

II. Why Should Established and Emerging Democratic Courts Conduct Program Evaluations?

This section examines some of the more salient reasons why democratic courts need to routinely integrate program evaluation methods into their judicial education structures.

Ongoing Program Evaluations Provide Feedback Regarding Court Program Performance

From a program management perspective, courts that conduct ongoing, continual program evaluations (as measured by outcomes and outcome indicators) are better situated to make interim programmatic adjustments and corrections. Just as computerized medical monitors give critical feedback to doctors regarding patients’ progress and daily status, courts’ program evaluations may indicate interim program corrections to improve performance and better ensure programs’ vitality and viability.

Often, ongoing program evaluations during the life of a project provide performance warning signals- much like regular physical exams detect high blood pressure. Without valuable performance feedback, programs, like humans, can blithely maintain deleterious or unproductive practices without taking necessary corrective remedies.⁶ Simply put, ongoing program evaluations allow courts to modify, correct and improve program performance. A related benefit to initiating or designing outcome evaluations during the program’s *infancy* is that it forces

⁴ *Handbook of Practical Program Evaluation* at 1

⁵ An outstanding guide to conducting public sector program outcome evaluation is Harry P.Hatry and Mary Kopczynski, *Guide to Program Outcome Measurement for the U.S. Department of Education.*(Washington, D.C.,: Urban Institute: U.S. Department of Education, Office of Educational Research and Improvement, Educational Resources Information Center, 1997).

⁶ When appropriate, outcome evaluations should be conducted *during the life of and subsequent to* the life of a relatively long-term judicial training program. Often, courts execute program evaluations only upon *completing* a long-term program. For the purposes of this paper, the term “outcome evaluations” includes both *ongoing* evaluations during the life of the program and *post-event* evaluations as well. Methods for both, discussed herein, do not significantly vary.

courts to identify key external and internal program stakeholders. Ultimately, during the life of the court program, it is the program's stakeholders who often influence the program's success or failure.

Program Evaluations Allow Courts to Cope with the "Scarcity Reality"

Both emerging and long-established democratic courts must compete for scarce public funds. Courts that can objectively demonstrate positive outcomes result from their programmatic efforts enjoy a competitive advantage when standing in line at the public sector appropriations' trough. With sound outcome evaluations in hand, courts can more effectively communicate with and persuade policy makers who make difficult decisions between appropriation priorities. Without program outcome evaluations, courts are more vulnerable to the changing political winds that prevail during public budget cycles. Relatedly, courts that establish a reliable history of providing policy makers with outcome evaluations are better situated to justify *future* budget requests for new program initiatives..

Similarly, on a worldwide basis, virtually every democratic court is confronted with growing mandates for demonstrated accountability- both in terms of expenditure efficiently and producing demonstrable results. In this regard, the "scarcity reality" situation for *newly emerging*, USAID-funded democratic court programs is somewhat different from the "scarcity realities" *long-established* democratic courts routinely face. Because the roles that program evaluations play in these two types of democratic court milieus vary, they need to be separately examined.

"Transitional" democracies⁷ often receive enormous funding support from many international sources. Rule of law initiatives in these countries provide prime examples of court-related projects which face constant results-based scrutiny. These countries' court-related programs must demonstrate their accountability to demanding stakeholders and international sponsors. For example, virtually every transitional democracy's USAID rule of law program must conduct, and satisfactorily pass muster, various externally imposed program evaluations. For this reason alone, transitional democracies' court institutions must learn how to conduct program evaluations, or at a minimum, learn how to critically analyze programmatic evaluations' methodologies and findings.

Moreover, a cornerstone of rule of law programs is to build the host country's capacity to continue their democratic reforms without excessive reliance upon future foreign aid or foreign experts. Consequently, transitional democracies inherently possess a justifiable parochial interest in demonstrating their internal capacities for institutionalizing stable rule of law notions. Their natural desire to prove their capacities in this regard can best be persuasively demonstrated via executing program evaluations. Therefore, whether transitional democracies are answering to external stakeholders'/benefactors' demands for accountability or, alternatively, demonstrating their internal capacity to institutionalize democratic reforms, effectively conducting program outcome evaluations becomes an extremely valuable asset.

Long-established democracies and their judicial institutions face equally challenging, albeit different, "scarcity realities". For example, years before the September 11, 2001 terrorists attack on the United States, spiraling and endemic state budget crises compelled state legislatures to decimate state judiciaries' operating funds. Many state and federal courts faced unprecedented

⁷ The term "transitional democracy" is often used by USAID-funded and United Nations Development Program-sponsored rule of law initiatives to describe newly emerging democratic governments.

competition for diminishing public resources. Indeed, this keen competition for diminishing public resources is not exclusive to American courts. Well-established foreign courts are facing similar fiscal pressures. Slowly, a growing number of state courts are embracing program evaluation methods to demonstrate their programs' value-added contributions to society and stabilize their programs' funding.

The "Consumer Movement" Demands That Courts Produce Results

Converging simultaneously with the worldwide "scarcity reality" phenomena is the public sector's trend towards borrowing "outcome evaluation" practices from the private sector. The private sector has long practiced the art and science of program outcome evaluations in order to remain competitive and to satisfy consumer demands. The growing "consumer movement" has recently come to increase citizens' scrutiny of public programs as well.⁸ Essentially, the public has become a more "sophisticated" and demanding consumer of public programs and services. Courts are not exempt from "consumer" expectations. Concomitantly, the "consumer movement" has elevated citizens' general expectations regarding the quality of justice delivered, court services provided and return on tax dollar investments in court programs.⁹

Used appropriately, courts can use program outcome evaluations to become more responsive to citizens' expectations of and demands for efficient administration of justice and court-sponsored programs. This is especially true when courts properly identify and include key citizen groups ("consumers") in the program evaluation process. Equally important from a political perspective, including key citizen stakeholder groups in the evaluation process also forges public support for court programs. Virtually all of us involved in the court system have occasionally misjudged citizens' expectations regarding court programs and the manner in which court programs are operated. Conducting ongoing program evaluations helps courts align program outputs and outcomes with public expectations.

III. Basic Evaluation Models and Evaluation Concepts for Courts

Before one can effectively apply the outcome evaluation methods and principles detailed in sections IV and V, below, it is helpful to first appreciate the general theory upon which most program evaluation models operate (this is equally true for judicial training program evaluations). This section provides a basic foundation towards understanding shared principles of most program evaluation models.

There are many training evaluation models prevailing in the education and training domain. The shared overarching goals of virtually every training evaluation model are:

⁸ See Raymond C. Carey and Emil J. Posavac, *Program Evaluation: Methods and Case Studies*, 2nd ed. (Englewood Cliffs, N.J.: Prentice Hall, Inc, 1985) at 16-20

⁹ Clearly, judicial educators produce an end product far different from their private sector "corporate trainer" colleagues. Judicial educators try to produce more elusively definable "outcomes" such as improved access to justice, greater public confidence in the courts, elimination of cultural bias in judges' decision-making process. These noble outcomes are hardly analogous to manufacturers' production line "widgets". Judicial training program outcomes are accordingly more difficult to measure or qualitatively evaluate than most private sector outcomes. The only publication, to the author's knowledge, devoted exclusively to outcome evaluations in judicial education contexts is, *Conducting Impact Evaluation for Judicial Branch Education, JERITT Monograph Eleven*. M. E. Conner, T.N. Langhorne. East Lansing, Michigan: Michigan State University, 2002. Copies can be obtained for twenty five dollars (U.S.) by calling The JERITT Project at 517.353.8603. Monograph Eleven also contains sample evaluation forms and worksheets .

*Measure students' satisfaction with their learning experience, class materials, instruction and/or, learning environment. This is often referred to as "process" evaluation. Process evaluations are generally empirically unsophisticated and are primarily used to modify subsequent trainings or curricula.

*Measure whether learning actually occurred during the learning event (regardless of whether or not the students changed their behaviors or applied what was learned).

*Measure whether students' behaviors, attitudes were changed or learned skills were applied in the workplace after attending the learning event, and,

*If learning occurred and if behaviors, attitudes or skills were subsequently applied in the workplace, measure how the organization was impacted, improved or changed as a result of the training.¹⁰

In the judicial branch education field, perhaps the most prevalent evaluation model used is the Kirkpatrick Four Level Evaluation model.¹¹ For this paper's purpose, the Kirkpatrick model provides a most appropriate and practical framework.

Kirkpatrick characterizes the first evaluation level (and lowest level from an empirical, scientific perspective) as "reaction evaluation". Similar to "process evaluation", *reaction* evaluation measures students' satisfaction with (students' reaction to) the learning event.¹² This paper will not address "reaction" (process) evaluation methods.¹³ Rather, this paper is devoted to a second general type of training program evaluation, that which is called "*outcome*" evaluation. Levels two, three and four of Kirkpatrick's evaluation model generally fall within the "outcome evaluation" category.

Kirkpatrick's termed the second level of evaluation as "learning evaluation". This level reflects an incrementally improved evaluation measurement from "reaction level" evaluation because it aims at determining whether learning actually took place in the classroom. Stating the obvious, all judicial branch educators know that high student satisfaction with /reaction to a training event does not necessarily predict whether actual learning occurred. The most elementary example of "learning evaluation" is that of post-training event testing of students to measure the learning event's efficacy and students' retention of the course's learning objectives.¹⁴

Kirkpatrick's level three evaluation focuses on students' post-training event's *behavior*. This level of evaluation is not so much concerned with whether learning occurred, but rather, did the student's behavior change as a result of what was learned. Consequently, level three evaluation methods are often employed in the workplace (e.g., courtroom, clerk's office, public waiting rooms, etc.) in order to evaluate whether the learned attitudes, behaviors and/or skills are being

¹⁰ See footnote 8 at 3-5

¹¹ *Evaluating Training Programs: The Four Levels*, Second Edition, D. L. Kirkpatrick (1998). See also, JERITT Monograph Eleven at 7-15 for a synthesis of Kirkpatrick's four levels of evaluation.

¹² JERITT Monograph Eleven at vi-vii

¹³ For excellent examples of "reaction" level evaluation methods and actual "reaction" level forms, surveys and questionnaires to measure your judicial branch students' satisfaction with their courses, course materials and instruction, please see JERITT Monograph One, *Judicial Education Needs Assessment and Evaluation*, J. K. Hudzik: East Lansing, MI. (Reprint 1999). See footnote nine for instructions to order Monograph One.

¹⁴ JERITT Monograph Eleven at vi-vii : For specific, replicable examples of levels two and three evaluation techniques, see JERITT Monograph One, Appendix A

applied in the workplace environment. Learning which changes behaviors is more likely to produce measurable outcomes.

At the summit of Kirkpatrick's four-leveled evaluations, in terms of empirical sophistication, scientific reliability and design costs, is what he terms "results" evaluation. In essence, "results" evaluations attempt to answer the following evaluative question: Assuming we know that our students actually learned from and applied new behaviors in the workplace as a result of the learning event, how did those changed behaviors and improved applied workplace skills improve (impact) the court? Are judges' decision-making skills improved? Are culturally diverse litigants being treated with greater respect by court personnel? Can physically challenged citizens gain better access court facilities? Has public confidence in the judiciary improved?

Complicating this inquiry is that two requisite criteria comprising true "results" (impact) evaluations make this level of evaluation impractical for most courts. First, true "results" evaluations are capable of confirming the training event was the *sole and exclusive causative factor (or variable)* resulting in the positive impacts. This obviously requires the "results" level evaluator to exclude all potential extraneous intervening factors (all factors or events other than the training event) as contributing factors causing the improved court performance. The resource and expertise commitments necessary to satisfy this first criterion are prohibitive for virtually all courts.

Secondly, most true "results" evaluations also contain a return on investment (ROI) criterion. The benefits-to-costs ratio accurately calculates the courts' expended costs to actual benefits returned as a result of the training event. Conducting "ROI" evaluations can be a daunting challenge for the even the most experienced researcher or professional evaluator. It is even more difficult for courts to do so given the amorphous nature of the commodity courts dispense—"justice."

The forgoing discussion of Kirkpatrick's four level evaluation model justifies the following conclusion: The majority of courts should focus their "outcome" evaluations to levels two and three evaluation efforts. Court program and court training evaluations must be contemplated within "real world" pragmatic constraints. When choosing an evaluation method or design, one overarching question should always be asked: *How much evaluative information is enough given the costs associated with conducting various types of program evaluations?* The answer to this question will vary among courts' policy makers and will often vary depending on the stakes at hand. The positive news is this: courts can conduct useful outcome evaluations that are affordable and that simultaneously provide useful programmatic feedback. Evaluation levels two and three on the Kirkpatrick scale can be conducted by courts for reasonable costs and devoted resources. First, however, we must address the necessary question of *what* should courts measure?

IV. What Programs Should Courts Evaluate?

Deciding *what* programs a court should evaluate is a thresh-hold, fundamental question fraught with complex, fleeting considerations. On rare occasions, the decision will be an easy one- e.g., when the decision is externally mandated or when political realities make the choice self-evident. Moreover, few courts can expend the resources necessary to evaluate more than a few programs within a given fiscal year. Accordingly, choosing which program(s) to evaluate has enormous consequences for courts. The following suggests a sound, practical, concrete process for answering this cardinal inquiry.

How should courts go about deciding which programs to evaluate? One way we can answer this “what” question is to first ask *why* would any court, citizen’s group or interested stakeholder want to conduct a court program evaluation? The answer may be much more practical and straightforward than one might initially believe.

Measuring how a court is performing should serve as the ultimate justification and motivation for conducting program evaluations. If we maintain that concept of court performance as our guiding principle, we can more easily set about determining which programs we should evaluate. Stated otherwise, resource realities compel us to identify programs that, if evaluated, can yield significant information regarding the court’s performance. Having suggested this pragmatic approach to determining which programs to evaluate, how do we then proceed in defining and selecting our court performance criteria? Fortunately, excellent work in defining trial court and appellate court performance criteria has been done for us.

Recall our definition of terms in Section One. Specifically recall our discussion and definition of “outcome indicators”. We said that outcome indicators are “...measures that help you monitor your judicial education program’s progress towards meeting your program’s stated goal(s). Outcome indicators are your selected pieces of information, however defined, that allow you to measure observable programmatic changes or improvements.”

Various industries and professions have developed their own, unique indicators by which they measure organizational performance. Likewise, in the mid-1990’s, the National Center for State Courts Institute for Court Management devised five basic court performance standards. Each of the five standards can be separately measured by approximately twenty performance measurements. What is now commonly referred to as the Trial Court Performance Standards (TCPS), these five performance standards and accompanying measures constitute the seminal guide in evaluating how American state trial courts are performing.¹⁵

While the architects of TCPS readily acknowledge that no single set of standards and measures should be blindly applied by all courts in all circumstances, TCPS prevail as the “gold standard” by which American trial courts and court administrators can gauge courts’ performance. Currently, TCPS and their measures are also routinely applied in foreign rule of law program initiatives to measure those countries’ reform progress.

The five basic performance standards used by courts for self-improvement and evaluation are as follows:

1. Access to Justice (Are courts open and accessible?)
2. Timeliness and Expedition (Are court actions timely, not delayed?)
3. Equality, Fairness and Integrity
4. Independence and Accountability (Marries the notion of judicial branch independence under the separation of powers doctrine with the need for public accountability)
5. Public Trust and Confidence

Dr. Ingo Keilitz, a leading TCPS expert, once explained these five standards to me in the following fashion. “These five performance standards are to court program evaluators what a

¹⁵ For a complete discussion of the TCPS and accompanying measures (outcome indicators), read *Trial Court Performance Standards with Commentary: The Trial Court Performance Standards Project*. National Center for State Courts. Williamsburg, VA. 1990. For guidance as to how copies of this publication can be obtained, call the National Center for State Courts, 757.253.2000.

car's dashboard indicators are to drivers- basic yet critical performance signs by which one can gauge how the car is generally performing.” I like to add that they serve as a court's pole stars, forever providing court administrators reliable guidance through placid and turbulent seas.

But courts need specific “outcome indicators” by which they can micro-measure their performance within any of the five standards. Fortunately, a concise and specific measurement system accompanies each standard.¹⁶

Candidly, mastering the complexities of the TCPS measurements and becoming adroit in applying them for program evaluations exceeds this paper's scope. Nevertheless, in Section Five, below, we will choose one of the five TCPS (timeliness and expedition) and provide practical examples as to how program evaluators can easily use TCPS's measurements (“outcome indicators”) to evaluate court performance.

What to evaluate? Applying principles learned

We began this section by asking, “What programs should courts evaluate? We initially suggested that understanding *why* we conduct court program evaluations helps us answer that threshold question. The driving rationale and motivation, we also suggested, was to measure how courts are *performing*. When we use TCPS as our “court's dashboard light indicators” for court performance, we can better focus on what programs to evaluate. We can therefore begin to decide what court programs to evaluate by asking the following questions, guided by the TCPS:

1. What, if any, programs do our courts have which have as their goal to ensure citizens have adequate *access to justice*? (Then examine and apply the specific TCPS measures accompanying the “Access to Justice” Standard to help evaluate those programs' outcomes and overall performance.)

Sample Measures:¹⁷

- A. Are the hearings transparent and open to public scrutiny?
 - B. Can observers understand court calendars and track proceedings?
 - C. Can participants hear and be heard in court?
 - D. Are the facilities physically accessible, safe and convenient?
 - E. Are court personnel accessible by telephone? Is information available on the Internet?
 - F. Do lay citizens find doing business with courts is relatively easy?
 - G. Is effective representation provided to litigants?
 - H. Are the hearing impaired or those who do not speak the court's native language provided interpreter services?
 - I. Are the costs of access to the court's proceedings, in terms of money, time or the procedures that must be followed, reasonable, fair and affordable?
 - J.
2. What, if any, programs do we have which have as their goal to ensure our cases are *timely resolved* and are not fraught with unnecessary delays? (Then examine and apply the specific TCPS measures accompanying the “Expedition and Timeliness” Standard to help evaluate those programs' outcomes and overall performance.

:

¹⁶ Ibid

¹⁷ These sample, abbreviated measures are borrowed from those measures that accompany the Trial Court Performance Standards. See footnote 15.

Sample Measures

- A. Time to disposition- does your court compare favorably to the local, state and national averages for disposing of cases?
- B. Case disposition ratio- Is the court disposing as many cases as are filed in a given year?
- C. Age of pending caseload- Is the average age growing or shrinking?
- D. Certainty of trial dates- This measures the frequency with which cases scheduled for trial are actually heard on the date originally scheduled.
- E. Are attorneys meeting the pre-trial “meaningful events” deadlines imposed by the court?

3. What, if any, programs do we have which have as their goal to ensure our courts *treat all litigants equally* in ways that ensures the *court’s integrity*? (Then examine and apply the specific TCPS measures accompanying the “Equality, Fairness and Integrity” Standard to help evaluate those programs’ outcomes and overall performance.)

Sample Measures:

- A. Jury lists and selected jurors are representative of the jurisdiction from which jurors are drawn.
- B. Courts give individual attention to cases, deciding them without undue disparity among like cases and decide only upon legally relevant matters.
- C. Bail/bond decisions are made on a fair and equitable basis.
- C. Practicing attorneys view the court as being equitable and fair.
- D. The court unambiguously addresses issues presented to it and makes clear how compliance can be achieved.
- E. Records of court decisions and actions are accurate and properly preserved.
- F. The court faithfully adheres to the relevant laws, procedural rules and established policies.

3. What, if any, programs do we have which have as their goal to ensure our courts are *accountable and remain independent*? (Then examine and apply the specific TCPS measures accompanying the “Independence and Accountability” Standard to help evaluate those programs’ outcomes and overall performance.)

Sample Measures:

- A. The court maintains its institutional integrity and observes its jurisdictional boundaries while respecting the decisions and jurisdictional force of other courts’ decisions (comity).
- B. The court has reasonable control over its budget development and expenditures.
- C. The courts’ allocated resources are spent in ways that meet their objectives.
- D. The court engages in adequate media relations and has transparent operations and fair policies.

4. What, if any, programs do we have which have as their goal maintain the *public trust confidence in the courts*? (Then examine and apply the specific TCPS measures accompanying the “Independence and Accountability” Standard to help evaluate those programs’ outcomes and overall performance.)

Sample Measures:

(The accompanying TCPS measures evaluate how various constituent groups and stakeholders perceive the court and measure their degree of trust and confidence in the courts. Various

qualitative and quantitative methods for measuring same are found in the TCPS' accompanying appendix.)¹⁸

V. Getting Started With Program Evaluation: How and Where Do I Begin?

We now present a model evaluation design and process that court administrators, judicial educators and novice evaluators (and experienced evaluators as well) can easily apply. I have taught this basic model to American and foreign professionals who have never before attempted programmatic evaluations. The following approach and techniques seem to have been very well received and replicable by those adopting it. This following model is especially recommended for those conducting outcome evaluations of judicial branch training programs. However, its general approach and process can be transplanted to evaluating non-training programs as well.

An important tenet of this model holds that the most effective program evaluation design begins during *the early curriculum design phase*. Stated otherwise, when courts initially begin designing a judicial branch training course, courts should simultaneously contemplate how they will eventually evaluate the course's impact on students' behaviors, specifically, and the court organization (and perhaps the public) generally. The following provides a methodical, step-by-step approach to guide you from the early evaluation design stage to the evaluation execution stage. Actual case examples accompany the discussion of each step to help readers apply the principles in a practical fashion.

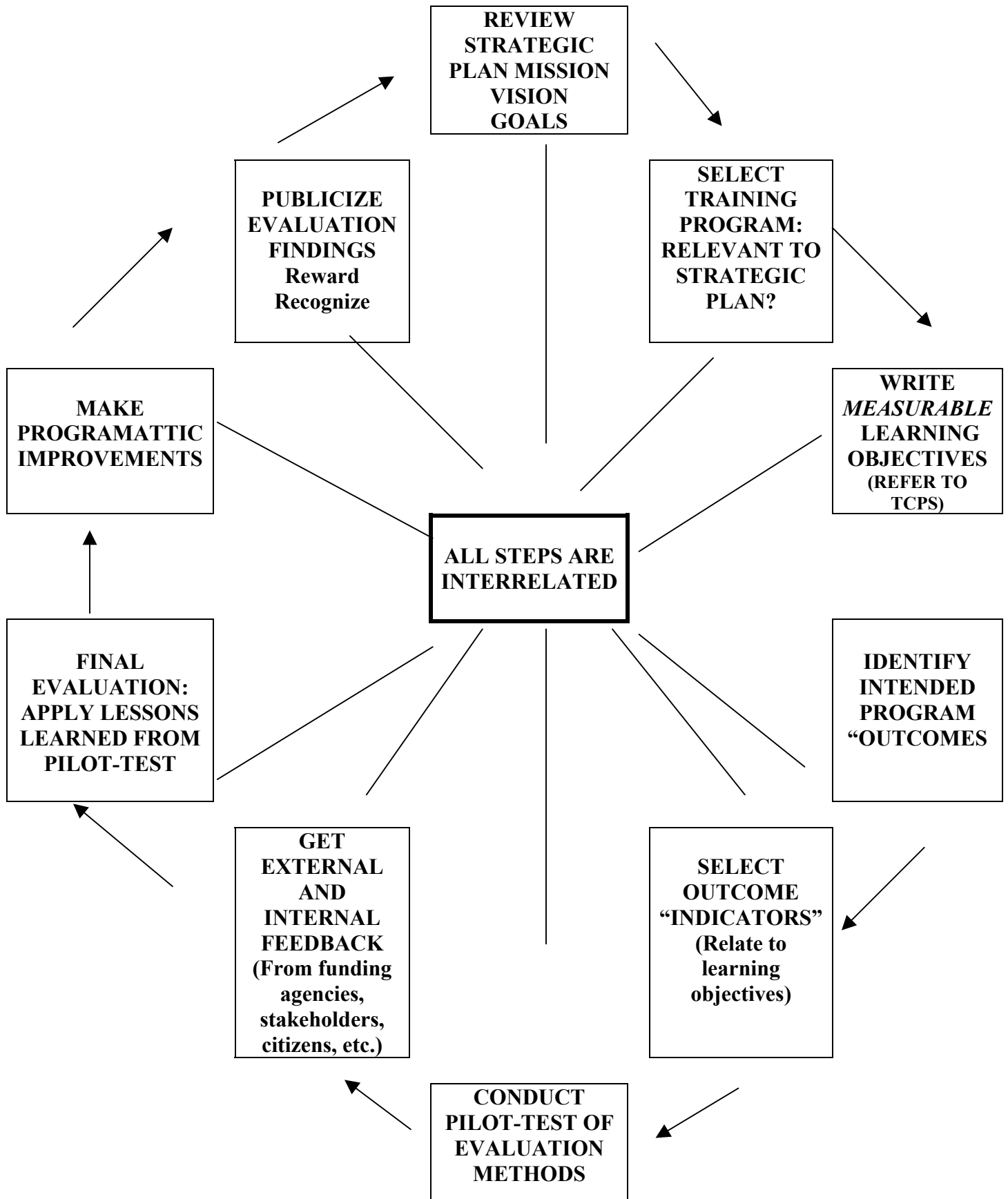
The Evaluation Circle

Figure 1, below, depicts our 360 Degree Evaluation Model. We will refer to various stages represented along the circle during our discussion of the evaluation process. We should first observe the importance of the model's configuration- the circle. A circle is chosen to illustrate the evaluation process for one practical and equally critical important reason: Ideally, the evaluation process should be continuous, always referring back and integrating information gathered from the steps along the circle. Consequently, the circle connotes the evaluation process is not merely linear. Instead, it is an integrated, ongoing process- the evaluator is always looking forward to the next step while looking in the evaluation process' rear view mirror for experiential guidance.

¹⁸ See footnote 16

Figure 1

THE 360 DEGREES EVALUATION CIRCLE



Step One- Refer to Your Court's Strategic Plan/Mission/Goals

Notice that the beginning point (and ending point) along the evaluation circle is represented by the court organization's "Strategic Plan/Mission/Vision/Goals". Most court organizations have adopted some form of a strategic plan by which it generally manages its program priorities and resources. Those courts that have not formally adopted a strategic plan virtually always have a mission statement(s), articulated organizational visions or goals that serve to articulate its primary objectives.

Keeping in mind that we are presently applying this 360 degrees evaluation model to evaluate the outcomes from judicial branch training program(s). Therefore, as a first step, begin by completely familiarizing yourself with your court's Strategic Plan. Appreciate, in a concrete way, what the Strategic Plan hopes to accomplish- whether it is larger aspirational goals or more discreet objectives. Be able to articulate and explain your court's goals or missions in a concrete way. To reiterate, this first step should be conducted before and during your course/curriculum design process. The reason for this is as practical as it is self-evident: Every course offered or curriculum designed should relate to and advance your court's goals. If your course/curriculum does not relate to or help advance your court goals, one must seriously scrutinize why you are offering or designing the course(s)! Continually relating back to your court's Strategic Plan ensures the course(s) relevance. Equally important, articulating an important connection between your court's goals and your course(s) improves your requests' appeal for training funds and empowers your training team.

As *Figure 1* depicts, it is useful to continuously seek and obtain internal and external feedback during the evaluation process. During the first information-gathering step, it is important to consult sources other than your court's Strategic Plan. For example, review recent legislation or appropriations bills containing narratives affecting your court's operations. Legislation often indicates what other important public bodies consider to be important goals for your court. Similarly, review your court's funding agencies' and other interested external stakeholders' strategic plans. Understanding their perceived priorities for court performance can be extremely useful. For example, an emerging or transitional democracy would want to review U.S. Congressional "rule of law" appropriations for various world regions to better understand basic performance expectations for newly created democratic courts. Likewise, examining USAID's or United Nations Development Projects' strategic plans often give clear indicia of future funding priorities. Of course, conducting face-to-face discussions or focus groups with public officials also provides rich information about overall goals and expectations affecting your courts.

Step 2- Write Your Training Course's Learning Objectives in Measurable Terms

After you have chosen a course that relates to and advances your court's goals, you must obviously begin writing your course learning objectives. The best practical advice to help you accurately draft your course learning objectives is to ask yourself the following question: "As a result of taking this course, our judges will be able to [insert an appropriate verb followed by a relevant object(ive)]." For example, suppose your court's major two-year strategic goal is to reduce case delay and backlog of cases. You rightfully decide to design and offer a "Effective Case Management and Delay Reduction Techniques" course for judges and administrators. Clearly, you have selected a course that advances a major court goal. When writing the course's

learning objectives, you must ask yourself what it is you want your judges to be able do as a result of taking this case management course.¹⁹ One could possibly choose to write the following measurable learning objectives: *As a result of taking this class, judges and administrators will be able to...*

1. (verb) Organize and group (object) similar cases together to create more streamlined and efficient pre-trial hearings.
2. (verb) Eliminate (object) by fifty percent the number of contested pre-trial discovery hearings order to expedite final case disposition.
3. (verb) Compel (object) more lawyers to timely meet pre-trial discovery deadlines.
4. (verb) Shorten children abuse cases' average time to final disposition from three years to eighteen months.

Forcing yourself to begin each learning objective with a verb and writing the object of that verb in measurable terms helps you accomplish the following: It frames learning objectives in a measurable way so that if the learning objectives are learned and applied in the workplace, the program's outcomes (e.g., fewer cases are delayed, case backlogs shrink, etc.) should also be observable and measurable during the subsequent program evaluation phase.

Step Three- Identify the Desired Program Outcomes Early in the Course Design

Recall that we define "outcomes" as those changed behaviors or improved skills that benefit the students, and hopefully, in turn benefits your court/public in observable, measurable ways. If the "thing" observed or measured does not represent a change in or tangible benefit to your students, it is not an outcome. This is because "outcomes" are not what the training program did (e.g., trained fifty judges in delay reduction techniques). Rather, outcomes are the consequential changes resulting from what the program did. If you have correctly completed step two by writing your learning objectives in observable and measurable terms, your task of identifying outcomes becomes relatively easy. In essence, correctly written learning objectives become a general statement of your intended outcomes. These general objectives can be refined and rewritten in more specifically measurable evaluation terms later in the evaluation process..

Step Four- Select Outcome Indicators

Each outcome identified in step three must be translated into specific outcome indicators that identify what is to be measured. These indicators will vary depending on your data source and data collection method used.²⁰ Recall that these indicators are specific measurements that allow you to chart your organization's improved performance. The primary question needed to asked when selecting your outcome indicators is, "What specifically ought to be measured?" It is helpful if you try to translate your indicators into numerical values that indicate progress toward achieving an outcome, such as a number, percentage or ratio. Hatry suggests several criteria for selecting outcome indicators and a useful checklist to ensure the criteria are appropriate.²¹

¹⁹ For a thorough discussion of writing measurable learning objectives see *Curriculum, Program and Faculty Development: Managing People, Process and Product*, JERITT Monograph Four. K. Waldrop, M.E. Conner. East Lansing, MI., Michigan State (1994).

Curriculum, Program and Faculty Development: Managing People, Process and Product, JERITT Monograph Four. K. Waldrop, M.E. Conner. East Lansing, MI., Michigan State (1994).

²⁰ *Guide to Program Outcome Measurement for the U.S. Department of Education at 31.*

²¹ *Guide to Program Outcome Measurement for the U.S. Department of Education at 31-32.*

Some Criteria for Selecting Outcome Indicators

1. Relevance to the program's mission/objectives and to the outcomes they are to measure.
2. Importance of what it measures.
3. The extent to which it might be duplicated by, or overlap with, other indicators.
4. Understandability of the indicator.
5. Feasibility and cost of collecting the indicator.

A Checklist for Outcome Indicators

1. Does each indicator measure some important aspect of the outcome?
2. Does each indicator begin with a numerical designation such as a number, percentage, rate, ratio, etc.?
3. Does your list of indicators cover all of the outcomes?
4. Does your list of indicators cover all the "quality" characteristics of concern to customers (users) of the program (such as timeliness)
5. Does your list of indicators include relevant feedback from program customers?
6. Is the wording of each indicator specific? (would a lay stranger understand your terminology?).

In our "Calendar Management and Delay Reduction" program, various outcome indicators satisfying the above criteria could be selected. One could use the actual number of cases awaiting trial in a given period, the average number of days a specific type of case has been pending after being filed with the court, the average number of contested pre-trial hearings conducted for particular case types. One might also measure the compliance rate by which important court-established pre-trial deadlines are timely met (e.g., for Court A, eighty five percent of all personal injury cases timely met the pre-trial deadlines for exchanging answers to discovery requests).

Step Five: Conduct Pre-training Event Evaluation Measurements

Optimally, if time and resources allow, courts should conduct pre-training event measures and observations. Conducting pre-training event measurements and observations provides a baseline to which post-training event evaluations can be compared. Taking pre-training event measurements and observations allows you to better measure improved organizational performance which occurs after the training event. For example, prior to delivering your "Calendar Management and Delay Reduction" course, you could research and determine the average time to final disposition in your students' courts. This could be accomplished fairly easily if those courts keep caseload statistics on a routine basis. If not, one could visit those courts and pull a statistically reliable sample of cases to determine the average time to final disposition. Several months after delivering your course, you would repeat those measurements to see if the average time to final disposition has been reduced.

Step Six: Identify the Sources of Your Outcome Indicators and Pilot the Data Collection

The sources of information for outcome indicators are so varied that it is beyond the scope of this paper to address them all. Similarly, thoroughly examining data collection methodologies are beyond this paper's scope. However, judicial training is essentially a human service aimed at, in turn, improving human (court) services. We can categorize data sources for human service programs into five basic types: records, specific individuals, the general public, trained observers and mechanical tests/measurements.²²

²² For detailed discussion of data collection methodologies and sample forms, surveys and qualitative research methods in program evaluations see JERITT Monograph Eleven at 37-98

Let us examine some examples of data sources for our “Calendar Management and Delay Reduction” course.

Records: Review case filings, motions, official caseload statistics, court dockets

Specific Individuals: Interview or administer written surveys to attorneys, litigants, court clerks, public defender offices, social services agency representatives.

General Public: Citizens who have recently appeared in court can be interviewed or surveyed regarding their experience or waiting periods.

Trained observers: Attend court hearings or sit in court public waiting areas and make observed measurements (using a stop watch, trained observers time the average number of minutes/hours citizens wait to have their case called or average time it takes to conduct a hearing).

After carefully choosing your sources of data and data collection methods, it is wise to “pilot-test” your evaluation design on one or a very few courts. Pilot-testing your evaluation methodology will allow you to accomplish the following prior to engaging in a full, formal evaluation project:

1. You will either discover flaws in your data collection methods or, at least, you will be positioned to extrapolate from the pilot-test experiences. This allows you to subsequently refine and improve your final program evaluation design.
2. You will eliminate extraneous evaluation processes and/or data collection in your final program evaluation.
3. You will more accurately estimate the costs of conducting your final evaluation based upon your pilot-test experience. This in turn allows to you modify the final evaluation design to better meet your funding realities and time constraints.
4. You will have a better assessment of the actual resources (personnel, materials, budgets, infrastructure) needed to conduct the final program evaluation.
5. Based upon what is learned and refined during your pilot-test evaluation, you can more persuasively convince policy makers for the need to conduct the final program evaluation.

Step Seven: After Conducting the Final Program Evaluation, Integrate Your Findings

After applying the forgoing principles and suggestions to both your “pilot-test” and your final program evaluation, you can use your evaluative findings in many useful ways. You will receive invaluable feedback from your measurements, observations, interviews and surveys. Those findings can be used to improve and refine the program you evaluated. In our case example, not only can the “Calendar Management” course can be improved, you can also integrate your evaluation findings into actual court processing and case management proceedings. This secondary benefit helps to positively change the administration of justice. For example, lessons learned during the evaluation process and the evaluation results/findings can be used to guide court staff and judges on an ongoing basis. Sharing the results can also be used to improve court morale and increase retention of employees. Where positive change in behaviors and actions can be measured, evaluation results can likewise be used to promote or advance exceptional court employees. Remember, evaluation results need not exclusively focus on areas needing improvement- it can also showcase where courts are excelling.

Publicly sharing your court’s evaluation results (both the positive and negative findings) further demonstrates your court’s commitment to transparent operations and public accountability. Allowing the public and policy makers to scrutinize your evaluation findings also distinguishes

your court during competitive public budgeting cycles. You will have shown your program does make a valuable difference in terms of measurable and positive outcomes. It additionally demonstrates your court can serve as responsible steward when spending its programmatic resources. All of this can be used to persuasively marshal support for increased funding requests. These beneficial byproducts, thanks to your program evaluation efforts, can now be objectively substantiated.

About the Author

Tom Langhorne is the co-founder of The Langhorne Group, Inc., an international consulting firm providing a wide range of consulting services to American courts and foreign countries' "rule of law " program initiatives. Some recently completed engagements include program monitoring and evaluation, curriculum design, and strategic planning for emerging democratic courts. An attorney, Tom Langhorne was a civil litigator with a national law firm and made formal appearances in many states' trial courts. He is currently an adjunct faculty and consultant to the University of Richmond's (Virginia) School of Law and that university's Business School. He is formerly the Director of Judicial Education for the Supreme Court of Virginia. In that capacity, he was charged with training all of Virginia's judges and court system personnel. The Langhorne Group's email address is rightinfo@aol.com. Tom can also be contacted by calling 804.323.3989.