



Office of Program Policy Analysis And Government Accountability

John W. Turcotte
Director

April 30, 1996

REVIEW OF THE DRAFT JUVENILE JUSTICE COST-EFFECTIVENESS INDEX

REPORT ABSTRACT

- Data compiled by the Department of Juvenile Justice for use in the draft juvenile justice index presents useful information about the performance of programs. However, when the Department adjusts and combines this data to compute a cost-effectiveness score, the results can be misleading due to the draft index design.
- The draft index does not provide a valid measure of program cost-effectiveness. It should not be used to rank juvenile justice commitment programs or evaluate their cost-effectiveness.
- The development of a more valid index will require the involvement of experts in cost-effectiveness evaluation and the collaboration of the Department of Juvenile Justice, program providers, and the Juvenile Justice Advisory Board.

PURPOSE OF REVIEW

This legislative session the House and the Senate initiated legislation (HB 405 and SB 738) requiring the Department of Juvenile Justice, in consultation with the Juvenile Justice Advisory Board (JJAB), to develop a cost-benefit analysis model to apply to all juvenile justice commitment programs. The Legislature requested our Office to evaluate the draft cost-effectiveness index that was developed by the Department with input from program providers and the JJAB. In this review we assess the validity of the draft index as a tool for ranking commitment programs and evaluating their cost-effectiveness.

BACKGROUND

The Department, program providers, and the JJAB identified information to consider when evaluating program cost-effectiveness. The Department selected five variables related to this information for which data is currently available through its existing data system. These five variables provide information on the seriousness of youths' offenses, the number of prior felony adjudications for program youth, the rate at which youth complete the program, the non-recidivism rate for youth released from the program, and information about program costs. The Department presents this information in data worksheets developed for use in the draft index.

The draft index computes program cost-effectiveness by calculating scores for each of the five identified variables and then adding them together to provide a total cost-effectiveness score. Appendix A (page 3) provides a detailed description of the draft index design and explains the procedures used for calculating the variable scores.

FINDINGS

Data Compiled by the Department Provides Useful Information. The Department assembles useful information in the form of data worksheets developed for use in the draft index. These worksheets move the evaluation of program performance beyond limited recidivism measures and represent a good start in the development of a cost-effectiveness index.

The Draft Juvenile Justice Cost-Effectiveness Index Does Not Provide a Valid Measure of Program Cost-effectiveness. Although the data worksheets provide useful information, the draft index itself lacks validity. The draft index adjusts and combines the data and computes program cost-effectiveness in a way that

results in misleading information about the comparative cost-effectiveness of programs. This is primarily due to the way in which the draft index is designed.

We reviewed the literature on cost-effectiveness and consulted experts in the area of evaluation design to identify basic criteria for assessing the usefulness and validity of the draft index. We found that the draft index does not adequately meet some of these criteria due primarily to a number of design limitations that include the following:

- The draft index adds up scores calculated for five variables to compute a total score labeled "cost-effectiveness index." This does not show the relationship between cost and effectiveness, which is conventionally expressed in a ratio.
- The draft index adjusts the data in a way that compresses results so that it is difficult to detect meaningful differences in the performance of programs.
- The draft index allows the influence of extremes or anomalies to skew cost-effectiveness results.
- The draft index does not provide a baseline to allow the comparison of program performance from one time period to the next.

In Appendix B (page 5) we describe in more detail the design limitations we identified in our review and explain how these limitations undermine the validity of the draft index. We also provide suggestions for improving the draft index design to alleviate these problems and develop a more valid index.

CONCLUSIONS AND RECOMMENDATIONS

The draft cost-effectiveness index does not produce the kind of information needed to make sound policy decisions. The manner in which the draft index computes cost-effectiveness results in information that can be misleading about the comparative cost-effectiveness of programs. This information should not be used to make funding decisions or rank programs. A more valid measure is needed for this purpose.

The Legislature has recognized the need for continual effort to develop a more valid index. To improve the draft index, a number of design and data quality problems will have to be addressed. Therefore, we recommend that the Department, in consultation with the Juvenile Justice Advisory Board, revise the draft index to address these validity and data quality concerns.

Cost-effectiveness evaluation typically requires the expertise of specialists. The development of a more valid index will require this level of expertise. We recommend that efforts to improve the draft index include the direct

involvement of a consultant or consultants with expertise in cost-effectiveness evaluation.

We also recommend that the Department and the Juvenile Justice Advisory Board continue to work in collaboration with program providers to further identify and refine variables to be included in the index.

AGENCY RESPONSE

The Secretary of the Department of Juvenile Justice (DJJ) provided the following written response to our review:

"Based on a legislative request, DJJ led a work group effort to develop a method of comparing program costs and effectiveness beyond just examining recidivism. The resulting index produces a high value for programs which deal with more serious offenders, are successful at getting youth through the programs and have lower recidivism rates, while demonstrating relatively low costs. The work group agreed upon what were thought to be the best currently available measures for these factors.

"The results of this method may not be as precise as we would wish, but they do allow us to distinguish excellent programs from those which are average and certainly from those which are performing poorly, especially when the index values are compared to DJJ Quality Assurance ratings. This allows DJJ to identify programs which may require special attention and gives an indication of areas where technical assistance is needed.

"We will continue to seek better methods to combine important variables into 'accountability scores' which can be used to compare program performance. We will stop referring to our measure as cost effectiveness to avoid confusion with cost benefit ratios, which have long been out of favor for policy analysis purposes due to problems of internal consistency. We feel that combining ordinal rankings is an approach which has more promise to provide meaningful results through a simplified technique while recognizing the limitations of current data.

"This effort to improve accountability has involved experts in the treatment of delinquents, program management, budgeting, measurement, statistics, criminology and economics. The department intends to continue to work collaboratively on improving the model. Additional outside expertise will be sought. The legislature has appropriated funding for development of a new juvenile justice information system. This system will vastly improve the quantity and quality of data needed as this accountability model evolves."

Appendix A

Description of the Draft Juvenile Justice Cost-Effectiveness Index Design

How Cost-Effectiveness Is Computed

The draft juvenile justice cost-effectiveness index is made up of five variables adjusted to fit between 0 and 1 and then added together to produce a total cost-effectiveness score. See Table 1 below.

The first two variables in the draft index are based on the delinquency histories of program youth and are intended to indicate the level of difficulty of the youth in the program. The third variable is

based on the rate at which youth complete the program. The fourth is based on the percentage of released youth who did not recidivate during the follow-up period. And the fifth variable is based on the average cost per youth completing the program. The scores for each of these variables are added to provide the total score in the sixth column labeled "cost-effectiveness index."

Table 1
Example of Department of Juvenile Justice Adjusted Variables
And Summary Draft Cost- Effectiveness Index

Level 6 Programs	Seriousness of Commitment Offense Value	Prior Felony Adjudications Value	Successful Program Completion Rate	Non-Recidivism Rate	Cost Per Successful Completion Value	Cost Effectiveness Index
Peace River						
Outward Bound	0.60	0.71	0.81	0.57	0.89	3.58
Withlacoochee						
Stop	0.55	0.36	0.86	0.65	0.94	3.35

Data Worksheets

Table 2 below shows an example of a portion of a worksheet with the data for each of the five variables used in the draft index before calculated scores have been adjusted to fit between 0 and 1.

Table 2
Example of Department of Juvenile Justice Data Worksheet

Level 6 Programs	Average Weighted Seriousness of Commitment Offense		Average Number of Prior Felony Adjudications		Successful Program Completions		Recidivism		Program Expenditures	Average Cost Per Successful Completion
	N	Average Weight	N	Average Adjudications	N	Percent	N	Percent		
Peace River										
Outward Bound	85	4.84	85	7.11	69	81.18%	30	43.33%	\$660,579	\$9,574
Withlacoochee										
Stop	120	4.37	119	3.59	103	85.83%	65	35.38%	563,354	5,469

Procedures for Calculating Adjusted Scores From the Data

The Department of Juvenile Justice (DJJ) provides the following explanation of the procedures used to calculate and adjust scores from the data:

- " ■ **Seriousness of Commitment Offense.** The method for calculating this measure was a simplified technique based upon the concept of rating offenses according to an estimate of relative seriousness. All points within the range were not used because current offense categories available in the existing information system do not allow for sufficient differentiation in terms of how serious comment offenses actually are, e.g. all types of burglary are subsumed within one single category designation of burglary. For each youth released during FY 1993-94, a point value corresponding to the most serious offense for which a youth was committed to a program, was assigned as follows: violent and sex felonies, 8 points; other felonies, 5 points; misdemeanors against persons, 2 points; and, other misdemeanors and "other delinquencies," 1 point. For each program, the assigned offense point values were summed and then divided by the total number of youth released from that program to get the average value for weighted seriousness of commitment offense. This average weighted seriousness of commitment offense for each program was then divided by 8 (the maximum point value) to create an index value between 0 and 1. Higher values indicate youth with more serious commitment offenses.
- **Prior Felony Adjudications.** The average number of prior felony adjudications for youth released from each program during the year was calculated. To obtain the measure used in calculating the index, the average for each program was divided by the highest average for any program within the same restrictiveness level. Higher values indicate youth with more extensive delinquent histories.
- **Successful Program Completion Rate.** This measure represents the number of youth who successfully completed each program divided by the total number of youth released from that program. A youth was counted as a successful completion if their reported release code indicated a program complete, being placed on furlough or post-commitment community control or that the youth was transferred to a commitment program at a lower restrictiveness level. Higher scores indicate higher degrees of success in getting youth through a program.
- **Non-Recidivism Rate.** The cohort that was tracked for recidivism consisted of those who successfully completed the program between January and June 1993. The value used in the calculation of the overall index was the number of youth who were not recommitted to DJJ, not placed on adult probation and not sentenced to adult prison for an offense committed within one year of their release. This number was then divided by the total number of youth who successfully completed the program to yield a rate of youth who did not meet the definition of recidivist. Higher values indicate more youth who did not meet that definition.
- **Cost Per Successful Completion.** The average cost per successful completion was determined by dividing the total cost of the program by the number of successful completions. Each average was divided by the highest average cost within the restrictiveness level in order to create an index value between 0 and 1. Finally, this figure was subtracted from 1 so that high scores indicate low cost per successful completion."

Source: Department of Juvenile Justice. Management Report Number 25: Cost-Effectiveness of Juvenile Justice Commitment Programs, March 1996.

Appendix B

Design and Data Limitations of the Draft Juvenile Justice Cost-Effectiveness Index

<i>DESIGN ISSUES</i>	D i s c u s s i o n	S u g g e s t i o n s
<i>The Index should show the relationship of cost to effectiveness.</i>	<p>Cost-effectiveness is usually expressed in a ratio to show the relationship between cost and effectiveness in a way that is easy to interpret. The draft index does not provide a cost-effectiveness ratio.</p> <p>The draft index adds the scores for five variables to provide a total "cost-effectiveness" score. But this total score does not describe the relationship between cost and effectiveness.</p>	<p>Use standard cost-effectiveness methods to compute a ratio like this:</p> $\frac{\text{computed cost}}{\text{effectiveness measure}}$
<i>The Index should adjust for differences in type of youth served.</i>	<p>In the draft index, two of the five scores added together to produce a total cost-effectiveness score are scores for the "level of difficulty" of the youth in the program. This treats difficulty as if it adds to or contributes to cost-effectiveness. However, this is not what is assumed about how difficulty affects program performance.</p> <p>A procedure for adjusting the effectiveness score to reflect the level of difficulty of the youth in the program is required. This would address the providers' concern that, even within levels, it would be unfair or misleading to compare programs as if the youth in different programs were all the same.</p>	<p>Use standard cost-effectiveness methods to compute a ratio adjusted for level of difficulty. This would change the above ratio to look like this:</p> $\frac{\text{computed cost}}{\text{effectiveness measure} \times \text{difficulty factor}}$ <p>Another option would be to group programs within each level according to difficulty of youth (high, moderate, low) and compute the cost-effectiveness ratio for programs in each group. This will allow the comparison of the cost-effectiveness of programs with comparable youth.</p>
<i>The Index should allow the detection of meaningful differences among programs.</i>	<p>The draft index is computed by adding scores that have been adjusted to fit between 0 and 1. This results in differences between program scores that may be too small to be meaningful or to help compare programs.</p> <p>It is sometimes useful to standardize or transform evaluation measures to keep them in scale. Differences in scale could overwhelm the calculation of cost-effectiveness and overstate differences. But the procedure used to transform scores to a similar scale should not understate differences either.</p>	<p>Determine if transformed scores are needed to avoid "out-of-scale" distortions. Use scores based on raw data if transformed scores are not necessary. If transformed scores <u>are</u> needed, use standard measurement methods to adjust for differences in scale.</p>

DESIGN ISSUES**D i s c u s s i o n****S u g g e s t i o n s**

The Index should reduce distortion from extremes or unusual cases.

The draft index is computed with averages and these averages are then divided by the highest average for any program within the same restrictiveness level. This allows extremes to drive the scores and may result in data that is skewed or misleading.

Avoid a design that allows averages or anomalies to distort or skew results.

The Index should base results on a meaningful number of cases.

The draft index allows scores based on very small numbers of youth to distort results. Broward Group Treatment Home, for example, released two youth who did not recidivate, and the program got a perfect (highest possible) score for non-recidivism: **1.0**. But if one of these youth had recidivated, the rate would be reduced by half to **.50**. When scores are computed on the basis of only a few cases, results like these can be very misleading. Programs may achieve a high total cost-effectiveness score based on what has happened to only one or two youth.

Conduct the study every two years to increase the number of cases. This will also help establish trend data to increase the reliability of results.

Another way to increase the number of cases to make recidivism rates more meaningful is to use the number of youth released for a full year instead of six months.

A way to deal with the effect of extremely small numbers without increasing the number of cases would be to group programs according to number of cases and compute the cost-effectiveness of programs within each group. This would allow the identification of less meaningful results based on small numbers, which should be used with caution.

The Index should allow the comparison of results from one time period to the next.

A baseline is needed to make useful comparisons about program performance from year to year. Without a baseline, one cannot tell if programs are improving over time or getting worse.

Do not use calculations that limit the comparison of scores to only one year.

The draft index uses highest average program scores to compute the cost-effectiveness of all the programs in each level. This means that results are relative to whatever the highest average score may be for that year and cannot be used to compare the cost-effectiveness of programs from year to year.

Use an expert review process to identify "benchmarks" or standards to use to compare programs with each other and over time.

**DATA QUALITY
CONCERNS**

D i s c u s s i o n

S u g g e s t i o n s

The Index should use more valid measures of program success.

"Successful" completion indicates how many youth completed a program, but it does not capture program effectiveness. Programs completing youth at a high rate may or may not be highly cost-effective.

Successful completion includes youth who are honorably discharged from a program or released or transferred to a less-restrictive program. But "honorable discharge" does not mean the same thing for all programs. In some programs, youth are not graduated until they meet demanding requirements such as increasing their reading level by two grade levels. Other programs release youth when their required length of stay is completed. Youth who are successful in a program but are referred to another program for additional treatment due to special needs may not be counted as "successful" despite gains they have made.

Avoid designs that treat program completion as a measure of program effectiveness.

Differences in discharge criteria could be used to provide more information about program effectiveness if used in the following way:

Survey programs to identify differences in honorable discharge criteria. Group programs accordingly when evaluating program effectiveness. This would provide additional information to compare the effectiveness of programs with different discharge criteria.

The Index should measure program costs with more reliable data.

Cost data is based on State Automated Management Accounting System (SAMAS) expenditures and cannot be used to compare all programs. Due to the way the Department assigns SAMAS codes, expenditures for some programs cannot be broken out from expenditures for other programs, so these programs are left entirely out of the draft index.

The costs to other state agencies are not included in the cost data. If these state costs contribute to program outcomes, they should be included.

The draft index does not include information about how many of the beds or slots in a program are being filled. Some beds and slots are paid for whether they are full or empty. Programs with many empty beds may have more resources to spend on each youth they serve. Programs like this may score more cost-effective than programs with no empty beds, but they may really be more wasteful.

Until cost issues are resolved, measure program effectiveness only.

Require separate contract numbers and SAMAS codes for every provider program.

Require providers to provide complete cost data for individual programs. Integrate this effort with the upgrading of the Department's data management system.

Include costs to other state agencies.

Until cost issues are resolved, consider other means of comparison. For example, use the number of budgeted client service days (number of days X number of clients per day). Using this option, one would include information about the utilization of funded beds as well as measures of program effectiveness.

**DATA QUALITY
CONCERNS**

D i s c u s s i o n

S u g g e s t i o n s

The Index should use more valid and well-defined measures of youth difficulty.

Level of difficulty has not been explicitly defined for use in the draft index. Does it mean more likely to recidivate, more difficult to handle, more likely to leave the program, or something else? The key to a valid and reliable index is the clear definition of terms or variables it includes.

If level of difficulty refers to the likelihood of re-offending, the draft index appears to assume things about seriousness of offense and the number of prior felony adjudications that may not be true. The number of prior misdemeanor adjudications for program youth is not included in the index. The literature on youth recidivism identifies non-violent chronic offenders of less serious crimes as being at high risk of re-offending. Youth who commit the most serious or violent crimes may not be as likely to re-offend.

The index does not include the age of youth in a program as a level of difficulty factor. The consensus among program providers is that older offenders are more responsive to treatment than younger offenders. Recent research has also identified a number of other factors that appear to contribute to delinquency and recidivism. These include peer group delinquency, level of family attachment, poor school performance, and history of abuse.

Decide what is meant by level of difficulty and measure it accordingly.

Test or validate assumptions about the effect of difficulty of youth on program outcomes before using these assumptions as a basis for the design of an index.

Include the number of prior misdemeanors in measures for likelihood of re-offending.

Test or validate assumptions about the effect of age on program performance. If appropriate, include age as a factor in the level of difficulty measure.

Consider the development of a more research-based level of difficulty measure.

This project was conducted in accordance with applicable evaluation standards. Copies of this report may be obtained by telephone (904/488-1023), by FAX (904/487-3804), in person (Claude Pepper Building, Room 312, 111 W. Madison St.), or by mail (OPPAGA Report Production, P.O. Box 1735, Tallahassee, FL 32302).

Project Supervised by:
Kathy McGuire, Policy Coordinator (904/487-9224)

Project Conducted by:
Louise Cobbe, Policy Analyst (904/487-9239)