Reconsidering the Project Greenlight Intervention: Why Thinking About Risk Matters

by James A. Wilson and Christine Zozula

Project Greenlight's negative outcomes disappointed stakeholders and puzzled researchers. A reexamination of Greenlight's data suggests that the intensity of the program may not have been well-suited for medium- and high-risk offenders.

The landscape of American corrections is littered with the bones of rehabilitative efforts that failed. This is certainly no surprise, given some of the novel efforts at rehabilitating criminal offenders, some of which, unfortunately, remain part of corrections even today.

In a 2008 seminar at the Institute for Excellence in Justice at the Ohio State University Criminal Justice Research Center, Ed Latessa of the University of Cincinnati reviewed several high profile programs that claimed to be rehabilitative.[1] These included such efforts as dance instruction for juveniles, drum circles for parolees, yoga for probationers, gardening, dog sledding and Handwriting Formation Therapy. To be clear, we have nothing against dance instruction, drum circles, yoga, gardening, dog sledding or handwriting, but their rehabilitative efficacy seems questionable. Although such programs are clearly not the norm, one has to wonder how well the concept of "evidence-based practice" has truly filtered down to inform correctional practice.

We have moved forward a great deal over the last decade in what we know about intervening with criminal offenders. The bulk of the research evidence clearly indicates that the programs most likely to produce robust results in reducing criminal recidivism have cognitive-behavioral foundations that target behaviors related to offending

and amenable to change, and that use social learning strategies.[2] In addition, the principles of correctional interventions suggest that programs should target medium- and high-risk offenders and that program implementation should be a key consideration for any new program.

However, even interventions with substantial empirical support do not always produce consistent results and may even be associated with negative outcomes.[3] One of the more prominent interventions recently linked to negative outcomes was Project Greenlight.[4] The original evaluation of Project Greenlight examined its effect on the recidivism rates of participants at 12 months compared with offenders who received standard prerelease programming and offenders who received no prerelease programming. It found that Greenlight participants had *higher* rates of arrests and parole revocations.

We reassessed Project Greenlight by analyzing data over a 30-month period. In this reanalysis, we specifically examined differences by the level of offender risk. Although the longer-term assessment confirmed the original evaluation findings, we also found that outcomes varied by risk level — low-risk offenders appeared to benefit most from the program, whereas medium- and high-risk offenders were harmed the most.

An Overview of Project Greenlight

The Greenlight program was developed and operated by the Vera Institute of Justice in conjunction with the New York State Department of Correctional Services and the Division of Parole. The program was built on the Reasoning and Rehabilitation (R & R) cognitive-behavioral program model. The

> ## The new results suggest a mismatch between the structure of the Greenlight program and the population to which it was delivered.

literature on correctional interventions shows that cognitive-behavioral approaches, such as R & R, are associated with reductions in recidivism rates. Cognitive-behavioral programs typically address attributes most related to criminal behavior and most amenable to change. These include such factors as impulsivity, maladaptive patterns of thinking, antisocial peers and attitudes, poor social skills, and drug use. In addition to the cognitive-behavioral foundation, the program also incorporated a number of other program elements with empirical or anecdotal support in reducing recidivism, including employment assistance, housing assistance, drug education and relapse prevention, development of a release plan, practical skills training, and release documentation that included identification and insurance coverage.

For the Greenlight intervention, the R & R program was modified in three important ways:

- The intervention period was shortened to eight weeks from four to six months.

- Class sizes were increased to 26 participants from the recommended eight to 10.

- Additional modules were incorporated, as outlined above.

As a result, the program can be considered more intensive than the standard formulation, and the compressed time frame and increased class sizes likely make it more diffi-

cult to deliver effectively. However, the restructured program's appeal should be obvious: more individuals can participate with the potential for sizeable reductions in cost.

The original assessment of Greenlight's effectiveness evaluated the combined rate of arrests and parole revocations 12 months after subjects were released from a correctional facility. In our reassessment, we looked at a longer follow-up period of 30 months, and reanalyzed the outcomes by the risk of the study participants. Principles of correctional intervention suggest that programming should be reserved for medium- and high-risk inmates, so it is plausible to think that the intervention might have differential effects by the risk level of the participants, with medium- or high-risk individuals showing some benefits.

Evaluation Design

The treatment group consisted of the 345 individuals transferred to the pilot facility and participating in the Greenlight intervention before release (GL). A second group of 278, who were also transferred to the pilot facility but assigned to the N.Y. Department of Corrections Transitional Services Program (TSP), constituted our primary control group. A third group met the criteria for participation, but these inmates were not transferred to the pilot facility due to space limitations. They were released directly from upstate facilities (UPS) and received no prerelease programming. The assignment process constitutes a relatively rigorous research design but has been described extensively elsewhere, so we do not discuss it here.[5]

Because both the GL and TSP groups were transferred to the pilot facility and had similar experiences with the

exception of the programming, we largely expected the intervention to account for any differences in outcomes. However, the UPS group deserves a short discussion because we can speculate that the effects could run in two different directions. To the degree that prerelease programming has net positive benefits, and UPS received no programming, we might expect the GL group (and to some degree, the TSP group) to do better. However, to the degree that the forced transfer and coerced participation in the program right before release might be disruptive and otherwise negatively experienced without achieving a therapeutic effect, we might expect the UPS group to do better than both GL and TSP.

Reconsidering the Evidence

The evaluation of Project Greenlight followed-up with inmates one year after they were released. At that time, the investigators' analysis found significant negative outcomes associated with the intervention — the GL participants had more arrests and parole revocations than either the TSP and UPS groups. In this reanalysis, we look at outcomes at 30 months and examine them by the participants' risk level. (See sidebar: Developing the Risk Instrument.)

Results by Risk Level

In Exhibit 1, we show the percentage of participants who were living in the community at 30 months and had not been rearrested.[6] Within each group, we examine the percentages by risk level. The data for the full sample, shown on the first row of the table, are consistent with the results of the one-year evaluation. Participants in the GL group had the highest recidivism rate, with less than half (47.5 percent with no rearrest) still in the community at 30 months. The difference of nearly 20 percentage points between it and

Developing the Risk Instrument

We developed our risk instrument from data on individual attributes that have been strongly associated with criminal recidivism.[1] Our data include numerous measures of criminal history such as current offense; numbers of misdemeanor and felony arrests and convictions; bench warrants; and database indicators for drugs, weapons and firearm offenses. Standard demographic data such as age, race/ethnicity and educational level, as well as some information on substance use were contained in the data files originally provided by the New York State Department of Correctional Services. We estimated a number of different models for constructing a risk scale. In doing so, we paid special attention to the literature on the predictors of offender recidivism, but we also tested all of the variables available to us and considered their potential meaning for respondent outcomes.

Following Gottfredson and Snyder, we used logistic regression to obtain unstandardized coefficients for variables that predicted new arrests.[2] Variables that were statistically associated included prior parole revocations, prior felony arrests, bench warrant indicators, substance use measures, release age and borough of release.[3] We included borough of release because it could potentially indicate opportunities and networks available to individuals recently released from prison. Given the lack of dynamic risk predictors (i.e., predictors amenable to change, such as antisocial attitudes and peer associations, substance use, poor social control/impulsivity, family environment, and education/employment), geographic location may be the next best thing because it suggests neighborhood characteristics such as employment opportunities, living arrangements and exposure to pro-social peers. Once our scale was constructed, we defined three risk levels for sample size reasons, but rather than simply dividing the scale into thirds, we selected the bottom 30 percent as "low risk," the top 30 percent as "high risk," and the middle 40 percent as "medium risk."[4]

Exhibit 1. Percent of Participants Without a Rearrest After 30 Months

Risk Level	Greenlight (N = 345)	TSP (N = 278)	UPS (N = 113)
Total Sample	47.5	51.8	66.4 **
Low-Risk	80.4	70.0 ^	86.4
Medium-Risk	44.0	51.7	69.0 **
High-Risk	23.7	33.8 *	32.1

Note: All comparisons of statistical significance are with the GL group.
^ p < .10; * p < .05; ** p < .01

Figure 1 is a very basic illustration of how well our risk instrument discriminates between those classified as low, medium and high risk. A 27 percentage point difference distinguishes the difference between low and medium risk, and a 22.3 percentage point difference exists between medium and high risk. The degree of discrimination across risk categories for any arrest is statistically significant. To our knowledge, no one has established specific criteria for what constitutes low-, medium- and high-risk offenders, but we believe our scale represents, to a reasonable degree, these conceptual categories.

Figure 1: Assessing the Risk Scale: Rearrest by Risk Level of Study Participants
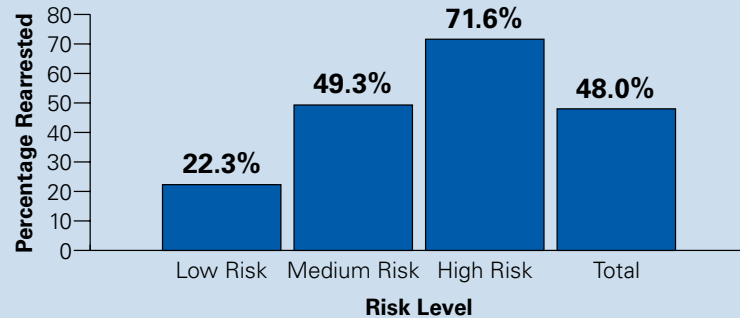
Note: Drawn from Figure 1b in Wilson and Zozula, "Risk, Recidivism and (Re)Habilitation: Another Look at Project Greenlight," *The Prison Journal* (forthcoming, 2012).

Notes

1. See, e.g., Andrews, D.A., Ivan Zinger, Robert D. Hoge, James Bonta, Paul Gendreau, and Francis T. Cullen, "Does Correctional Treatment Work? A Clinically Relevant and Psychologically Informed Meta-Analysis," *Criminology* 28 (1990): 369-404. Andrews, D.A., and James Bonta, *The Psychology of Criminal Conduct,* 4th Edition, Cincinnati, OH: Anderson Publishing (2006). Gendreau, Paul, Tracy Little, and Claire Goggin, "A Meta-Analysis of the Predictors of Adult Offender Recidivism: What Works!" *Criminology* 34 (1996): 575-608.

2. Gottfredson, Don M., and Howard M. Snyder, *The Mathematics of Risk Classification: Changing Data into Valid Instruments for Juvenile Courts,* OJJDP Report, Washington, D.C.: Office of Juvenile Justice and Delinquency Prevention, July 2005, NCJ 209158, available at https://www.ncjrs.gov/pdffiles1/ojjdp/209158.pdf.

3. Educational level and race/ethnicity were not included because they were not predictive in any of the models tested and because the use of race/ethnicity variables in such scales raises ethical concerns.

4. As one might expect, whether we divided our risk levels into thirds, quartiles or some other grouping made little difference. We ultimately decided on the 30-40-30 distribution in order to capture those who were at slightly lower and slightly higher risk, but in practical terms, other divisions did not yield different results.

the UPS group (66.4 percent with no rearrest) is statistically significant.

However, recidivism rates vary depending on risk level. The "risk principle" suggests that the most intensive programming should be reserved for medium- and high-risk offenders, but it is the low-risk offenders who appear to benefit most from the GL program. In contrast, high-risk TSP participants were more likely to avoid rearrest than high-risk GL participants.

Individuals in the UPS group were less likely to be arrested again

compared with either GL or TSP participants for every risk level except high, in which results for UPS and TSP participants were similar. Further, despite the lack of statistical significance (largely due to inadequate statistical power due to small sample sizes), most of the contrasts suggest reasonable reductions in recidivism. The 25 percentage point difference between medium-risk GL and UPS (44 percent to 69 percent) offenders is substantial. The question is, how do we explain these differences and what are the implications for correctional programming?

Making Sense of the Results

A number of explanations are possible for the results we present here. The most obvious is that the research design was flawed and that individuals who are more prone to crime were differentially assigned to each of the three groups — in short, the GL group has more high-risk participants than the TSP group, which has more than the UPS group. Although some differences in risk levels are evident, the strength of the research design and multivariate analyses with controls suggest that demographic and criminal history

differences don't account for the differences in recidivism rates. We also affirm that attrition is not at issue: All individuals assigned to the treatment group completed the mandatory GL programming and were followed for the full period after release.

In the initial evaluation, discussions about the negative effects associated with the GL program centered on program design and implementation. The new results suggest a mismatch between the structure of the GL program and the population to which it was delivered.

Several factors support this conclusion. First, speculation about poor program implementation was bolstered by evidence that certain GL case managers accounted for nearly all of the negative program effects reflected in the original one-year follow-up figures. Such differences would suggest problems with the delivery of the program. However, in the most recent assessment of the data, variation among case managers shows much smaller differences across the board.

In addition, if the program were poorly structured or poorly delivered, it seems reasonable to think that the negative effects of the program due to problems with implementation would apply to all risk levels. At the very least, one might expect the lowest risk individuals to be most negatively affected if the program had been poorly structured or poorly delivered. However, in this case, the lowest risk individuals don't exhibit the same negative effects as the medium- and high-risk offenders when the comparison is between GL and TSP.

So what can explain our findings? We would argue that the 30-month findings show low-risk individuals are the most amenable to the inten-

sity of the Greenlight intervention. By definition, low-risk individuals are likely to be less impulsive, have better attention spans, better cognitive skills, better social skills and better verbal ability — in short, they are more likely to have the skills that serve one best in a classroom environment. Thus, it seems reasonable to think they would be better situated to process the more intensive and more compressed intervention that Project Greenlight provided.

Why would the medium- and high-risk individuals do so much more poorly with Greenlight? Perhaps, just as low-risk individuals possess the attributes that make them more suited for such intensive and compressed programming, medium- and high-risk individuals are more likely to possess traits that make them less suitable. The risk principle holds that the most intensive programming should be reserved for those who are at medium-to-high risk. However, treatment programs should be delivered in a style and mode consistent with the ability and learning style of the offender (the responsivity principle). As we have already noted, the GL intervention might be considered "very" intensive given its compressed delivery time, increased class sizes and additional program elements. This intensive programming, however, may not have been clinically appropriate. With high-risk offenders, programming can initially engender more resistance, creating anger, resentment and frustration at being forced to participate.[7] Wilson and Davis[8] noted that "if the intervention is not of sufficient length for a therapeutic effect to be realized, offenders may be released directly to the community still suffering the ill effects of coerced programming" rather than its intended therapeutic effects. In other words, the program might just be too short for intervening with high-risk offenders.

The other major question is why the UPS group, released directly from prison with no prerelease programming whatsoever, did so well, compared not only with the TSP group, but also to the GL group. For lack of a more plausible explanation at this point, one must consider the possibility that transferring individuals right at the end of their incarceration and coerced programming might be detrimental to their well-being. To the degree that inmates form social bonds and networks, are embedded within a specific community and a stable institutional life, and have some semblance of control over their lives, an involuntary transfer to another facility, with coerced programming to follow, may be disruptive and counterproductive. A diverse literature suggests that situations and events that create stress, especially those that generate a sense of powerlessness such as involuntary moves, can negatively impact a host of life outcomes, including recidivism. GL program designers assumed that transferring individuals to an institution in their home community right before release would help them in the prerelease planning process, especially in connecting participants to community-based service providers. Our data suggest that prison transfers or coerced programming just before release, or some combination of the two, might be counterproductive in significant ways.

Lessons for the Future

We believe the patterns of success between the three different groups across the different risk levels suggest important considerations for correctional program developers. It seems clear in hindsight that the GL developers failed to consider several important principles of effective correctional programming despite drawing from that literature.

One of the most important failures was to ignore participants' risk levels. Despite the notion that the most intensive interventions should be reserved for medium- and high-risk individuals, a notion that is intuitively and theoretically sound, our analysis suggests that some intensive interventions, especially those that are compressed into a very short time frame, may not be suitable for such offenders. They simply may not be capable of processing large amounts of material in such a compressed period of time. The structure of the GL program seems to have been much more suitable for the abilities of those at lowest risk. The positive performance of the low-risk group also suggests that such condensed programming has potential for rehabilitative efforts with such individuals.

We also note that our findings may not be too disparate from other segments of the literature regarding correctional interventions. At least one meta-analytic review reports that results from evaluations of the R & R program show positive effects for both low- and high-risk offenders, with slightly stronger effects for low-risk offenders, although differences between the two groups are not statistically significant.[9] In this case, the more condensed "intensive" program might still have yielded positive effects for low-risk inmates, but exceeded the tipping point for what is suitable for medium- and high-risk individuals.

Our analysis also raises questions about the wisdom of forced transfers and coerced programming

immediately before release. Despite the potential benefits of connecting offenders to local service providers, disrupting social networks and existing routines, and creating or heightening any number of negative emotional states may be counterproductive, especially if sufficient time isn't allotted to counteract the more negative effects. At the very least, we think this explanation for the worse outcomes of the GL and TSP groups compared with the group that was not transferred is plausible and that these issues warrant a harder look.

About the authors: James A. Wilson is the senior program officer at the Russell Sage Foundation. Christine Zozula is a graduate student of sociology at the University of Connecticut.

NCJ 235890

Notes

1. Latessa, Edward J., "Improving the Effectiveness of Correctional Programs Through Research," *Institute for Excellence in Justice Seminar* (2008). http://www.drc.state.oh.us/web/iej_seminars.htm (accessed August 4, 2011).

2. Andrews, Don A., "Enhancing Adherence to Risk-Need-Responsivity: Making Quality a Matter of Policy," *Criminology and Public Policy* 5 (2006): 595-602. Cullen, Francis T., and Paul Gendreau, "Assessing Correctional Rehabilitation: Policy, Practice, and Prospects," in *Policies, Processes, and Decisions of the Criminal Justice System,* vol. 3, ed. J. Horney, Washington, DC: U.S. Department of Justice, 2000, NCJ 185530.

3. Barnoski, Robert, *Outcome Evaluation of Washington State's Research-Based Programs for Juvenile Offenders,* Olympia, WA: Washington State Institute for Public Policy, 2004. http://www.wsipp.wa.gov/pub.asp?docid=04-01-1201 (accessed August 8, 2011). Lowenkamp, Christopher T., Edward J. Latessa, and Paula Smith, "Does Correctional Program Quality Really Matter? The Impact of Adhering to the Principles of Effective Intervention," *Criminology and Public Policy* 5 (2006): 575-594.

4. Ritter, Nancy, "No Shortcuts to Successful Reentry: The Failings of Project Greenlight," *Corrections Today* 94 (December 2006): 94-97. Wilson, James A., "Habilitation or Harm: Project Greenlight and the Potential Consequences of Correctional

Programming," *NIJ Journal* 257 (2006): 2-7. Wilson, James A., and Robert C. Davis, "Good Intentions Meet Hard Realities: An Evaluation of the Project Greenlight Reentry Program," *Criminology and Public Policy* 5 (2006): 303-338.

5. Brown, Brenner, Robin Campbell, James A. Wilson, Yury Cheryachukin, Robert C. Davis, Jean Dauphinee, Robert Hope, and Kajal Gehi, *Smoothing the Path From Prison to Home,* Final report to the National Institute of Justice, grant number 2002-RT-BX-1001, April 2006, NCJ 213714, available at https://www.ncjrs.gov/pdffiles1/nij/grants/213714.pdf.

6. Simple percentages like this may be complicated by the differences in time spent at risk in the community. If one group has more time at risk, it might have higher percentages of rearrest. All participants

had at least 30 months at risk and we censored all cases at 30 months. In doing so, we essentially controlled for differences in time at risk in the community that might account for differences in rearrests.

7. Porporino, Frank, and Elizabeth Fabiano, *Reasoning and Rehabilitation Revised: Theory and Application,* Ottawa, Ontario, Canada: T3 Associates (2000).

8. Wilson and Davis, "Good Intentions Meet Hard Realities: An Evaluation of the Project Greenlight Reentry Program," *Criminology and Public Policy* 5 (2006): 303-338.

9. Tong, L. S. Joy, and David P. Farrington, "How Effective Is the 'Reasoning and Rehabilitation' Programme in Reducing Reoffending? A Meta-Analysis of Evaluations in Four Countries," *Psychology, Crime and Law* 12 (2006): 3-24.

Learn more about Project Greenlight and its evaluation at http://www.nij.gov/journals/257/habilitation-or-harm.html.

Listen to Edward Latessa talk about evidence-based knowledge in corrections: http://www.nij.gov/multimedia/presenter/presenter-latessa.

Read about cognitive-behavioral therapy at http://www.nij.gov/journals/265/therapy.htm.