

©2008

Avi-Yonah M. Schwab

ALL RIGHTS RESERVED

DOES GRADUATION FROM A TOKEN ECONOMY
PREDICT LONG-TERM OUTCOMES
OF A RESIDENTIAL TREATMENT PROGRAM?

By

AVI-YONAH M. SCHWAB

A Dissertation submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Psychology

written under the direction of

Prof. Maurice J. Elias

and approved by

New Brunswick, New Jersey

May, 2008

ABSTRACT OF THE DISSERTATION

Does Graduation from a Token Economy Predict
Long-Term Outcomes of a Residential Treatment Program?

By AVI-YONAH M. SCHWAB

Dissertation Director:
Prof. Maurice J. Elias

Token economies are efficacious and widely-used treatments for disruptive adolescents in residential settings. However, the literature suggests that positive effects of these treatments may not maintain or generalize when artificial reinforcement is no longer available. Theory and research from both social and behavioral psychology suggest that generalization could be enhanced if acute behavior change in token economies is followed by a phase that emphasizes self-regulation training and natural and logical consequences over token reinforcement. At Girls and Boys Town's family-style residential program for youth with multiple behavioral and environmental problems, youth who succeed in the program progress through motivation system levels, from traditional token economy levels to a motivation system that emphasizes self-regulation and natural and logical consequences. It was hypothesized that youth who ended their treatment at higher motivation system levels would have better outcomes at 5-16-year follow-up, even when controlling for a proxy of general success in the treatment program (Favorable Departure Rating). One cohort of former residents was surveyed 16 years, on average, after departure ($n = 207$; 42% of the total cohort), and another cohort of former

residents was surveyed 5 years after departure ($n = 131$; 61% of the total cohort). Data on respondents were also collected from treatment records. Analyses of the 16-year follow-up group suggested that discharge from higher motivation systems positively predicted educational attainment, household income, and military service and marginally predicted locus of control and happiness. However, when controlling for Favorable Departure, motivation system marginally predicted only two outcomes. Favorable Departure was a more reliable predictor of happiness, highest grade completed, and (marginally) respondents' retroactive evaluation of the treatment at follow-up than motivation system was. Neither measure significantly predicted outcome on 13 other long-term outcome measures. In the 5-year follow-up group, motivation system was again a strong predictor of military service, even when controlling for Favorable Departure. However, neither motivation system nor Favorable Departure was a significant predictor of any of the other 16 long-term outcome measures. Exploratory analyses of predictors of post-treatment success and long-term follow-up outcomes of residential treatment were also conducted. Extensive recommendations for follow-up research are offered.

Acknowledgements

I gratefully acknowledge my mentor, Maurice J. Elias, Ph.D., for his ceaseless dedication and superlative guidance throughout this project and, indeed, throughout my graduate career. A deep thank you is owed to Ronald W. Thompson, Ph.D., who generously invited me into the Girls and Boys Town family, offered me the use of two precious and unique datasets, and allocated many hours of his time and his staff's time to bring this project to fruition. Thank you to the other committee members, Sandra L. Harris, Ph.D., and Brian C. Chu, Ph.D., for their thoughtful comments and suggestions. I am very grateful to the research staff at Girls and Boys Town, including Jonathan C. Huefner, Ph.D., Jay L. Ringle, M. Beth Chmelka, and Connie Spath, for their hospitality, patience, and time. I am especially thankful to the youth and staff of Girls and Boys Town for welcoming me and teaching me about their special community. I owe a debt of gratitude to my parents, Aviva and Joel Schwab, and my in-laws, Linda and Larry Rezak, for their support in so many different ways. Thank you to my children, Hillel, Sarit, and Ezra, for bring added meaning to my life and occasionally allowing me to work on my dissertation. Finally, thank you to Jessica Rezak Schwab, my partner in life. Without her limitless support and love, I could not have done this work or reached this milestone.

Table of Contents

Abstract	ii
Acknowledgments	iv
Table of Contents	v
Introduction	1
Method	25
Results	51
Discussion	74
References	93
Tables	102
Curriculum Vita	167

Introduction

The most widely studied and empirically supported interventions for youth with conduct and disruptive behavior problems are behavioral treatments (Brestan & Eyberg, 1998; Farmer, Compton, Burns, & Robertson, 2002). These treatments rely heavily on operant conditioning, specifically positive reinforcement for compliant behavior and punishment for rule-breaking and noncompliance. The most intense form of behavioral treatment for disruptive behavior is a token economy. Token economies are system-wide contingent reinforcement programs that are used frequently in residential and inpatient treatment of youth with behavior problems. Research has shown that token economies can improve behavior significantly, at least within the confines of the treatment context, for many different populations (see Dickerson, Tenhula, & Green-Paden, 2005; Hersen, 1976; Kazdin, 1977, 1982; Kazdin & Bootzin, 1972; Sullivan & O'Leary, 1990), including disruptive youth (e.g., Foxx, 1998; Hobbs & Holt, 1976; Kirigin, Braukmann, Atwater, & Wolf, 1982; Wolfe, Dattilo, & Gast, 2003).

The outcomes of behavioral treatments in general and token economies in particular are robust but not universal. Typically, 20-40% of children treated with these approaches drop out of treatment, and 30-50% of completers remain in the clinical range at the end of treatment (Greene, Ablon, & Goring, 2003; Kazdin, Siegel, & Bass, 1992; Schuhmann, Foote, Eyberg, Boggs, & Algina, 1998). Though follow-up studies have not been conducted with great frequency, anywhere from a quarter to more than half of completers are typically reported to be in the clinical range at follow-up (Kazdin et al., 1992; Reid, Webster-Stratton, & Hammond, 2003). Maintenance and generalization are major challenges of operant-oriented approaches. The difficulty in maintaining behavior

improvements outside of the treatment context can undermine the long-term effectiveness of behavioral methods.

Token economies, in particular, have a history of limited success with generalization (e.g., Kirigin et al., 1982). The main problem is stimulus generalization. Stimulus generalization is the transfer of learned responses to new stimulus conditions. In the context of token economies for disruptive children, stimulus generalization refers to the child's behaving well when the token economy is not in force. Response generalization, which is the increased use of desirable behaviors that were not specifically targeted by the token economy, is also a concern as it would be pragmatically impossible to target every possible desirable behavior. In both cases, the behaviorist literature is clear. One cannot expect that generalization will occur on its own. Stokes and Baer call this "train and hope" (1977, p. 351), and there is an extensive literature of little generalization and only very short maintenance of behavior after a token economy ceases to operate. Generalization must be programmed explicitly, or one can assume that the end of the token economy is an extinction operation (Kazdin, 1977; Kazdin & Bootzin, 1972; Kendall, 1989; Levine & Fasnacht, 1974; Stokes & Baer, 1977). Despite the fact that the challenges of generalization were first articulated over 35 years ago, there is surprisingly little research that systematically examines procedures for promoting generalization in the token economy literature.

One might think that standard operant procedures – rewards and punishments – would promote generalization on their own. If a behavior becomes associated with a salient reward, operant conditioning theory tells us that it will increase. Likewise, if it is associated with punishment, it will decrease. Classical conditioning suggests that, even

when the rewards and punishments are no longer available, the positive associations between a behavior and a reward and the negative associations between a behavior and a punishment will persist, at least for a time. This should make the behavior self-reinforcing. However, the general lack of spontaneous generalization suggests otherwise. Two, mostly compatible, models may explain why classical conditioning often does not promote generalization in this way. The first is behavioral learning theory itself. The other is Self-Determination Theory (SDT; Deci & Ryan, 2000) and findings on the undermining effects of rewards on internal motivation (Deci, Koestner, & Ryan, 1999). This paper will not attempt to differentiate between these accounts or support one over the other; rather, it will attempt to integrate them to enhance our understanding of the phenomena.

According to SDT, external reinforcement influences two competing needs in the child, the need for competence and the need for autonomy. The need for competence is satisfied when the child receives information about the efficacy of his or her behavior. According to this theory, external reinforcement can increase the frequency of a target behavior when a contingency is in place partly because it is positive feedback; it supplies useful information, serving the child's need for competence. However, that is not the only effect of external reinforcement. Once the operant contingency is no longer in place (e.g., rewards are no longer available to reinforce the target behavior), research has demonstrated that the rate of the target behavior generally drops *below* the baseline level if the baseline level was above zero (Deci et al., 1999). In behavioral terms, this is a "behavioral contrast effect." SDT posits that a contrast effect occurs (or is enhanced) when the contingency is experienced as controlling, thereby impinging on the child's

perception of autonomy. In other words, the child may comply under controlling circumstances, as long as the behavior is externally reinforced. However, he or she is unlikely to continue this behavior when external reinforcement is unavailable. In behavioral terms, perhaps the “controllingness” of the contingency is aversive. Perhaps behaviors that are externally controlled acquire an *aversive* association through classical conditioning, rather than a rewarding association, thereby punishing future behavior when the reward and its desirable qualities are no longer available.

The first study of this phenomenon in children was reported by Lepper, Greene, and Nisbett (1973). Preschoolers in the “expected-reward” condition were told in advance that they would receive a reward contingent on drawing with magic markers, and, after drawing, they received a reward. Children in the “unexpected-reward” condition were not told in advance that they would receive a reward (but they did), and the children in the “no-reward” group were neither told they would receive a reward nor did they receive one. After a single session of drawing in which these different contingencies were in place, the magic markers were made accessible in the classroom during free play for a number of days, and the rate of playing with magic markers was measured. During this free-choice period, children in the expected-reward condition played with the markers significantly *less* than children in the other conditions. SDT suggests that this occurred because establishing a reward contingency was perceived as controlling and, therefore, reduced internal motivation for magic marker play relative to the no-reward control condition. The unexpected reward did not cause a drop in intrinsic motivation, it is argued, because there was no *a priori* contingency, so the children in this group did not experience the activity as controlling. They attributed their magic marker

drawing to internal, rather than external, motivation. A behavioral account of these findings might argue, similarly, that the expected-reward group's use of magic markers was reduced relative to the no-reward group because of a contrast with the previous reinforcement schedule. It is more difficult to explain the unexpected-reward group within a behavioral framework rather than from an attributional perspective, but there are a couple of behavioral possibilities. First, the unexpected-reward group may have had a different outcome because they were operating on something akin to an intermittent schedule of reinforcement; they did not expect a reward the first time and did not know when to expect one again. An alternative explanation is that the unexpected-reward group drew with magic markers in the initial session without a contingency in place, so, unlike the expected-reward group, the behavior was not associated with negative feeling of being controlled.

A direct analogy can be made between the expected rewards in this study and the use of reinforcement in treatment. For example, a behavioral contingency may be established in which a child is rewarded for sharing. If the child has a base rate above zero for sharing before the contingency goes into effect, then it is somewhat internally regulated to begin with. Effectively reinforcing sharing will increase the rate of sharing while the contingency is in effect (e.g., when adults are monitoring the behavior and reinforcing it, sharing should increase). However, SDT suggests that this focus on external motivation will reduce internal motivation. Therefore, when the contingency is removed – or when adults are not present to monitor and reinforce – the rate of sharing will likely drop *below* the base rate. This is precisely what happened in a well-designed classroom-based study of a token economy (Greene, Sternberg, & Lepper, 1976). When

children were reinforced for participating in specific educational activities, their rate of engagement with those activities increased significantly. During a 13-day withdrawal phase, however, children who showed high initial interest in the activities and children who chose the activities they would be rewarded for engaged with their target activities at below-baseline levels. Children who showed low initial interest returned to their baseline levels, which was significantly lower than the engagement time of a non-reinforced control group. This suggests that token reinforcement can undermine long-term performance on activities that are both high and low in initial interest and even on activities that are chosen by the children for reinforcement. This raises the question of what a token economy truly accomplishes, unless a child can be continually monitored and reinforced. Because generalization and maintenance are so difficult in token economies, some have suggested that token economies should only be used to shape previously nonexistent behaviors rather than serve as a general motivation and treatment system to motivate all desired behaviors (Hersen, 1976; Levine & Fasnacht, 1974).

A continuing issue in the field of behavioral programming has been the lack of attention given to the mechanisms for maintenance and generalization, and, especially, differentiation between many of the conditions under which maintenance and generalization are more or less likely to occur. Unfortunately, some of the most common approaches to programming for maintenance may not be optimal for long-term outcomes. One common approach to maintenance is to taper reinforcement so that target behaviors are reinforced on an intermittent schedule that slowly decreases in rate. A problem with this approach is that it does not take context into account. Though this may work within the treatment context (i.e., when the child is being monitored and knows that he or she is

on an intermittent schedule), it is unlikely to be effective when the child leaves the treatment facility or is not being monitored by an adult, unless the behavior is somehow naturally reinforcing. In behavioral terms, this is simple discrimination learning. Another commonly cited and used generalization technique in research and practice is what Stokes and Baer call “sequential modification” (1977, p. 352). Sequential modification transfers the behavior modification procedures, or similar ones, to each context in which generalization is desired. While this is likely an effective strategy to modify behavior in new contexts, it does not meet the definition of stimulus generalization because the most salient stimuli, those which are associated with the operant procedures, are now present in the “generalization” context. Generalization in this paper refers to transferring behavior to contexts in which artificial reinforcement, reinforcement that is part of an intentional system of behavior modification but is otherwise unrelated to the behavior, is not available. It is the generalization of behavior to situations beyond the reach of any token economy or behavior modification system and when no reinforcing agents (that is, therapists, teachers, parents, or other adults who dispense reinforcement in a behavior modification system) are present or could reinforce retroactively. In other words, generalization here refers to behaving well without an artificial system, with only the normal contingencies in the environment serving as the motivation. This is what disruptive children need if they are to ever live responsible, independent lives.

If behavioral procedures alone cannot account for the success of behavioral treatment, how, then, are some target behaviors in some children maintained long after behavioral treatment? I will offer two conjectures. First, SDT helps to deepen our

understanding of what might enhance or inhibit maintenance and generalization. SDT predicts that if feedback provides useful information about a behavior without being perceived as controlling by the child, it is more likely to foster internalization of the behavior. Internal regulation, in turn, is more likely than external regulation to lead to lasting change in the intended direction. After 35 years of empirical study of rewards and internal motivation, a number of conditions related to rewards have been identified that either increase or decrease the perception of “controllingness” and, therefore, internal motivation (Deci et al, 1999). As demonstrated by Lepper et al. (1973), when rewards are expected, they are more likely to reduce internalization than if they are unexpected. Tangible rewards are more likely than verbal rewards (i.e., praise) to undermine internal regulation. Rewards that are contingent on simple engagement or completion of an activity – as opposed to unusually good performance on the activity – are more likely to reduce internal motivation. Finally, rewards that are delivered in a controlling manner are more likely to undermine internalization than those that are delivered neutrally or support the child’s need for autonomy (Deci et al., 1999). To the degree that target behaviors are reinforced by unexpected and/or verbal rewards, that rewards are given for excelling at rather than just doing a task, and that the contingencies are not perceived as controlling or capricious, motivation is more likely to be internalized and the behaviors maintained over time and generalized to different contexts. The types of reinforcement that are less likely to undermine internal regulation probably occur at least some of the time in therapeutic contexts, even though behavioral programs do not systematically manipulate these variables. Addressing these variables directly, it would seem, could

enhance the long-term effectiveness of reinforcement without sacrificing its short-term efficacy.

The second explanation for how behavioral approaches sometimes lead to maintenance and generalization is that changes in most target behaviors are also maintained by the natural and social worlds. Target behaviors are initially changed through “intentional” reinforcement paradigms, such as rewards and punishments administered by adults, but they are also reinforced by physical and social contingencies that are not designed or carried out intentionally by therapeutic agents (Stokes & Baer, 1977). Returning to the sharing example, a child may be reinforced for sharing with some tangible reward. However, sharing also leads to social reinforcement in the form of peers being more likely to play with the child. Social consequences similarly rely on the social order. If a child insults a friend, and the friend stops playing with her, she has received a social consequence. Natural reinforcement and consequences are analogous except that, instead of the social order, they depend on the physical world; they occur without any human intervention. For example, if a child prepares a sandwich for himself, he is rewarded with the food. This is natural reinforcement. Conversely, if a child plays roughly with a toy, and the toy breaks, she no longer has use of the toy. This is a natural consequence. Natural and social reinforcers have the advantage of continuing to reinforce behavior well beyond the treatment context – when no specifically designated agent is monitoring and reinforcing. These natural and social contingencies are not always salient enough to modify behavior on their own (at least in the short-term), but they can support maintenance and generalization as long as the intentional reinforcers are

not so salient as to overshadow and block the reinforcement power of the natural or social reinforcer.

When behaviorists review the literature and make general recommendations about programming generalization, they invariably put natural reinforcement at the top of the list of strategies. In Stokes and Baer's (1977) seminal article on generalization, they write,

Perhaps the most dependable of all generalization programming mechanisms is one that hardly deserves the name: the transfer of behavioral control from the teacher-experimenter to stable, natural contingencies that can be trusted to operate in the environment to which the subject will return, or already occupies. (p. 353)

Out of Stokes and Osnes's (1989) twelve tactics to facilitate generalization, the first three are contact natural consequences, recruit natural consequences, and modify natural consequences. They write that "perhaps the most fundamental guideline of behavior programming, as well as generalization programming, is to teach behaviors that are likely to come into contact with powerful reinforcing consequences that do not need to be programmed by a therapist or behavior change agent" (p. 341). In his review of a number of articles on generalization, Kendall (1989) concludes that "teaching skills that will be maintained *naturally* was a reappearing suggestion that can help guide practitioner's selection of skills for training" (p. 360, emphasis his).

While it is ideal to simply put the individual in contact with existing natural reinforcement and consequences in the environment that need no programming, these are often not salient enough or reliable enough on their own to change or maintain behavior (Kazdin, 1977). Therefore, it may be necessary to modify the natural environment to

bring it into control of the target behavior. This is called using “logical consequences” (Dreikurs, & Loren, 1968) or logical reinforcement. Logical consequences are similar to natural and social reinforcement, but they require some agentic manipulation. The goal is to mimic the natural or social consequences of a behavior as closely as possible. The logical consequence, therefore, must be logically related to the behavior it is supposed to modify. Horner (1971) described an excellent example of this type of manipulation in a therapeutic setting. A 5-year-old institutionalized retarded boy was taught to walk with crutches. Rather than using token reinforcement to reward the generalization of this behavior beyond the experimental setting, caretakers who had previously pushed his wheelchair around were simply prompted to withhold that help. The reinforcement for walking with crutches was a “logical” one, to be able to participate in all of the activities he walked to. Within 15 days, he had reached the ceiling of 12 walking trips per day. Spot checks over the next 35 days showed that he performed 100% of his walking trips every day. “Generalization programming seems well served by providing the least artificial, least cumbersome, and most natural positive consequences in programming interventions. Such programming most closely matches naturally occurring consequences and their entrapment potential” (Stokes & Osnes, 1989, p. 341).

Behavioral treatment programs focus on the saliency and availability of reinforcers but largely ignore, or do not systematically address, the distinctions described here between reinforcers that foster or inhibit internalization as well as the differences between artificial reinforcers that require adult monitoring and natural and social reinforcers that are always present in the natural environment; logical reinforcers are rarely mentioned. The reinforcement used in existing behavioral treatment often falls

into the more effective categories, but this is usually serendipitous. Because selecting reinforcers based on their generalization potential is not yet an important concern in behavioral treatment, some of the techniques that are typically used because they are so effective and easy to administer in the short-term may disrupt long-term maintenance and generalization. While token economies and other behavioral treatments for disruptive behavior are designed to give accurate feedback to the child about his or her behavior (supporting the need for competence), they also employ tactics that can be perceived as controlling (impinging on the need for autonomy). Furthermore, they are useful for acute behavior change (such as teaching a new behavior or modifying a very dangerous behavior quickly) when natural and social reinforcement are not salient enough, but, if used too long, they can block the natural reinforcement that would help to maintain and generalize the behavior throughout life. The quandary facing mental health professionals is how to change noncompliant behavior rapidly in the short-term without preventing internalization and, therefore, undermining maintenance and generalization.

The generalization literature makes a number of suggestions. Sequential modification and intermittent reinforcement have already been discussed. Delayed and thinning schedules of reinforcement have similar limitations. Other suggestions include training with sufficient stimulus exemplars and response exemplars. There is some evidence that one need not train in *every* context; training in a few (perhaps as few as two) may be sufficient for generalization in some cases. Training in diverse circumstances (training “loosely”) is a similar tactic. Also suggested is incorporating functional mediators, such as physical and social stimuli that cue the behavior in different environments (Kazdin & Bootzin, 1972; Stokes & Baer, 1977; Stokes & Osnes, 1989).

These all appear to be useful modification procedures; however, all of these suggestions still rely on the motivational power of artificial reinforcement in contexts in which that reinforcement is either not available or only occasionally available. “Intermittent schedules may only delay extinction, rather than prevent it” (Kazdin & Bootzin, 1972, p. 363). This likely applies to all of these techniques because, eventually, if the child is not reinforced somehow, that is, if the child does not find the behavior valuable, relying on past artificial reinforcement for motivation will probably not suffice.

There are two ways in which these behaviors can be valued; the child can learn to enjoy the behavior intrinsically, or the child can receive some benefit from the natural environment for doing it. If the behavior is intrinsically motivating, it becomes its own reward. This is obviously not the case for many target behaviors, or generalization would not be a problem. Furthermore, external reinforcement has been shown to reduce, not enhance, intrinsic motivation. The generalization literature suggests an alternative to external reinforcement, however, that can enhance internal regulation (if not intrinsic motivation). Self-control and self-management procedures, when they are designed carefully, promote autonomy while maintaining a high level of informative feedback. According to SDT, this should make them more effective in promoting long-term behavior change because they should facilitate rather than inhibit internalization. Self-mediated stimuli, such as self-monitoring, self-talk, and self-reinforcement, all fall into this category. There have been a number of successful attempts to integrate self-management into token programs.

Turkewitz, O'Leary, and Ironsmith (1975) modified the after-school classroom behavior of eight disruptive children using a traditional token economy. The children

were then taught to set goals and self-evaluate, and tokens were used to reinforce accurate self-evaluation. Over time, the requirement to match one's self-evaluation with the teacher's evaluation was faded, and then the availability of backup reinforcement was also faded. Over the 5-day period at the end of the study in which no back reinforcement was available, disruptive behavior levels were maintained below baseline and only slightly above the level achieved when self-evaluation was reinforced. Generalization was observed to the times during the class in which the token economy did not operate. However, these behavioral improvements did not generalize to the children's public school classrooms.

Neilans and Israel (1981) also applied self-management to an after-school class environment. In this case, it was in a residential treatment group home for disruptive children that employed a token economy. After baseline and token economy phases, self-charting, self-goal-setting, self-evaluation, and self-reinforcement were sequentially taught. After a phase of being reinforced for accurate self-evaluations, the children were allowed to self-reinforce. During the self-regulation phase, disruptive behavior decreased below the token economy phase, and it maintained at this low level during 11 post-treatment days. Changes in on-task behavior were parallel. Interestingly, while behavior during the math portion of the class (which was targeted) improved during the token economy phase, behavior during reading on those same days deteriorated, suggesting a contrast effect. There was no contrast effect during the self-regulation system. Neilans and Israel concluded that "the present design suggests caution and further investigation as to whether an other-controlled system is a necessary precursor to self-regulation training" (p. 193).

Using a similar multiple baseline design, Wood and Flynn (1978) compared the effects of a typical token system to a self-evaluation token system on the room-cleaning behaviors of six pre-delinquent youth in family-style residential treatment. In the self-evaluation condition, youth were rewarded for accurate self-evaluations as part of their training. External token reinforcement was effective in increasing room-cleaning behavior, but the behaviors rapidly returned to baseline when tokens were not available. However, all six youth maintained their high levels of performance after being trained to self-evaluate. Maintenance was measured in the treatment home for as long as 60 days (the longest baseline period in the study) after tokens were no longer available. Though a number of mechanisms may be at work, these three studies seem to support SDT's prediction that autonomy-enhancing procedures facilitate generalization within token economies.

Ultimately, self-management is necessary but not sufficient for long-term maintenance. The ability to self-regulate makes the behavior possible outside of the treatment context, and it gives the child self-efficacy in that domain, which may make it intrinsically reinforcing for a while; however, it may not motivate the child to perform a behavior that is not inherently interesting or useful. Motivation comes from the child's finding that the behavior has tangible benefits, that is, if it results in reinforcement from the natural or social worlds. Ultimately, that is how long-term maintenance of behavior change occurs far from the treatment context. The child must learn to self-regulate with regards to the behavior *and* be motivated to use those skills by the natural contingencies in the environment. Writing specifically about token economies, Katzin and Bootzin (1972) argue that, "although there are a number of procedures for potentially increasing

generalization, it is our guess that the most fruitful techniques will be the ones that emphasize programming the natural environment” (p. 364).

The question remains as to how to change behavior quickly without undermining long-term maintenance and generalization. Are self-management skills and natural reinforcement enough? What if the child’s behaviors are very destructive and resistant to change with these techniques? One potential solution is to employ a traditional token economy at the beginning of treatment to change behavior as quickly as possible. Then, once the child has demonstrated an ability to behave properly with the help of external contingencies, shift to self-regulation training that is backed up by less controlling and more natural reinforcement contingencies. Such contingencies would differ from the first phase in that they would not take the form of tangible rewards or punishments that are artificially connected to the behavior and determined solely by the token economy system or by the adults administering it. Contingencies in the second phase would take the form of natural and logical consequences, which rely, as much as possible, on the “real-world” contingencies in the environment. It is hypothesized that children experience these types of consequences as less controlling than traditional contingencies because they are not arbitrarily chosen but rather endogenous to (that is, a natural result of) the behavior in question. They are also hypothesized to promote generalization because they are similar to or the same as the real-world contingencies in effect outside of the treatment context.

Residential Treatment

It is both pragmatic and theoretically interesting to test this theory – that natural and logical consequences and self-regulation will enhance the maintenance and generalization of behavioral gains – in the residential treatment modality. Residential

treatment programs serve approximately a quarter of a million children and adolescents at any one time (Leichtman, 2006). Though their programs and cultures are quite heterogeneous, Leichtman (2006) argues that they have had three features in common since the modality was first defined in the 1950's. First, they all provide psychotherapeutic interventions. Originally, this meant psychoanalytic therapy, but it now includes behavioral treatments, family therapy, group therapy, pharmacotherapy, and so on. Second, residential treatment programs focus not just on the time residents spend in "therapy" but on the entire experience of living in the residence. The term "milieu therapy" is amorphous, and its meaning varies greatly from program to program, but it attempts to capture the idea that the entire experience of residential treatment – 24 hours a day – is supposed to be therapeutic. Programs emphasize helping residents to learn to negotiate all of life's daily activities adaptively. Third, residential treatment programs are intended to address critical events as they happen. While it might be convenient for children to schedule their crises for therapy hours, this rarely happens. Therefore, all members of the staff must learn to intervene therapeutically whenever problems arise and to provide a consistent, unified approach at all times. These last two pillars are unique to residential treatment (Leichtman, 2006).

For many, perhaps most, residential programs, a structured behavioral reinforcement system (often a token economy) is part of the treatment "milieu" and is administered by all staff at all times. It is often supposed to frame, or partly frame, the systematic responses to crises and problems. While these tenets are clearly a strength of residential treatment, if the above-noted limitations to such behavioral systems are accurate, this may account for the challenges that many such programs find in helping

residents to maintain and generalize their gains after discharge (Kirigin et al., 1982; Leichtman, 2006). The stepped-care model described above provides a theoretical framework for systematically improving behavioral treatments that could have important implications for the long-term mental health of the hundreds of thousands of children and adolescents who are treated in residential settings. A residential facility is also an ideal place to study this model because of the clinical relevance of the modality and the unique ability to control the children's total environments in such a setting.

Girls and Boys Town

While token economies are empirically supported and widely used in residential treatment, this stepped-care approach, in which a token economy is followed by a phase of self-regulation and natural and logical consequences to promote internalization and generalization, is not currently common practice, nor has it been tested systematically. However, it is the model of Girls and Boys Town's long-term adolescent residential treatment program. Girls and Boys Town (GBT) was one of the first treatment programs of its kind to employ a token economy, beginning in the 1970s. In that program, adolescents were reinforced on a complex motivational point system with a daily reinforcement schedule. Despite the token economy's effectiveness while youth were at GBT, behavior change was not generalizing well when youth left the program. Generalization is particularly important at GBT because many youth stay until graduation from high school. When they leave, they are emancipated adults. Their ability to behave prosocially and responsibly without constant external monitoring is key to their success in living independently (R. W. Thompson, personal communication, July 2006).

In an attempt to improve generalization, GBT instituted a stepped-care approach similar to the one described here. Youth who are successful in the daily reinforcement schedule “graduate” to a weekly reinforcement schedule. After demonstrating success with the weekly reinforcement schedule, teens graduate to an “achievement” system in which they are no longer subject to the token economy, for the most part. At this stage, natural and logical consequences and self-regulation are emphasized.

Though no research has been done on the motivation system levels component of GBT’s treatment, the overall treatment model, which has many other components, has been studied. GBT uses the Teaching-Family Model (TFM), which was pioneered at Achievement Place by behavioral researchers from the University of Kansas (see, e.g., Phillips, 1968; Phillips, Phillips, Fixsen, & Wolf, 1974; Wolf et al., 1976). The TFM is structured around homes for 6-8 youth in which a married couple serves as surrogate parents and the primary treatment agents. One of the reasons that treatment is structured around a family environment is to promote generalization to the family contexts that youth return to or create after leaving GBT. A token economy is used to reinforce the explicit teaching of social, emotional, and life skills. The TFM also focuses on developing positive relationships, academic growth, and self-government and self-determination. The TFM is one of the most common models of treatment for juvenile offenders and behaviorally disordered youth (Larzelere, Daly, Davis, Chmelka, & Handwerk, 2004). In a large-scale early study of the TFM, Kirigin and colleagues (1982) found that youth in TFM treatment had lower rates of criminal behavior and rated staff more highly during treatment compared to a control group, but there were no differences between the groups at 1-year follow-up on criminality or reinstitutionalization.

Some of the research on the TFM has been done at GBT (for a review, see Friman, 2000). Larzelere and colleagues (2004) studied 440 youth who were discharged from GBT between October 1998 and September 2000. Measures were taken at admission, discharge, and 3-month follow-up. Youth tended to leave GBT for a less restrictive environment than the one they came from, and this lower level of restrictiveness was maintained at follow-up. During treatment, youth improved on the broad-band scales of the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001) by approximately one standard deviation. They also had fewer psychiatric diagnoses overall and in every category (disruptive, anxiety, depressive, and substance abuse) 12 months after intake. Fifty-nine percent of youth left GBT after completing the program or graduating from high school, and 87.1% of problems at intake were rated as improved at departure. At follow-up, 81% of youth reported improved relationships with peers, and 90% reported that GBT had a positive or very positive effect. In comparing follow-up respondents to national samples of youth in the same age range, former residents were functioning at the same level across a variety of measures as nationally representative samples.

Thompson and colleagues (1996) examined the educational outcomes of GBT residents compared to a control group of children who were accepted to GBT but did not attend. Outcomes were measured during treatment and continuing for an average of four years after departure. GBT residents had significantly higher grade point averages than comparison youth. Though grades dropped when they left the program, they remained significantly higher than the comparison group's grades throughout the follow-up. GBT youth completed years of school at a faster rate, and significantly more of them

completed high school or received a GED (83% vs. 69%). Attitudes about attending college and frequency of requesting help with homework were both significantly higher for the treatment group. Friman and colleagues (1996) compared the same two groups at baseline and 3 months after admission on a number of attitude measures. They found that, contrary to negative societal beliefs about residential treatment, youth at GBT reported that they were receiving more helpful treatment than comparison youth, they were more satisfied with the adults they were living with, and they reported feeling less isolated from family and friends. They also reported a sense of control that was marginally significantly higher than comparison youth.

GBT's outcomes extend beyond post-treatment and 3-month follow-up. The treatment and comparison groups studied by Thompson and colleagues (1996) and Friman and colleagues (1996) were sought for a long-term follow-up survey an average of 16 years after discharge for the residential group. Thompson, Huefner, Ringle, and Daly (2005) compared residents who remained in the program for at least 18 months to the comparison group and national normative data. All three groups were equally likely to have attained a high school diploma or GED, but the comparison group had a higher proportion who achieved this through a GED than the residential group. The residential group had a much higher rate of military service than the national norms, and the comparison group was somewhat higher than national averages. On other measures, such as being arrested in the past 12 months and using illicit drugs, both the residential and comparison groups were worse off than national norms. Finally, there were a number of areas, including family income and percentage of adults with health care coverage, in which residential, comparison, and national samples had similar rates. Huefner, Ringle,

Chmelka, and Ingram (2007) examined the rates of intimate partner violence in this sample. Residential groups who stayed at GBT for any length of time had significantly lower intimate partner violence rates than the comparison group and were comparable to national norms. For residents in the sample with childhood abuse histories, they reported significantly lower rates of intimate partner violence (14.5%) than the literature predicts for this population (36-42%). Residents who had childhood abuse histories and stayed greater than 18 months had intimate partner violence rates (9.1%) equivalent to the general American population (8.4%).

GBT's treatment model has been replicated with positive outcomes in other settings, including a residential treatment center at GBT Home Campus (Larzelere, Dinges, et al., 2001), a residential treatment center in the northeast (O'Brien, Ringle, & Larzelere, 2001), a children's inpatient psychiatric center (Woodlock, Juliano, & Ringle, 2002), and a day treatment program (Hicks & Munger, 1990).

Because GBT's Teaching-Family Model has a history of positive research outcomes and because GBT uses a stepped-care approach that is similar to the one recommended in this paper to promote internalization and generalization, it is an appropriate setting to test the effectiveness of this approach.

Present Study and Hypotheses

The important test for this theory is not outcome at post-treatment but outcome at follow-up. Traditional token economies have demonstrated the ability to motivate good behavioral outcomes while the child is at the facility. Maintenance and generalization can only be measured after the child has left. Ideally, behavioral outcomes should be measured long after treatment, when the youth reaches adulthood and is no longer subject

to artificial contingency management by parents, guardians, or successive agencies. GBT has conducted two long-term follow-up surveys of former residents. The first, conducted in 2002, targeted youth who entered GBT between 1981 and 1985 and had left GBT 16 years earlier, on average. These former residents were 27-37 years old at the time of follow-up. Outcomes from this sample were described earlier in this section (Thompson et al., 2005; Huefner et al., 2007). The second, conducted in 2005, targeted all youth who were discharged in the year 2000, that is, 5 years after departure. These former residents were 17-24 years old at the time of follow-up.

Since this study was conducted in an existing treatment setting and promotion to higher levels of the token economy was based on merit, it was impossible to randomly assign youth to conditions or motivation system levels. Therefore, this study tested whether the motivation system level that the youth departed on (typically, this was the highest one achieved while at GBT) was predictive of long-term follow-up outcomes. Because the motivation system variable represents more than just the treatment itself – it is also reflective of the child’s success and progress in the program and judgments about the child’s internalization – it was important to control for those components of the variable. Therefore, a measure of success in the program, Favorable Departure Rating, was included in a parallel set of predictor models to see whether or not any outcomes predicted by motivation system level were better accounted for by the child’s general success in treatment, as opposed to the particular effects of the motivation system. Favorable Departure Rating is a mean of a number of rating scales assigned to each youth upon departure from GBT. The child’s achievement of treatment goals, the favorability of his or her departure, and the staff’s predictions of the child’s future success influence

the rating. It shares a good deal of variance with the motivation system variable because both capture, in part, the child's successful progress through the treatment program and his or her internalization of the skills that GBT teaches. Since the goal of this study was to examine the effects of the different motivation systems as treatments, rather than as markers of children's success, Favorable Departure Rating was entered into the analyses in order to control for treatment success and related constructs on long-term outcomes. If the motivation system variable predicts outcomes even when controlling for success in treatment and the staff's predictions about the youth's future success, a conservative test given the length of follow-up and the fact that the Favorable Departure raters had access to the youth's motivation system level when assigning their ratings, then this may suggest that the actual motivation system had an influence beyond its correlation with successful treatment. Other predictors from baseline and post-treatment were entered into the models in a pragmatic way to control for otherwise unexplained variance. There were no specific hypotheses associated with these predictors, though it was interesting to see which predictors were associated with long-term outcomes.

The hypothesis of this study is that residents who reached higher motivation levels by the time of departure from GBT would have more positive behavioral outcomes over the long-term (5-16 years post-treatment) than those who did not. This prediction was expected even after controlling for another measure of success in the program (i.e., Favorable Departure Rating).

Method

Participants and Setting

The Father Flanagan's Boys' Home was founded as an orphanage for boys in Omaha, Nebraska, in 1917. Father Flanagan had the revolutionary idea that, if helped when they were young, boys in trouble would not become homeless men or criminals when they grew up. Because the demand was great and he did not want to turn away any boy who needed help, a few years later he purchased a 160-acre farm 10 miles west of Omaha to found Boys Town. In 1936, it was recognized by the state as its own village. The mayor and council members of the village are elected by and from the adolescent residents as part of a self-government system. Over the years, Boys Town shifted its focus from orphans in general to youth needing residential placement because of multiple personal and systemic problems, such as behavior disorders, histories of abuse, criminal records, and unpredictable environments. In 1979, girls began to receive treatment at Boys Town, and, in 2000, the residents voted to change the name to Girls and Boys Town. There are now nearly 500 residents at the Home Campus in Boys Town, Nebraska. As many as 47,000 children receive treatment each year at 19 GBT sites across the United States.

Long-term follow-up surveys were administered to two cohorts of residents at different periods of time after leaving GBT. Demographic, pre-treatment, and post-treatment variables for the follow-up survey respondents were taken from treatment and previous research records. Previous residents eligible for the study represented the great majority of residents at GBT during the periods described.

16-Year Follow-Up

The first cohort was admitted to GBT between June 21, 1981, and July 9, 1985. These residents departed GBT between November 1981 and April 1991. Four hundred ninety-seven residents met the criteria for inclusion and agreed to participate in a longitudinal study that began when they entered GBT and ended in 1989. Inclusion criteria have been described extensively elsewhere (Friman et al., 1996). The cohort consisted of 459 (92%) males and 38 (8%) females, reflecting the pattern of admission in that era. Most were Caucasian ($n = 345$, 69%), but a significant proportion were African American ($n = 98$, 20%), Hispanic ($n = 29$, 6%), and “Other” ($n = 25$, 5%). The adolescents were 9-18 years old at admission, with a mean of 14.5 years and a standard deviation of 1.6 years.

In 2002, attempts were made to contact 410 of these former residents because they had *not* previously requested “no further contact.” After extensive efforts, 243 were found (59% find rate). Of those found, 19 had passed away, and 4 refused to participate. A total of 220 consented to participate, and 207 of them completed the follow-up survey. The number of respondents represents 92% of living individuals found, 50% of attempted contacts, and 42% of the entire cohort. (It should be noted that an additional 3 former residents completed the follow-up survey, but they were not eligible for this study because one had not participated in the original longitudinal study, and the other two had been part of a comparison group before entering GBT, so their baseline data were not at the time of admission to GBT.) Of the 207 survey respondents, 87 completed the survey by mail, and 120 did so by telephone.

Surveys were conducted an average of more than 16 years after discharge for these residents. Survey completers were 27-37 years old, with a mean age of 32.7 years ($SD = 1.9$). They were first admitted when they were 9-17 years old, with a mean age of 14.3 years ($SD = 1.7$). The racial distribution of the respondents was almost identical to the overall cohort with 71% Caucasian, 19% African American, 5% Hispanic, and 5% classified as a different race. Again, the participants were predominantly male (90%). Participants resided at GBT for an average of 27.0 months ($SD = 19.5$), but the range was very broad, from 13 days to 105.6 months.

Those who participated consented to do so under the guidelines set by the Father Flanagan's Boys Home Human Rights Review Committee (GBT's IRB). Participants were offered \$50 for completing the survey.

5-Year Follow-Up

The second cohort consisted of individuals who were discharged from GBT Home Campus anytime in the year 2000. This yielded a potential subject pool of 214 former residents. (One individual, who was admitted and discharged on the same day, was excluded.) The cohort consisted of 132 (62%) males and 82 (38%) females. More than half were Caucasian ($n = 125$, 58%), but a significant proportion were African American ($n = 49$, 23%), Hispanic ($n = 22$, 10%), and "Other" ($n = 18$, 8%). The adolescents were 8-17 years old at admission, with a mean of 15.2 years and a standard deviation of 1.6 years.

Follow-up data were collected in 2005, 5 years after discharge. After extensive efforts to find and contact all former residents who departed in 2000, 148 were found (69% find rate). Of those found, 131 completed a follow-up survey (89% response rate;

61% overall response). Of those who completed the survey, 83 did so by mail or on the Internet, and 48 did so by telephone. Participants were 17-24 years old, with a mean age of 21.9 years. Most were Caucasian (63%), 23% were African American, 5% were Hispanic, 2% were Native American, 1% was Middle Eastern, and 6% were multi-ethnic. The sample was half female (49.6%) and half male (50.4%). Participants resided at GBT for an average of 20.2 months ($SD = 17.5$), and the range was again very broad – 17 days to 116.7 months.

Those who participated consented to do so under the guidelines set by the Father Flanagan's Boys Home Human Rights Review Committee. Participants were offered \$25 for completing the survey.

Consent Procedures and Deidentification of Data

This study was an analysis of existing data from two ongoing research projects being conducted by the Girls and Boys Town National Research Institute for Child and Family Studies. The studies were 5-year and 16-year follow-up studies of treatment outcomes of GBT's residential treatment program. All respondents consented to complete a questionnaire and allow their GBT treatment records to be included in the studies. The data came from those sources – the follow-up questionnaires and treatment records. To further protect the rights of the participants, any information that could identify participants was removed by GBT before researchers outside of GBT viewed or analyzed the data. No one outside of GBT had the ability to match the data to identifying information.

GBT Long-Term Residential Program

The goal of this study was to test differential long-term outcomes based on one component of GBT's treatment model, the motivation systems. Those systems are one aspect of a comprehensive treatment approach. The overall model is summarized here followed by a discussion of the various motivation systems.

The treatment model, called the Teaching-Family Model (TFM), is based on the well-described Achievement Place model (e.g., Phillips, 1968; Phillips, Phillips, Fixsen, & Wolf, 1974; Wolf et al., 1976), which was designed by behavioral researchers at the University of Kansas and has been studied and replicated in diverse treatment settings (Gottschalk, Davidson, Mayer, & Gensheimer, 1987; Hicks & Munger, 1990; Jones, Weinrott, & Howard, 1981; Kirigin et al., 1982; Larzelere et al., 2004).

The TFM places 6-8 youth in homes with highly trained married couples ("Family-Teachers"), who supervise treatment and administer the program. There are five core elements of the model. First, youth are proactively taught social skills and life skills. Specific target skills are reinforced with the motivation systems, described below. Second, an emphasis is placed on residents' developing healthy, positive relationships at GBT. Third, GBT promotes residents' moral and spiritual development through weekly attendance at religious services and coursework on values in school. Fourth, the TFM strives to create a positive, realistic family environment for every adolescent, and the youth are also encouraged to strengthen relationships with their natural families. Fifth, youth are given many opportunities for self-government and self-determination (J. Davis & Daly, 2003). The TFM is designed to create an all-encompassing environment that promotes positive behavior and healthy relationships. This emphasis on the child's

whole ecology is intended to teach and reinforce positive behavior in multiple domains in order to promote generalization. Residents attend GBT's on-campus middle and high schools, which operate within the same therapeutic structure and coordinate closely with family-teachers. GBT also works with residents' families to build healthy relationships. Parents are required to complete a six-week parenting course based on the TFM before their child's first home visit. The treatment is described in a detailed treatment manual (J. Davis & Daly, 2003) and other training documents, and treatment fidelity is maintained through continuous training, supervision, consultation, and evaluation (Larzelere et al., 2004).

In addition to the TFM milieu, more traditional treatments, such as pharmacotherapy, psychotherapy, and 12-step sobriety group treatments, are provided to youth who need them. The typical duration of residence at GBT is 18-24 months. Youth are referred by their families and by many agencies but are only accepted if the youth themselves consent to enroll. Generally, they are encouraged to stay until they complete high school at GBT.

Motivation Systems

GBT's motivation systems are complex; they are described in detail in the *Girls and Boys Town Long-Term Residential Treatment Manual* (J. Davis & Daly, 2003). A summary follows.

Daily Points System. When youth enter GBT, they begin on a Daily Points System in which they receive points for good behavior and lose points for bad behavior throughout a 24-hour period. Points are tallied on "point cards" that the youth carry with them at all times. Points are added or subtracted in "teaching interactions," during which

a staff member and the child discuss the specific positive or negative behavior at hand. Target behaviors are taken from the curriculum of social, emotional, and personal responsibility skills the youth are taught. Family-teachers are expected to alter their rate of interaction with the residents based on the youth's level of functioning. A novice youth may participate in 15-18 interactions per day while a veteran may be involved in as few as 8. Staff members are expected to maintain a ratio of four positive teaching interactions for each negative one (Field, Nash, Handwerk, & Friman, 2004).

At the end of each 24-hour period on the Daily System, residents sit down individually with their family-teachers for a "card conference." The card is reviewed, and points earned and lost are totaled up. On days in which children earn a certain net total of points, they receive access to five privileges over the following 24 hours. These privileges are basic household privileges (such as listening to the radio in their rooms), sweet or salty snacks, television, telephone, and free time on campus. In addition, for each day that youth have positive point balances, they move closer to graduating to the Weekly Points System. When they have reached the criteria for graduation from the Daily System and the treatment team agrees that progress has been reliable and that the child is ready, he or she progresses to the Weekly System.

Weekly Points System. The Weekly Points System has similar mechanics. Residents carry daily point cards and meet with family-teachers to review their cards on a daily basis. However, the points are totaled weekly, and privileges are earned for the following 7-day period. New privileges are also available, including free time off campus, special privileges (e.g., later bedtime, renting a video), and "bonds," which track progress toward graduating from the weekly system. When the resident is new to the

weekly system, he or she must still reach a certain level of net points daily to continue to receive that week's privileges. As the adolescent is successful with the weekly system, the daily requirements are gradually reduced to 0, though daily conferences are held to maintain daily individual contact with the family-teacher and to continue to focus on treatment goals. Youth can graduate from the Weekly System when they accumulate a specified number of bonds and the treatment team agrees that progress has been reliable, treatment goals are being met, and the resident is ready to succeed on the next system.

Achievement System. When youth graduate from the Weekly System, they reach the Achievement System. Achievement is supposed to be closer to the model of a typical family rather than a therapeutic environment. On Achievement, residents do not carry around point cards, and they receive all of the privileges available on the Daily and Weekly Systems without having to earn them through points. They continue to receive feedback and consequences in a number of ways that are unique to the system. First, family-teachers are expected to continue to initiate teaching interactions throughout the day in which they offer praise or correction, but these interactions are less frequent than on the other systems. Second, feedback is often delayed until the daily conference with the family-teachers rather than given immediately. Third, at the daily conference, youth are asked to talk about progress toward treatment goals and self-report problems that happened out of view of their family-teachers. Fourth, points are given and taken away at the daily conference for overall behavior and progress toward treatment goals, but these points are not connected to daily privileges. They are stored in an "Achievement System Savings Book" and can be used periodically to purchase special or creative privileges, such as extra free time, special games, additional television time, choosing the

restaurant the “family” goes out to, attending a concert, and so on. Fifth, when youth do not meet expectations, the consequences are often natural or logical. For example, if a resident on Achievement fails to hand in an assignment in school because she spent extra time at the gym, the consequence might be that, for the next two days, she cannot go to the gym and has to complete her homework at home immediately after school. Youth at this level are expected to take more responsibility for their actions and are given more autonomy. Consequently, they tend to have a higher status within the home.

A primary goal of the Achievement System is to begin the process of stimulus generalization to the post-treatment environment. Much of its structure is based on suggestions from the generalization literature for social learning theory-based interventions (e.g., Kazdin & Bootzin, 1972; Stokes & Baer, 1977; Stokes & Osnes, 1989), such as the use of social reinforcement, fading token reinforcement, delay of token reinforcement, delay between receiving tokens and exchanging them for back-up reinforcers, and the use of naturally occurring reinforcers. The concept underlying this level is that residents’ behavior is primarily motivated by the natural and social reinforcement in the environment. Furthermore, increased autonomy, trust, and responsibility are granted, which support the residents’ need for autonomy. The logical consequence for misusing that autonomy, trust, and responsibility is the loss of the same, which is part of the system. If a resident on Achievement misbehaves in a serious way, he or she may be removed from the Achievement level temporarily or permanently.

Though partially based on the concepts described in the introduction to this paper, natural and logical consequences and self-evaluation are not the only means of motivation in Achievement. Some artificial, nonendogenous reinforcers (including

points) continue to be used. Therefore, the Achievement System should be considered a hybrid between a token economy and an endogenous system, meant to serve as a transition between a token economy and a typical family environment.

During the time that the participants were at GBT, there was another motivation system called “Natural and Logical,” a term borrowed from the type of consequences by which the adolescents were supposed to be motivated. However, this system was not as well defined as the others, and youth were not placed in it, as opposed to Achievement, in any systematic way. Therefore, it was subsumed within Achievement in the current treatment manual (J. Davis & Daly, 2003). Because they are not theoretically or practically distinct, youth who were in the Achievement and Natural and Logical groups were combined when analyzing the data for this study.

Subsystems. In addition to the Daily, Weekly, and Achievement systems, youth who commit severe misbehaviors, such as substance use, theft, or running away, are placed on a Subsystem. Subsystems are the most restrictive motivation systems at GBT. There are a number of Subsystem types, but the basic feature is that they all require higher daily point totals than the Daily System to earn fewer privileges. Generally, youth are not on a Subsystem for more than a month at a time before returning to either their previous motivation systems or to the Daily System. Youth who end their residence at GBT on a Subsystem often have committed a serious offense soon before leaving. (In some cases, that misbehavior precipitated the discharge.) When someone leaves while on a Subsystem, he or she has apparently failed to internalize the skills that GBT aims to teach.

Measures

The two cohorts studied here began their tenures at GBT about 15 years apart, on average. Therefore, they represent different populations, they received treatment under different treatment manuals, the data collected about them was somewhat different, and the follow-up surveys they completed differed and were completed at different life stages. Therefore, the analyses cannot be combined. This paper should be viewed as two studies with the potential for replication, not as one large study with two samples. For both studies, the independent variable was the motivation system of the children at departure from GBT. The dependent variables were various self-reported outcomes at long-term follow-up taken from the surveys that the former residents completed. In addition, demographic, pre-treatment, treatment, and post-treatment variables were entered into the analyses as control variables. The variables used for the 16-year follow-up group are described first.

16-Year Follow-Up

Demographics. Demographic and pre-treatment variables for the 16-year follow-up cohort are summarized in Table 1. Gender, ethnicity (including minority status), and age at admission were taken from GBT admissions records. The respondents' age at follow-up was taken from the follow-up survey.

History. Pre-treatment variables were organized into conceptual groupings. The first group, which dealt with the child's systemic and academic history, was composed of the Restrictiveness of Living Environment Scale (ROLES; Hawkins, Almeida, Fabry, & Reitz, 1992), the number of prior placements, the number of presenting problems, and estimated GPA. The first three variables in this group were derived from the residents'

admission records, and estimated GPA was taken from an item on the pre-admission questionnaire, described below.

The current version of the ROLES is a 30-point scale of the restrictiveness of the living situation that the child came from immediately before arriving at GBT. Each level is a distinct category ranging from 1 (jail) to 25 (independent living by self). (Note that 5 of the levels are not integers. For example, military is 13.5.) The ROLES for each child was completed retroactively by researchers based on data from the admissions files. Because of the limited data and the limited options for placement when this cohort entered GBT, only 6 of the ROLES categories were, in practice, endorsed.

The number of prior placements variable represents the total number of both formal and informal placements the child experienced before arriving at GBT. Formal placements include foster homes, detention centers, and psychiatric hospitals. Informal placements are placements with relatives. The number of presenting problems is a count of all of the items endorsed on a checklist of 16 problems, such as parental rejection, youth “out of control,” school attendance problems, delinquency, substance involvement, and running away. Both number of prior placements and presenting problems were recorded by the admissions counselors upon the child’s entry to GBT.

As part of the original longitudinal study, youth in this cohort completed an extensive pre-treatment questionnaire, which was administered in an interview format. Estimated GPA, the Aggressive Offense Scale, the Property Offense Scale, the Drug Scale, the Consumer Scale, the Locus of Control Scale at admission, and the Happiness Rating were all taken from this questionnaire. These scales, their coding, the original sources of the questions, and the internal reliability in this sample are described in Table

1. Further reliability information, including alpha coefficients and test-retest reliability in the original test samples, can be found in Chmelka and Noss (1990) and in the source questionnaire reports (Elliott, Ageton, Canter, & Huizinga, 1979; Nowicki & Strickland, 1973; Phillips et al., 1974). When individual item responses were missing in a scale, data were imputed only for cases with 70% or greater of the items completed. If 70% of the items were completed, the average of the available data points was taken. If less than 70% of the items were completed, the case was excluded from analyses of that variable. Unless otherwise noted, this is how data were imputed for all measures in this paper.

Self-reported delinquency. The Aggressive Offense Scale, the Property Offense Scale, and the Drug Scale comprised the self-reported delinquency group. These scales were all derived from specific yes-no questions about delinquent acts on the pre-admissions interview. Affirmative responses were coded as 1, and negative responses were coded as 0. The scales were averages of their component items.

Psychological scales. The Consumer Scale, the Locus of Control Scale at admission, and the Happiness Rating composed the psychological scales group. The Consumer Scale was the mean of four 7-point questions about the fairness, concern, effectiveness, and pleasantness of the adults with whom the child was living before admission. The Locus of Control Scale was composed of eight dichotomous items that measured the degree to which the child's locus of control was internal or external at admission. The Happiness Rating was a single 3-point item about overall happiness.

Treatment data. Treatment and post-treatment data are summarized in Table 2. Length of stay was calculated based on the total number of days the resident spent at GBT. For the 60 youth who had two distinct stays at GBT and for the 3 youth who had 3

distinct stays at GBT, the total number of days in residence was summed because the time in between stays was usually much smaller than the time in residence and because treatment generally continued where it had left off when adolescents returned. Length of stay has been shown to be a positive predictor of outcome in previous evaluations of GBT, including with this sample (see Thompson et al., 2005).

Motivation system at departure was taken from residents' electronic or paper records. While the hypotheses of this paper would be better served if this variable strictly represented a type of intervention, this variable is not that pure. First, while youth may have left on a particular motivation system, no data are available about how long that child had spent on the motivation system prior to leaving or how long the child spent on other motivation systems during his or her stay. It is possible that some youth spent only a short period of time in the last motivation system before departing. Second, motivation system was not randomly assigned. On the contrary, it was assigned based on performance criteria and clinical judgment of a child's readiness and degree of internalization. Therefore, the motivation system variable does not only tell what intervention the child received, it is also an evaluation of the child's progress in the program. Third, while there are specific criteria in the treatment manual for moving between motivation systems, clinical judgment is a necessary part of the decision-making process. Therefore, there is some subjectivity in assigning motivation systems. Furthermore, clinical realities, such as a youth's anticipated departure, may have influenced the assignment of motivation system. For example, clinical staff may have accelerated some youth to Achievement before graduating from high school in order to promote generalization. On the other hand, some adolescents who had been progressing

well in the program may have “done something stupid” shortly before leaving, which would have affected their motivation system levels at departure. Fourth, progress to higher motivation systems is somewhat time-dependent. Even a child succeeding in the program would take a number of months – typically more than a year – to reach Achievement. The mean length of stay for residents who left at the Achievement level was nearly 3 years, in contrast to an average length of stay of 20 months for those who left under other motivation systems. Because of the many influences on this variable, it was important to control for other proxies of success in the program in order to attempt to partial out the effects of the treatment. However, because of the high intercorrelations between motivation system and other treatment success variables and the fact that other variables were affected by it, this made the primary analyses a conservative test of the hypothesis.

Post-treatment. Favorable Departure Rating is the mean of three 5-point ratings given by the youth’s clinical specialist (the staff member overseeing the child’s treatment) at departure. The three ratings are Favorable Conditions of Leaving, Goal Achievement, and Prediction of Future Success. These ratings take into account the child’s behavioral status at departure, progress on treatment goals, the youth’s success within the treatment environment, perceptions of the child’s internalization and generalization of positive behavior, and so forth. The motivation system level reached is taken into account in these ratings. In addition to current behavior and perceived internalization, Prediction of Future Success also takes into account the supportiveness and restrictiveness of the environment the child is going to, parental involvement, the youth’s educational achievement, and his or her post-treatment plans. Favorable

Departure is GBT's global assessment of treatment success at post-treatment. It is not independent of motivation system at departure because that is part of the judgment.

Follow-up survey. The 16-year follow-up participants completed a 151-item follow-up survey that was compiled from items used on eight national surveys that each attempted to characterize the population as a whole: the Behavioral Risk Factor Surveillance System (2003), the General Social Survey (J. A. Davis, Smith, & Marsden, 2003), the National Survey of Families and Households (Sweet & Bumpass, 1996), the Social Capital Community Benchmark Survey (2000), the National Household Survey on Drug Abuse (National Household Survey on Drug Abuse, 1985), the National Youth Survey (Elliott et al., 1979), the National Survey of America's Families (Wang Cantor, & Vaden-Kiernan, 2000), and the 2000 Census (United States Census Bureau, 2003). The 151 questions on the GBT follow-up survey were divided into eight topic areas: (1) living environment and community involvement; (2) physical and mental health and well-being; (3) substance use; (4) household composition and family relationships; (5) safety, victimization, and criminality; (6) friendships and social activities; (7) education and employment; and (8) current perceptions of the intervention.

Items and scales from the 16-year follow-up survey chosen for this study are summarized in Table 3. They were chosen for their relevance as social, behavioral, or psychological outcomes. Summary variables were created to reduce the number of variables and to combine similar items. Efforts were made to ensure that the internal reliability of scales was acceptable. The Perpetration Index alpha coefficient is low (.56) because it is a sum of crimes committed in the past 12 months across a number of

domains. While the frequency of perpetration in the various domains was not highly related, the overall count of crimes committed has face validity.

Data were imputed for all scales of 4 or more items as described above except in the case of the Social Involvement Count. That variable is based on the number of groups (out of 4) that a respondent reported socializing with at least monthly. If respondents had 2 or fewer responses, his or her data were excluded. If respondents had 3 valid responses and 1 “not applicable,” the total was calculated based on the missing item scored as 0 because it was assumed that the respondent did not socialize with that group.

Outcomes variables derived from the 16-year follow-up survey were also grouped into conceptual categories. *Psychological outcomes* included the mental health index, recent substance abuse, locus of control at follow-up, the satisfaction index, happiness at follow-up, and whether or not the person was deceased at follow-up. These were chosen to correspond, as much as possible, to the psychological scales at baseline. *Criminality* was measured by whether or not the individual was arrested in the past 12 months and the number of crimes perpetrated in the past 12 months (Perpetration Index). *Education* was measured by the highest grade completed to date and whether or not the person was in school in the past 12 months. The *employment and earnings* group consisted of whether or not the person was currently working, a job satisfaction scale, household income, whether or not the person was unable to pay basic bills in the past 12 months, and whether or not the individual had served in the military. This last variable, military service, is an important outcome for GBT alumni because many of GBT’s most successful graduates go on to military service. The respondent’s involvement in his or

her *social system* was measured by whether or not he or she was currently married or living with a partner, the child involvement index, a count of the number of domains of the person's social involvement, and self-reported spirituality. Finally, the respondent's retrospective *evaluation of GBT* was measured with the GBT Index.

5-Year Follow-Up

Demographics. Demographic and pre-treatment variables for the 5-year follow-up cohort are summarized in Table 4. Gender, ethnicity (including minority status), age at admission, and age at departure were taken from GBT records. Since the survey was administered roughly 5 years after departure for this group, the age at departure varies systematically with age at follow-up. The IQ score was taken from a paper file review of available admissions data conducted by research staff at GBT to ensure the reliability of certain scales. However, not all paper files were available. Out of 214 members of the cohort (including both respondents and nonrespondents), 179 have file review data.

History. Pre-treatment variables were organized into conceptual groupings. The first group, which dealt with the child's systemic and academic history, was composed of ROLES at admission, number of formal prior placements, number of presenting problems, number of family problems, and whether or not the child had an arrest history. ROLES (Hawkins et al., 1992) at admission, the number of formal prior placements, and arrest history were taken from the file review. ROLES was described above; in the file review, youth received a greater variability of scores than youth in the first cohort. The number of formal prior placements represents the total number of formal placements the child experienced before arriving at GBT. As above, formal placements included foster homes, detention centers, and psychiatric hospitals. Informal placements were not

included in this count. The file review also yielded a dichotomous variable indicating whether or not the child had ever been arrested before arriving at GBT.

The number of presenting problems and the number of family problems were each taken from residents' admission records. The total number of presenting problems in this cohort was collected in a similar way as the previous cohort except that the number of potential categories was 38 instead of 16. In addition, a checklist of 34 family problems (e.g., alcohol/drug use immediate family, suicide in immediate family, domestic violence, incarceration of parent) was recorded by the admissions counselor; the number of family problems is the total number of categories endorsed on that list.

Psychopathology at admission. The National Institutes of Mental Health Diagnostic Interview Schedule for Children (DISC; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000) is a computer-based structured diagnostic interview. The DISC yields DSM diagnoses based on the youth's responses to the items in the interview. The DISC was administered to youth upon admission to GBT. Because GBT was a test site for the DISC, a number of different versions of the DISC were administered to the adolescents, depending upon when they entered GBT. The DISC-IV Generic was a major revision of the DISC 2.3 Generic. The DISC 2.3 (administered to 35 youth) generally assesses diagnoses over the past 6 months, whereas the DISC-IV Generic (104 youth) assesses them over the past year. The DISC-IV also includes some additional diagnoses that were not included in the DISC 2.3, such as manic disorders. In addition, the Present State DISC-IV, which generally assesses symptoms over the past month, was administered to 66 youth. The Present State DISC-IV can be administered in one of two ways. One way is for an interviewer to read the prompts on a computer program and

enter the child's responses (DISC-IV-PS; 30 youth). The other way is for the computer program to speak the questions automatically (through wave files) and allow the child to respond without mediation (DISC-IV-PS-Voice; 36 youth). No matter how administered, a computer algorithm assigns the diagnoses.

One study (with a sample from GBT) found that rates of conduct disorder were similar across different versions of the DISC, suggesting that the assignment of disruptive behavior disorders by different versions of the DISC is roughly equivalent (Jewell, Handwerk, Almquist, and Lucas, 2004). Unlike other disorders, the timeframes for these diagnoses are similar across versions in order to match the timeframes in the DSM, and many of the questions are worded similarly in different versions. Nevertheless, collapsing across the versions of the DISC must be done cautiously as there are no data available to corroborate that finding with other classes of disorders, and the different timeframes used in the different versions may have weakened reliability across versions. Therefore, individual diagnoses were not examined. Rather, total number of common diagnoses across versions, total number of common internalizing diagnoses, and total number of common externalizing diagnoses were used. A dichotomous variable of "any substance abuse disorder" was also used.

Treatment data. Treatment and post-treatment data are summarized in Table 5. Length of stay was calculated based on the total number of days the resident spent at GBT. There were no youth with multiple stays in this cohort.

Motivation system at departure was taken from residents' records. All of the caveats described above in relation to this variable apply here, as well.

Post-treatment. Favorable Departure Rating in this sample is the sum of four 7-point ratings given by the youth's clinical specialist at departure. In addition to the three ratings, Favorable Conditions of Leaving, Goal Achievement, and Prediction of Future Success, used with the earlier cohort, an Overall Behavior rating was added to characterize the youth's general behavior at GBT before leaving. Again, these ratings were not independent of motivation system at departure because that was part of the judgment. In a previous study (Larzelere, Smith, Jorgensen, Daly, & Handwerk, 2001), treatment success was measured using the Departure Success Scale. That scale summed the Favorable Departure Rating (that is, the sum of its 4 constituent ratings, each with a 1-7 scale) with a transformed version of the ROLES (range 1-6), and a Program Completion Indicator (1 for complete, 0 for not complete) yielding a total scale range of 5-35. Though Larzelere, Smith, and colleagues (2001) found strong internal reliability for the scale ($\alpha = .89$), that was not the case for the present sample ($\alpha = .45$). Therefore, the Favorable Departure Rating ($\alpha = .94$) was used to control for treatment success in this study.

Follow-up survey. The 5-year follow-up participants completed a 93-item survey that was compiled from items used on three national surveys. Most items were from the National Longitudinal Survey of Youth 1997 (NLSY97; Center for Human Resource Research, The Ohio State University, 2005), but some were from the National Survey on Drug Use and Health (Research Triangle Institute, 2004) and the National Health and Nutrition Examination Survey (National Center for Health Statistics, Centers for Disease Control and Prevention, 2007). The 93 items on the GBT follow-up survey were divided into eight topic areas: (1) living environment; (2) family relationships and social support;

(3) religion, health, and well-being; (4) criminality; (5) substance use; (6) education; (7) employment and income; and (8) current perceptions of the intervention.

Items and scales from the 5-year follow-up survey chosen for this study are summarized in Table 6. They were chosen for their relevance as social, behavioral, or psychological outcomes. Summary variables were created to reduce the number of variables and to combine similar items. Efforts were made to ensure that the internal reliability of scales was acceptable. Outcomes variables derived from the 5-year follow-up survey were also grouped into conceptual categories. *Psychological outcomes* included the mental health index, recent substance abuse, and friends' substance use. These were chosen to correspond, as much as possible, to the psychological scales at baseline. *Criminality* was measured by whether or not the individual self-reported any crime in the past 12 months, whether or not the he or she was arrested in the past 12 months, and whether or not he or she was currently incarcerated at the time of the survey. *Education* was measured by the highest grade completed to date and whether or not the person was enrolled in school in the past 12 months. The *employment and earnings* group consisted of whether or not the person worked during the previous week, household income, and whether or not the individual had served in the military. The respondent's involvement in his or her *social system* was measured by whether or not he or she was currently married or living with a partner, the closeness to that partner, the amount of conflict with that partner, the amount of child involvement, whether or not the person had a supportive parent, and the importance of religion.

While there was an effort to create similar outcome measures across the two long-term follow-up surveys, the general content and specific wording of items made this

difficult. Therefore, even if items or scales on the two surveys have similar or identical names, one should not assume that they are identical in nature. In the vast majority of cases, the precise content of the items was different. Furthermore, these samples cannot be expected to be similar. In addition to their having received treatment at GBT, on average, 15 years apart, the respondents were also at very different life stages at follow-up. The earlier cohort was 27-37 years old, while the later cohort was just 17-24 years old.

Data Analytic Strategy

Because of the many differences between the populations and the eras in which they resided at GBT, data were analyzed separately for each cohort. Nevertheless, a similar strategy was used for each.

Preliminary Analyses

In order to characterize how well the survey respondents represented the cohorts of residents from those years, survey respondents and nonrespondents during the corresponding years were compared across all demographic, baseline, treatment, and post-treatment variables.

Primary Analyses and Hypothesis

Residents of Girls and Boys Town who were at the Achievement System motivation level at the time of departure were hypothesized to have better long-term outcomes (as measured by the 5- and 16-year follow-up surveys) than residents who were discharged while on a more structured motivation system. Likewise, residents discharged while on the Weekly Point System were hypothesized to have better long-term outcomes than residents who departed while on a more restrictive motivation system. Therefore,

motivation system at departure was treated as a 4-point ordinal variable with Achievement as the highest level.

Within each group of outcome variables derived from the follow-up surveys, outcome variables were individually regressed onto motivation system at departure with other predictors in a simultaneous linear or logistic regression. Predictors were taken from the demographic, baseline, treatment, and post-treatment variables. Predictors for a given dependent variable were identified by their bivariate correlations with that dependent variable. This technique has been suggested to control for variability in the dependent variable in order to increase the likelihood of observing a relationship, if any, between the independent variable (in this case, motivation system) and the dependent variable. Rather than enter all of the potential predictors, only predictors that correlated with the dependent variable were entered into the regression in order to preserve degrees of freedom and, because of intercorrelations between predictors, to protect against threats to tolerance. If the goal had been to systematically test the relationships between the predictors and the follow-up outcomes, then predictors would have been entered on theoretical grounds in planned analyses. However, since this was a test of motivation system and Favorable Departure Rating, not the predictors, the latter were included strictly on practical grounds.

The first regression performed with each long-term outcome variable was a relatively liberal test of motivation system because treatment success was not systematically controlled for. Therefore, it was possible that significant findings for motivation system could have been artifacts of general success in treatment rather than the motivation system intervention itself. To compensate for this, a second linear or

logistic regression was performed with each outcome variable that was identical to the first analysis except that Favorable Departure Rating was included as a predictor. Favorable Departure represents the child's success in treatment at GBT, perceived internalization of the skills learned, and the child's prospects for maintaining gains after departing from GBT. It is a marker both for the state of the child at post-treatment but also for the child's success at navigating and succeeding in a complex treatment environment. An important component of the measure is the staff's assessment of future success based not only on the child's treatment history and current functioning but also on the supportiveness of the environment the youth is returning to and the strength of his or her future plans. Motivation system is also a marker for success in treatment and internalization because residents were not permitted to reach higher levels without some success and internalization on previous levels. Because the motivation system variable represents not only the intervention implied by the system but also these other factors, Favorable Departure was included to try to parcel out the variability in motivation system related to the child's treatment success and personal strengths. Of course, it is impossible to completely disentangle these two variables because they take so many common factors into account. Furthermore, Favorable Departure Rating includes elements beyond success in treatment that likely gave it some advantages over motivation system. Specifically, Favorable Departure Rating includes a prediction of the child's future success based, in part, on the youth's perceived internalization at departure, his or her future plans, his or her next placement, and the supportiveness of his or her post-treatment environment. It also takes the motivation system at departure into account because that is part of the data available to the rater at the time of departure. Finally,

Favorable Departure Rating may account for more variance because it contains more variance by definition. It is a 13-point scale in the 16-year follow-up and a 25-point scale in the 5-year follow-up, compared to motivation system, which is a 4-point scale in each. Clearly, only variance highly distinct to motivation system would emerge as a significant effect when controlling for Favorable Departure Rating. Therefore, this was considered a conservative test of the hypothesis to counterbalance the liberal test described above. Where these analyses are presented in tables, the respective approaches are both included and labeled as Liberal and Conservative.

Motivation system at departure was hypothesized to predict long-term outcomes even when controlling for measures of success in the program (i.e., Favorable Departure Rating). Compared to other predictors of long-term outcomes, level of motivation system was hypothesized to be relatively robust.

Secondary Analyses

A series of exploratory analyses looked at which demographic and baseline variables predicted post-treatment outcome, particularly motivation system at departure, Favorable Departure Rating, and length of stay. Each of these treatment and post-treatment variables was regressed onto demographic and pre-treatment variables to begin to develop a model of the relevant admissions measures that predict a child's likelihood of successfully completing treatment at GBT, based on both within-setting and external definitions of "success."

Results

16-Year Follow-Up

Preliminary Analyses

Descriptive statistics for the demographics were reported in the Method section. Means, standard deviations, and sample sizes for all other variables are reported in Table 7.

Long-term follow-up survey responders were compared to nonresponders in the same cohort on all demographic, baseline, treatment, and post-treatment variables to determine their representativeness of the cohort. Chi-square tests were used for categorical and dichotomous variables, and MANOVAs were used for groups of continuous variables. There were no differences on any demographic or baseline variables except for age at admission. At admission, respondents (mean = 14.3 years) were approximately 5 months younger than nonrespondents (mean = 14.7 years), $F(1, 495) = 8.31, p < .01$. Respondents were also more likely to have multiple stays at GBT than nonrespondents (means = 1.18 vs. 1.10), $F(1, 495) = 5.92, p < .05$. These differences may be a function of the systematic difference in length of stay between respondents (mean = 823.14 days) and nonrespondents (mean = 706.51 days), $F(1, 495) = 5.85, p < .05$. Respondents also had significantly higher Favorable Departure Ratings (means = 2.43 vs. 2.17), $F(1, 467) = 5.70, p < .05$. There was no significant difference between the respondents (mean = 1.25) and nonrespondents (means = 1.23) on motivation system, $F(1, 444) = 3.59, p > .05$. It is not surprising that youth who stayed at GBT longer and had more favorable outcomes were more likely to respond to the survey.

However, these significant findings may also be a result of the large number of variables compared.

Primary Analyses

Psychological outcomes. Bivariate correlations between each long-term outcome variable and each demographic, pre-treatment, treatment, and post-treatment variable were analyzed in order to identify potential predictors. Predictors that were significantly correlated with the dependent variable were entered into two linear or logistic regressions. The first, “liberal” analysis included motivation system. The second, “conservative” analysis included both motivation system and Favorable Departure Rating. In all cases, analyses were conducted only with those participants who had data for both Favorable Departure Rating and motivation system, to keep these analyses parallel. When no variable, including motivation system and Favorable Departure Rating, was significantly related to the dependent variable, no further analysis was done with that outcome measure. This was the case with the first psychological outcome measure, mental health index.

Recent substance abuse was regressed onto gender, the Property Offense Scale at admission, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 8). The liberal model (i.e., without Favorable Departure Rating) was significantly predictive of recent substance abuse, $\chi^2(3, n = 177) = 19.09, p < .001$. Men were more likely to report substance abuse within the 12 months preceding the follow-up study than women ($OR = 10.47$), and higher Property Offense Scale scores at baseline predicted substance abuse at follow up ($OR = 24.08$). Motivation system was not a significant predictor. The regression was repeated with Favorable Departure Rating

included as a predictor (the conservative test). The conservative model was significant, $\chi^2(4, n = 177) = 19.16, p < .001$, but neither motivation system nor Favorable Departure Rating was a significant predictor. One outlier was identified in these analyses.

Throughout this paper, whenever outliers were detected, the regressions were repeated without the outliers. Unless otherwise noted, there were no differences in outcome, and only the original regressions are reported.

Locus of control at follow-up was regressed onto minority status, the Aggressive Offense Scale at admission, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 9). The liberal model was significantly predictive of locus of control at follow-up, $R^2 = .064, F(3, 172) = 3.93, p = .01$. Higher Aggressive Offense Scale scores at baseline significantly predicted more external locus of control at follow up, and minority status was a marginally significant predictor ($p = .090$) with minorities reporting higher internal locus of control. Motivation system was also a marginally significant predictor ($p = .094$) with higher motivation system predicting more internal locus of control at follow-up. The conservative model was significant, $R^2 = .064, F(3, 172) = 3.93, p = .01$, but neither motivation system nor Favorable Departure Rating was a significant predictor. When controlling for success in treatment, motivation system was no longer a unique predictor.

The Satisfaction Index was regressed onto the Drug Scale at admission, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 10). The liberal model was not significantly predictive the Satisfaction Index, $R^2 = .024, F(2, 173) = 2.15, p = .12$. The conservative model was also not significant, $R^2 = .034, F(3, 172) = 2.01, p = .12$.

Happiness Rating at follow-up was regressed onto number of presenting problems, length of stay, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 11). The liberal model was significantly predictive of Happiness Rating at follow-up, $R^2 = .053$, $F(3, 173) = 3.23$, $p < .05$. However, no single predictor reached statistical significance. Number of presenting problems was a marginally significant predictor ($p = .076$), with fewer presenting problems at baseline predicting greater happiness at follow-up. Motivation system was also a marginally significant predictor ($p = .077$) with higher motivation system predicting more happiness at follow-up. The conservative model was significant, $R^2 = .074$, $F(4, 172) = 3.45$, $p < .01$. Not only was Favorable Departure a significantly positive predictor of happiness at follow-up, but the marginal significance of motivation system and number of presenting problems both disappeared in the conservative model. The variance in happiness was better accounted for by Favorable Departure Rating.

No predictors tested correlated significantly with whether or not the former resident was deceased by the time of the long-term follow-up; therefore, no further analyses were performed with this as the dependent variable.

Criminality. Arrested in the past 12 months was regressed onto minority status, age at follow-up, the Drug Scale at admission, number of presenting problems, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 12). The liberal model was significantly predictive of having been arrested, $\chi^2 (5, n = 177) = 17.91$, $p < .01$, though no single predictor reached significance individually. Minority status was marginally significant ($p = .096$) with minorities more likely to report an arrest ($OR = .51$). Motivation system was not a significant predictor. The

conservative model was significant, $\chi^2 (6, n = 177) = 18.57, p < .01$, and minority status remained a marginally significant predictor ($p = .093$). In this test, which controlled for Favorable Departure Rating, motivation system was a marginally significant predictor ($p = .094$) with higher motivation systems predicting less likelihood of arrests ($OR = .66$).

Perpetration Index was regressed onto age at admission, ROLES at admission, Locus of Control Scale at admission, length of stay, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 13). The liberal model was significantly predictive of Perpetration Index, $R^2 = .115, F(5, 154) = 4.01, p < .01$. However, no single predictor reached statistical significance. ROLES at admission was a marginally significant predictor ($p = .067$) with higher ROLES scores predicting fewer crimes perpetrated at follow-up. The Locus of Control Scale was also marginally significant with more internal locus of control predicting fewer perpetrations. Motivation system was not significant. The conservative model was significant, $R^2 = .115, F(6, 153) = 3.33, p < .01$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

Education. Highest grade completed was regressed onto Locus of Control Scale at admission, length of stay, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 14). The liberal model was significantly predictive of highest grade completed at follow-up, $R^2 = .103, F(3, 160) = 6.10, p < .001$. Locus of Control Scale was marginally significant ($p = .081$), with more internal locus of control predicting higher educational achievement, and length of stay was not a significant predictor. Motivation system was the only significant predictor in this model with higher motivation systems predicting higher educational achievement. The conservative model

was also significant, $R^2 = .136$, $F(4, 159) = 6.27$, $p < .001$. Not only was Favorable Departure a significantly positive predictor of educational attainment, but the significance of motivation system and marginal significance of the Locus of Control Scale both disappeared in the conservative model. The variance in highest grade completed was better accounted for by the Favorable Departure Rating.

In school in the past 12 months was regressed onto ROLES at admission, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 15). The liberal model was marginally predictive of having attended school in the past 12 months, $\chi^2 (2, n = 173) = 5.68$, $p = .059$. Higher ROLES scores significantly predicted greater likelihood of school attendance ($OR = 1.06$). Motivation system was not a significant predictor. The conservative model was not significant, $\chi^2 (3, n = 173) = 5.68$, $p = .13$.

Employment and earnings. Currently working was regressed onto minority status, ROLES at admission, length of stay, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 16). The liberal model was predictive of currently having a job, $\chi^2 (4, n = 173) = 14.84$, $p < .01$. Caucasians were significantly more likely to have a job ($OR = 2.28$), and higher ROLES scores significantly predicted greater likelihood of currently working ($OR = 1.05$). Length of stay was a marginal, positive predictor of having a job ($OR > 1.00$). Motivation system was not a significant predictor. The conservative model was significant, $\chi^2 (5, n = 173) = 15.27$, $p < .01$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

No predictors tested correlated significantly with job satisfaction; therefore, no further analyses were performed with this as the dependent variable.

Household income was regressed onto number of presenting problems, length of stay, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 17). The liberal model was significantly predictive of household income at follow-up, $R^2 = .096$, $F(3, 164) = 5.80$, $p < .001$. Number of presenting problems at baseline significantly predicted lower household income at follow up, and length of stay was not a significant predictor. Motivation system was a significant predictor with higher motivation systems predicting higher household income. The conservative model was significant, $R^2 = .097$, $F(4, 163) = 4.39$, $p < .01$, but neither motivation system nor Favorable Departure Rating was a significant predictor. When controlling for success in treatment, motivation system was no longer a unique predictor.

No predictors tested correlated significantly with inability to pay important bills within the last 12 months; therefore, no further analyses were performed with this as the dependent variable.

Military service was regressed onto Happiness Rating at admission, length of stay, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 18). The liberal model was significantly predictive of military service, $\chi^2(3, n = 175) = 26.52$, $p < .001$. Happiness Rating was a significant predictor with those reporting greater happiness at baseline *less* likely to have served in the military at follow-up ($OR = .49$). Length of stay was not significant, but motivation system was. Higher motivation systems predicted greater likelihood of military service ($OR = 1.91$). The conservative model was also significant, $\chi^2(4, n = 175) = 27.07$, $p < .001$. Favorable Departure Rating was not a significant predictor, and motivation system was now a

marginally significant predictor ($p = .057$). When controlling for success in treatment, motivation system moved from being a significant to a marginally significant predictor.

Social systems. Currently living with a spouse or partner was regressed onto age at follow-up, Happiness Rating at admission, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 19). The liberal model was predictive of living with a partner at follow-up, $\chi^2(3, n = 171) = 9.46, p < .05$. Older respondents were significantly more likely to live with a partner ($OR = 1.26$). Happiness Rating and motivation system were not significant predictors. The conservative model was significant, $\chi^2(4, n = 171) = 9.95, p < .05$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

No predictors tested correlated significantly with the Child Involvement Index or the Social Involvement Count; therefore, no further analyses were performed with these as dependent variables.

Spirituality was regressed onto the Aggressive Offense Scale at admission, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 20). The liberal model was not significantly predictive the spirituality at follow-up, $R^2 = .023, F(2, 172) = 2.05, p = .13$. The conservative model was not significant either, $R^2 = .028, F(3, 171) = 1.66, p = .18$.

The GBT Index (the average of 5 items on the follow-up survey measuring respondents' evaluation of GBT) was regressed onto minority status, length of stay, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 21). The liberal model was significantly predictive of the evaluation of GBT at follow-up, $R^2 = .064, F(3, 171) = 3.88, p < .05$. Length of stay significantly predicted the

evaluation of GBT at follow up with longer stays associated with higher ratings of GBT. Minority status and motivation system were not significant predictors. The conservative model was also significant, $R^2 = .079$, $F(4, 170) = 3.66$, $p < .01$. Length of stay continued to positively and significantly predict GBT Index. Minority status was marginally significant ($p = .093$) with minority respondents rating GBT higher than Caucasian respondents did. Motivation system was not significant, but Favorable Departure Rating was marginally significant ($p = .092$) with higher departure ratings predicting a better evaluation of GBT at follow-up.

Secondary Analyses

In order to identify which demographic and pre-treatment variables predicted outcomes during and immediately post-treatment, a series of exploratory analyses regressed motivation system at departure, Favorable Departure Rating, and length of stay onto potential predictors. The predictors entered into a given analysis were chosen based on their bivariate correlations with the dependent variables.

Motivation system at departure. Motivation system at departure was regressed onto gender, minority status, age at admission, ROLES at admission, number of prior placements, the Aggressive Offense Scale at admission, the Property Offense Scale at admission, and the number of presenting problems (see Table 22). The linear regression model was significant, $R^2 = .100$, $F(8, 344) = 4.78$, $p < .001$. Age at admission was a significantly positive predictor of motivation system at departure, and number of presenting problems was a negative predictor. Gender was a marginally significant predictor ($p = .071$), with girls attaining higher motivation systems. Minority status was also marginally significant ($p = .084$), with Caucasians attaining higher motivation

systems. In this model, ROLES at admission, number of prior placements, the Aggressive Offense Scale, and the Property Offense Scale were not significant predictors.

Favorable Departure Rating. Favorable Departure Rating was regressed onto gender, minority status, age at admission, ROLES at admission, number of prior placements, the Aggressive Offense Scale, and the number of presenting problems (see Table 23). The linear regression model was significant, $R^2 = .110$, $F(7, 369) = 6.49$, $p < .001$. Gender was a significant predictor of Favorable Departure with girls attaining higher ratings. Age at admission was also a significantly positive predictor of Favorable Departure Rating, and number of presenting problems was a negative predictor. Minority status was marginally significant ($p = .097$), with Caucasians attaining higher Favorable Departure Ratings. Number of prior placements was marginally ($p = .096$) negatively predictive. In this model, ROLES at admission and the Aggressive Offense Scale were not significant predictors.

Length of stay. Length of stay was regressed onto gender, African American status (which was more strongly correlated with length of stay than minority status), age at admission, the Property Offense Scale, the Drug Scale, estimated GPA, the Consumer Scale, and the Happiness Rating (see Table 24). The linear regression model was significant, $R^2 = .146$, $F(8, 420) = 9.00$, $p < .001$. The only significant predictor was age at admission, which was a negative predictor of length of stay. African American status was marginally significant ($p = .086$), with African Americans staying longer. None of the other predictors were significant in this model. To reduce shared variance between the predictors, gender, the Property Offense Scale, estimated GPA, and Happiness Rating were removed one-by-one until all of the predictors in the model were significant. The

final model was significant, $R^2 = .137$, $F(4, 461) = 18.258$, $p < .001$, with a loss of only 0.9% variance accounted for compared with the first model. In this model, age at admission and the Drug Scale were significantly negative predictors of length of stay, and African American status and the Consumer Scale were significantly positive predictors of length of stay.

5-Year Follow-Up

Preliminary Analyses

Descriptive statistics for the demographics were reported in the method section except for IQ. Means, standard deviations, and sample sizes for IQ and all other variables are reported in Table 25.

Long-term follow-up survey responders were compared to nonresponders in the same cohort on all demographic, baseline, treatment, and post-treatment variables to determine their representativeness of the cohort. Chi-square tests were used for categorical and dichotomous variables, and MANOVAs were used for groups of continuous variables. The groups differed significantly in gender, $\chi^2(1, n = 214) = 18.25$, $p < .001$ (see Table 26). Women were much more likely to respond to the survey than men. Ethnicity was also significantly different across responder groups, $\chi^2(3, n = 214) = 12.32$, $p < .01$ (see Table 27). Hispanics were less likely to respond, and individuals in the “other” category were more likely to respond to the survey than African Americans or Caucasians. The responders and nonresponders were equivalent on all other measures, including all pre-treatment, treatment, and post-treatment variables. The groups did not differ significantly on motivation system, $F(1, 210) = 2.32$, *n.s.*

Primary Analyses

Psychological outcomes. Liberal and conservative analyses for each outcome variable in the 5-year follow-up were conducted with the same method used for the 16-year follow-up. When no variable, including motivation system and Favorable Departure Rating, was significantly related to the dependent variable, no further analysis was done with that outcome measure.

Mental health index was regressed onto number of internalizing diagnoses at admission, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 28). The liberal model was significantly predictive of mental health at follow-up, $R^2 = .121$, $F(2, 121) = 8.32$, $p < .001$. Number of internalizing diagnoses at baseline significantly and negatively predicted mental health at follow up. This relationship is not surprising given that the mental health index measures internalizing symptoms. Motivation system was not a significant predictor. The conservative model was significant, $R^2 = .125$, $F(3, 120) = 5.74$, $p < .01$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

Recent substance abuse was regressed onto age at departure, length of stay, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 29). The liberal model was significantly predictive of recent substance abuse, $\chi^2(3, n = 131) = 9.00$, $p < .05$. No single predictor was significant, but age at departure was marginally significant ($p = .075$). Youth who left GBT when they were relatively older (and were, therefore, relatively older at follow-up) were less likely than younger responders to report substance abuse within the 12 months preceding the follow-up study ($OR = 0.76$). Motivation system and length of stay were not significant predictors in this

model. The conservative model was significant, $\chi^2(4, n = 131) = 9.67, p < .05$, but neither motivation system nor Favorable Departure Rating was a significant predictor. One outlier was identified in these analyses, and the regressions were repeated without the outlier. Removing the outlier did not change the predictive strengths of motivation system or Favorable Departure, though length of stay became a significant predictor, and age at departure was no longer marginally significant.

Friends' substance use was regressed onto gender, Hispanic status, ever arrested at admission, any substance abuse diagnosis at admission, presenting problems percent improved upon, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 30). The liberal model was significantly predictive of friends' substance use at follow-up, $R^2 = .200, F(6, 97) = 4.05, p < .01$. Ever arrested at baseline and any substance abuse diagnosis at baseline significantly and positively predicted friends' substance use at follow-up. Gender was a marginally significant predictor ($p = .095$) with men reporting more substance use among their friends than women reported. Motivation system was not a significant predictor. The conservative model was significant, $R^2 = .200, F(7, 96) = 3.44, p < .01$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

Criminality. Whether the respondent perpetrated any crime in the past 12 months was regressed onto age at departure, ever arrested, ROLES at departure, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 31). The liberal model was significantly predictive of perpetration, $\chi^2(4, n = 109) = 20.08, p < .001$. Age at departure was significant, with youth who left GBT when they were relatively older (and were, therefore, relatively older at follow-up) less likely than

younger respondents to report perpetration within the 12 months preceding the follow-up study ($OR = 0.65$). Not surprisingly, having been arrested before arriving at GBT was a positive predictor of perpetrating a crime at follow-up ($OR = 2.90$). ROLES at departure was marginally significant, with higher ROLES scores predicting a lower likelihood of perpetration ($OR = 0.95$). Motivation system was not a significant predictor. The conservative model was significant, $\chi^2(5, n = 109) = 20.10, p < .01$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

Whether the respondent reported an arrest in the past 12 months was regressed onto age at departure, presenting problems percentage improved, ROLES at departure, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 32). The liberal model was significantly predictive of arrest in the year before follow-up, $\chi^2(4, n = 128) = 12.59, p < .05$. Age at departure was the only significant predictor in the model, with youth who left GBT when they were relatively older (and were, therefore, relatively older at follow-up) less likely than younger respondents to report an arrest ($OR = 0.69$). Motivation system was not a significant predictor. The conservative model was significant, $\chi^2(5, n = 128) = 12.62, p < .05$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

Current incarceration was regressed onto gender, presenting problems percent improved, ROLES at departure, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 33). The liberal model was significantly predictive of incarceration, $\chi^2(4, n = 128) = 21.13, p < .001$. Men were more likely to be currently incarcerated at follow-up study than women ($OR = 15.51$). ROLES at departure was marginally predictive, with higher ROLES scores predicting less likelihood for

incarceration ($OR = .91$). Motivation system was not a significant predictor. The conservative model was significant, $\chi^2 (5, n = 128) = 22.47, p < .001$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

Education. Highest grade completed was regressed onto age at departure, IQ, length of stay, presenting problems percent improved upon, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 34). The liberal model was significantly predictive of educational attainment at follow-up, $R^2 = .171, F(5, 82) = 3.39, p < .01$. Age at departure was the only significant predictor, with youth who left GBT when they were relatively older (and were, therefore, relatively older at follow-up) having completed more years of education. Motivation system was not a significant predictor. The conservative model was significant, $R^2 = .172, F(6, 81) = 2.80, p < .05$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

Whether the respondent reported attending school in the past 12 months was regressed onto ever arrested at admission, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 35). The liberal model was marginally significantly predictive of school enrollment, $\chi^2 (2, n = 110) = 4.75, p = .093$. Having been arrested before arriving at GBT was marginally predictive ($p = .065$) of a lower likelihood of school attendance ($OR = .45$). Motivation system was not a significant predictor. The conservative model was not significant, $\chi^2 (3, n = 110) = 5.00, p = .17$.

Employment and earnings. No predictors tested correlated significantly with whether or not the former resident worked during the week before the follow-up survey was completed; therefore, no further analyses were performed with this as the dependent variable.

Household income was regressed onto number of family problems, number of internalizing diagnoses, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 36). The liberal model was significantly predictive of household income at follow-up, $R^2 = .074$, $F(3, 107) = 2.83$, $p < .05$. Number of internalizing diagnoses at baseline negatively predicted household income at follow-up. Number of family problems was marginally significantly predictive ($p = .071$) of household income, with fewer family problems at baseline predicting higher incomes at follow-up. Motivation system was not a significant predictor. The conservative model was marginally significant, $R^2 = .075$, $F(4, 106) = 2.13$, $p = .082$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

Military service was regressed onto gender, age at departure, ROLES at admission, ever arrested, number of externalizing diagnoses, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 37). The liberal model was significantly predictive of military service, $\chi^2 (6, n = 107) = 37.89$, $p < .001$. Having an arrest history before arriving at GBT predicted a lower likelihood of military service at follow-up ($OR = .03$). Interestingly, number of externalizing diagnoses at baseline *positively* predicted military service ($OR = 17.64$). Gender was a marginal predictor ($p = .060$), with men more likely to have served in the military than women. Motivation system was not a significant predictor in this model. The conservative model was also significant, $\chi^2 (7, n = 107) = 37.93$, $p < .001$, but neither motivation system nor Favorable Departure Rating was a significant predictor. Two outliers were identified in these analyses, and the regressions were repeated without the outliers. However, this yielded an infinite odds ratio for gender, number of externalizing diagnoses, and motivation

system, indicating that there was no variability in those predictor variables in one of the dependent variable categories.

Since motivation system was the variable of interest in this study, the motivation system of each respondent with military experience was examined. Of the 12 reporting military service, 10 had been on Achievement at departure, and 2 were on Weekly. (These latter two were the outliers.) No youth who departed on Daily or Subsystem reported military service. In order to determine if motivation system was systematically different between the military and nonmilitary groups, a t-test was run comparing those with military service to those without on motivation system level at departure. Respondents with military experience left on significantly higher motivation systems (mean = 2.83, $SD = .39$) than those without (mean = 2.08, $SD = 1.06$), $t(128) = -2.43$, $p < .05$. Since the variances were unequal, a t-test with equal variances not assumed was run. The outcome was the same, $t(32.07) = -5.03$, $p < .001$. To compare this to Favorable Departure Rating, a second t-test was run with military service as the grouping variable and Favorable Departure Rating as the dependent variable. The difference between military (mean = 22.75, $SD = 3.70$) and nonmilitary (mean = 20.53, $SD = 5.62$) groups on Favorable Departure Rating was not significant, $t(128) = -1.34$, $p = .183$. To see whether motivation system's prediction of military service was robust even when controlling for Favorable Departure, an ANCOVA was run. The grouping variable was military service, and the dependent variable was motivation system. Favorable Departure was added as a covariate. This method was used to partial out the variance in the motivation system variable accounted for by Favorable Departure. The difference between the groups remained significant, $F(1, 127) = 4.49$, $p < .05$, suggesting that motivation system

predicts military service over and above the variance associated with Favorable Departure Rating. (No outliers were found in this analysis, Cook's maximum distance = .18.)

Social systems. Currently living with a spouse or partner was regressed onto minority status, ever arrested, any substance abuse diagnosis, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 38). The liberal model was significantly predictive of marriage or cohabitation at follow-up, $\chi^2(4, n = 107) = 13.94, p < .01$. Any substance abuse diagnosis at baseline was the only significant predictor in this model; those with a substance abuse diagnosis at baseline were more likely to be married or living with a partner at follow-up ($OR = 2.74$). Motivation system was not a significant predictor. The conservative model was significant, $\chi^2(5, n = 107) = 14.43, p < .05$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

Closeness to one's spouse or partner was regressed onto "other" ethnicity, age at departure, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 39). The liberal model was significantly predictive of closeness to one's spouse or partner at follow-up, $R^2 = .248, F(3, 53) = 5.82, p < .01$. Respondents who were not Caucasian, African American, or Hispanic reported significantly less closeness. Age at departure was a significant predictor, with youth who left GBT when they were relatively older (and were, therefore, relatively older at follow-up) reporting closer relationships with their spouses and partners. Motivation system was not a significant predictor. The conservative model was significant, $R^2 = .249, F(4, 52) = 4.31, p < .01$, but neither motivation system nor Favorable Departure Rating was a significant

predictor. Two outliers were identified in these analyses, and the regressions were repeated without the outliers (see Table 39). Without the outliers, neither the liberal test, $R^2 = .052$, $F(3, 51) = .94$, *n.s.*, nor the conservative test, $R^2 = .053$, $F(4, 50) = .69$, *n.s.*, was significant.

Conflict with one's spouse or partner was regressed onto number of family problems, total number of diagnoses, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 40). The liberal model was significantly predictive of conflict with one's spouse or partner at follow-up, $R^2 = .164$, $F(3, 48) = 3.15$, $p < .05$. No individual predictor was significant, but number of family problems and total number of diagnoses were each marginally significant predictors (p 's = .085, .053) with higher scores on each predicting more conflict. Motivation system was not a significant predictor. The conservative model was marginally significant, $R^2 = .172$, $F(4, 47) = 2.45$, $p = .059$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

Average child involvement was regressed onto gender, number of family problems, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 41). The liberal model was significantly predictive of child involvement at follow-up, $R^2 = .149$, $F(3, 57) = 3.33$, $p < .05$. Women were significantly more involved with their children than men were. Number of family problems was marginally predictive ($p = .072$) of child involvement, with, counter-intuitively, higher numbers of family problems at baseline predicting more frequent involvement with kids at follow-up. Motivation system was not a significant predictor. The conservative model

was significant, $R^2 = .157$, $F(4, 56) = 2.62$, $p < .05$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

Having support from at least one parent was regressed onto number of family problems at admission, ever arrested, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 42). The liberal model was significantly predictive of parental support at follow-up, $\chi^2(3, n = 106) = 10.63$, $p < .05$. A greater number of family problems at baseline predicted a lower likelihood of having a supportive parent at follow-up ($OR = .85$). Having an arrest history before arriving at GBT was, counter-intuitively, positively predictive of having a supportive parent ($OR = 4.32$). Motivation system was not a significant predictor. The conservative model was significant, $\chi^2(4, n = 106) = 10.80$, $p < .05$, but neither motivation system nor Favorable Departure Rating was a significant predictor. One outlier was identified in these analyses, and the regressions were repeated without the outlier (see Table 42). The liberal model was again significant, $\chi^2(3, n = 105) = 9.02$, $p < .05$. Arrest history continued to positively predict parental support ($OR = 5.20$), but number of family problems was no longer significantly related. The conservative model was significant, $\chi^2(4, n = 105) = 10.10$, $p < .05$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

The importance of religion to the respondent was regressed onto IQ, motivation system, and (in the conservative analysis) Favorable Departure Rating (see Table 43). The liberal model was significantly predictive of the importance of religion at follow-up, $R^2 = .149$, $F(2, 78) = 6.49$, $p < .01$. Higher IQs were associated with *lower* importance of religion scores. Motivation system was not a significant predictor. The conservative

model was significant, $R^2 = .171$, $F(3, 77) = 5.28$, $p < .01$, but neither motivation system nor Favorable Departure Rating was a significant predictor.

Secondary Analyses

In order to identify which demographic and pre-treatment variables predicted outcomes during and immediately post-treatment, a series of exploratory analyses regressed motivation system at departure, Favorable Departure Rating, the Departure Success Scale, length of stay, and ROLES at departure onto potential predictors. The predictors entered into a given analysis were chosen based on their bivariate correlations with the dependent variables.

Motivation system at departure. Out of 10 potential predictors, only gender was significantly correlated with motivation system at departure. IQ was marginally correlated ($p = .053$). Therefore, motivation system was regressed onto gender in the first step, IQ was added in the second step, and the other 8 predictors were added in the third step to see if any of them improved the strength of the model (see Table 44). The first step was significant, $R^2 = .052$, $F(1, 137) = 7.58$, $p < .01$. Girls ended on significantly higher motivation systems than boys. The second step was marginally significant, $\Delta R^2 = .025$, $F(1, 136) = 3.75$, $p = .055$. In that model, gender remained a significant predictor, and IQ was marginally significant ($p = .055$), with higher IQs predicting higher motivation system levels. When the remaining 8 predictors were entered in step 3, the additional variance accounted for was not significant, $\Delta R^2 = .057$, $F(8, 128) = 1.06$, $n.s.$

Favorable Departure Rating. Out of 10 potential predictors, only gender and number of formal prior placements were significantly correlated with Favorable Departure Rating. No predictor was marginally correlated. Therefore, Favorable

Departure was regressed onto gender and number of formal prior placements in the first step, and the other 8 predictors were added in the second step to see if any of them improved the strength of the model (see Table 45). The first step was significant, $R^2 = .121$, $F(2, 138) = 9.46$, $p < .001$. Girls received significantly higher Favorable Departure Ratings than boys. A higher number of formal prior placements was a negative predictor of Favorable Departure. The second step did not add significantly to the variance accounted for in the first step, $\Delta R^2 = .050$, $F(8, 130) = .986$, $n.s.$

Departure Success Scale. Though it was not included in the primary analyses in this paper because of its poor internal reliability and high correspondence with Favorable Departure Rating, testing for the predictors of success on the Departure Success Scale was interesting as a replication of a previous study on the baseline predictors of this scale (Larzelere, Smith, et al., 2001). Not surprisingly, the outcome of this analysis was virtually identical to the one done with Favorable Departure Rating. Out of 10 potential predictors, only gender and number of formal prior placements were significantly correlated with the Departure Success Scale. No predictor was marginally correlated. Therefore, Departure Success was regressed onto gender and number of formal prior placements in the first step, and the other 8 predictors were added in the second step to see if any of them improved the strength of the model (see Table 46). The first step was significant, $R^2 = .124$, $F(2, 138) = 9.74$, $p < .001$. Girls received significantly higher Departure Success Scale scores than boys. A higher number of formal prior placements was a negative predictor of Departure Success. The second step did not add significantly to the variance accounted for in the first step, $\Delta R^2 = .055$, $F(8, 130) = 1.09$, $n.s.$

Length of stay. Length of stay was regressed onto Hispanic status, age at admission, number of formal prior placements, number of presenting problems, and any substance abuse diagnosis (see Table 47). The linear regression model was significant, $R^2 = .075$, $F(5, 169) = 2.73$, $p < .05$. The only significant predictor was age at admission, which was a negative predictor of length of stay. Because the goal of the program was for youth to stay until high school graduation, it is not surprising that children who arrived at younger ages stayed longer. None of the other predictors were significant in this model.

ROLES at departure. Out of 10 potential predictors, only number of presenting problems, ROLES at admission, and number of formal prior placements were significantly correlated with ROLES at departure. Therefore, ROLES at departure was regressed onto those three predictors in the first step, and the other 7 predictors were added in the second step to see if any of them improved the strength of the model (see Table 48). The first step was significant, $R^2 = .077$, $F(3, 137) = 3.80$, $p < .05$. ROLES at admission positively predicted ROLES at departure, and number of presenting problems and number of formal prior placements both negatively predicted ROLES at departure. The second step was marginally significant, $\Delta R^2 = .084$, $F(7, 130) = 1.85$, $p = .083$. In the model with all 10 predictors, ROLES at admission and number of formal prior placements both remained significant, but number of presenting problems was no longer significant. Gender was also significant in this model, with girls achieving higher ROLES scores at departure than boys.

Discussion

Predictors of Long-Term Follow-Up Outcomes

Based on the generalization and internalization literatures, it was hypothesized that youth who ended their treatment at GBT on higher motivation system levels would have better outcomes at 5-16-year follow-up. Two parallel sets of regression analyses were run with motivation system as the independent variable and various long-term outcomes derived from two long-term follow-up surveys as the dependent variables. The “liberal” analyses did not control for general success in treatment, and the “conservative” analyses included a proxy of general success in the treatment program (Favorable Departure Rating). In the 16-year follow-up survey, liberal analyses suggested that leaving GBT from a higher motivation system level significantly and positively predicted highest grade completed, household income at follow-up, and military service. Motivation system also marginally predicted more internal locus of control and greater happiness at follow-up when not controlling for Favorable Departure. In the conservative analyses, Favorable Departure was a more reliable predictor of happiness and highest grade completed at follow-up than motivation system was. Furthermore, motivation system’s significant prediction of household income and marginally significant prediction of locus of control disappeared in the conservative analyses. After being strongly predictive in the liberal analysis, motivation system remained a marginal predictor of military service in the conservative analysis. However, it was a marginal negative predictor of arrests in the past 12 months only when controlling for Favorable Departure. Neither measure significantly predicted outcomes on 13 other long-term outcome measures (for a summary, see Table 49). In sum, motivation system was a significant or

marginally significant predictor of six outcome variables in either the liberal or conservative analyses, but, when controlling for Favorable Departure, it was only a marginal predictor of two of them (i.e., military service, arrested in the past 12 months). In the 5-year follow-up survey, motivation system was a strong, positive predictor of military service, even when controlling for Favorable Departure. However, neither motivation system nor Favorable Departure was a significant predictor of any of the other 16 long-term outcome measures (for a summary, see Table 50).

Other demographic, baseline, and post-treatment variables were entered into the analyses when they correlated significantly with the outcome measures in order to control for extraneous variance. Using this pragmatic method, only one of these other variables emerged as a frequent predictor of long-term outcome in the 16-year follow-up study. Minority status was a significant predictor of one outcome (whether currently working at follow-up) and a marginal predictor of three outcomes (locus of control, whether arrested in the past 12 months, and evaluation of GBT); interestingly, being a member of a minority group predicted positive outcomes in two of these and negative outcomes in the other two. All of the other variables were predictors in two or fewer models. Compared to most other variables, motivation system was a relatively frequent predictor. In the 5-year follow-up study, four variables emerged as relatively frequent predictors. Out of 17 follow-up outcome measures, age at departure and arrest history at admission each significantly predicted five outcomes and marginally predicted one. Gender predicted two outcomes significantly and two marginally. Number of family problems at admission significantly predicted one and marginally predicted three outcomes. There

was no discernable pattern as to when these predictors overlapped or appeared independently.

Predictors. It should not be surprising that it was difficult to predict long-term outcomes of residential treatment. While I found no studies that attempted to look at predictors of follow-up residential treatment outcomes, even prediction models of post-treatment outcome account for a relatively small amount of variance. In a sample from GBT, Larzelere, Smith, and colleagues (2001) were able to account for 15% of variance in their models of departure success based on demographic and baseline variables. It should be expected that the strength of predictions over 5-16-year post-treatment periods would be even weaker. Interestingly, the specific predictors (and non-predictors) in this study were unanticipated. First, one would expect that baseline and outcome variables in the same domain would be related. For example, substance use at follow-up was expected to be best predicted by substance involvement (i.e., Drug Scale in the 16-year, any substance abuse diagnosis on the DISC in the 5-year) at admission. However, these substance-related pre-treatment variables were not related to substance abuse at follow-up, and a similar pattern was frequently observed for other conceptually related pre-treatment and follow-up variables. The direction of the predictors was generally expected (i.e., better pre-treatment functioning predicted better follow-up outcomes, and vice versa), and a post-hoc explanation could be formulated for each significant predictor, but it would have been impossible to identify the specific predictors for most of the dependent variables *a priori*. Second, previous research at GBT has found that length of stay is an important predictor of follow-up success in studies following up residents after 4-16 years (e.g., Huefner et al., 2007; Thompson et al., 2005; Thompson et al., 1996), but

it was not a frequent predictor in this study. That neither length of stay nor the Favorable Departure Rating predicted a single outcome in the 5-year follow-up was especially surprising. After treating an adolescent in residence for two years, the typical length of stay, one would expect that the youth's success in the program and the clinical judgments made about how well prepared the child was to leave treatment (both of which were important factors in the Favorable Departure Rating) would be predictive of at least some of the 5-year outcomes.

16-year follow-up. In terms of the primary hypothesis, that leaving on a higher motivation system would predict better long-term outcomes, the results were, at best, suggestive. In the 16-year follow-up, motivation system was a significant or marginally significant predictor of six outcomes out of 20, an impressive number for a study with a 16-year gap between measurement points. However, it was only a marginally significant predictor of two outcomes when controlling for Favorable Departure. Favorable Departure Rating uniquely contributed to three models, happiness, highest grade completed, and evaluation of GBT, the latter marginally. (While evaluation of GBT is an interesting variable about satisfaction with treatment, it is probably not comparable to the psychological and behavioral items as a measure of treatment outcome.) This suggests that, perhaps, motivation system is as strong as Favorable Departure as a predictor of long-term outcome, even though the latter takes the former (as well as a lot of other data about the child) into account. However, Favorable Departure achieved significance or marginal significance in its three models even after controlling for motivation system. In two of those cases, motivation system was a significant or marginal predictor in the liberal model, but its influence was erased in the conservative model with the inclusion of

Favorable Departure. In contrast, four of the five cases (locus of control, happiness, highest grade completed, and household income) in which motivation system was a significant or marginal predictor in the liberal model, the addition of Favorable Departure caused motivation system's influence to disappear because of the multicollinearity it shared with Favorable Departure. This suggests that motivation system was not a unique predictor of those outcomes when controlling for Favorable Departure. Interestingly, in predicting whether or not the follow-up respondent had been arrested in the past 12 months, motivation system only approached marginal significance ($p = .104$) in the liberal model but attained marginal significance ($p = .94$) in the conservative model. This suggests that motivation system negatively predicted arrests at long-term follow-up when controlling for Favorable Departure. Military service, which is discussed below, was the only outcome for which motivation system was an unequivocal, unique predictor.

The simplest explanation for the lack of consistent predictors, including motivation system and Favorable Departure, in the 16-year follow-up was the length of the follow-up. Even if the GBT experience was very influential, the environmental and idiographic influences on behavior over the 16 years since treatment likely contributed an enormous amount of variance. That motivation system came up more frequently than any other predictor may suggest, rather tentatively, that it has a unique predictive role in outcome. On the other hand, because this was an exploratory study and the variance accounted for in the models and by individual predictors was generally small, no corrections were made to control for Type I error. It is possible that these findings are the result of chance.

5-year follow-up. The near-complete lack of significant results in the 5-year follow-up with regard to motivation system, Favorable Departure, and length of stay was surprising. The latter two variables have been found to be predictors in past GBT studies (e.g., Larzelere, Smith, et al., 2001; Thompson et al., 2005). In the present study, no treatment variable was predictive, and only one post-treatment variable (ROLES at departure) was marginally predictive of any long-term outcome. On the other hand, demographic (specifically, gender and age at departure) and pre-treatment variables (specifically, two “history” variables, number of family problems and arrest history) were far more consistently predictive in this sample than in the 16-year group. How is it possible that success in the program, across measures, predicted so little at 5-year follow-up? Why were treatment and post-treatment variables better predictors in the longer-term follow-up? And why did demographics and pre-treatment history variables emerge as more consistent in the 5-year group than in the 16-year group? Perhaps measuring outcomes of treatment at 5 years is premature. Former residents were just 17-24 years old at follow-up. The late teens to early 20’s is a tumultuous time of constant change. Young adults are going to college, working, and living independently for the first time. They are also forging independent identities as adults and beginning to develop long-term romantic partnerships. Some negative behaviors, like binge drinking, are far more common during this developmental period than they are later. Perhaps this group is too “unsettled” for treatment outcomes to yet be reliable, and this variance is reflected in the data here. Perhaps predictions to the 16-year follow-up group, though not strong, were more reliable because this group was older, 27-37 years old, and, therefore, more “settled.” Perhaps the history variables in the 5-year group (i.e., number of family

problems and arrest history) were more influential than the history variables in the 16-year group because the respondents were younger, so their pre-treatment histories were more recent. These younger respondents were probably also in closer contact with the environments that influenced them before they arrived at GBT and that contributed to their pre-treatment risk factors. Furthermore, the age range represented by this group is very wide. Seventeen-year-olds and 24-year-olds lead very different lives and some outcomes, such as living with a spouse or partner, mean very different things at different ends of that spectrum. One might consider a 24-year-old living with a partner in a committed relationship a good outcome. On the other hand, an 18-year-old living with a partner suggests that the person is not in college nor does he or she live with parents. At this age, living with a partner might be a marker for a negative outcome. Perhaps age at departure, which varied closely with age at follow-up, was a frequent predictor because it captured some of the variability in maturity and life-stage across this sample. Gender was also a common predictor with females having consistently better outcomes than males. Perhaps this is because young women in this age group responded better to treatment, or perhaps they were simply more mature than their male counterparts.

Military service. The most striking finding related to the hypothesis was the consistent positive prediction of military service by motivation system, even when controlling for Favorable Departure, in both samples. (Motivation system was a marginally significant predictor when controlling for Favorable Departure in the 16-year follow-up.) Military service was included as a dependent variable in this study because it may represent a good outcome for graduates of GBT. The military provides great opportunities for these disadvantaged young adults; it prevents them from having to live

in the often-dysfunctional environments that they grew up in, and it is a natural intermediate step between the rigorous structure of GBT and independent civilian life. Leaving GBT on a higher motivation system was a strong, unique predictor of military service, even when controlling for Favorable Departure, and this was replicated across both samples. This is a suggestive finding for motivation system, but it is also ironic because the Achievement System aims to reduce the influence of a token economy on motivation, and a token economy (though not as tightly controlled) is the primary method of motivation in the military. It seems that those who ended up on Achievement were more likely to seek (or be accepted into) the military's structured environment. Perhaps this is a coincidence, and joining the military does not represent a desire to return to a structured behavioral system but, rather, a desire to better oneself, have greater opportunities, and learn marketable skills. In this case, experience in Achievement may have helped these graduates to make a responsible decision, despite its substantial costs and risks. On the other hand, this relationship could be a reflection of certain youth's superior abilities to conform to a strict behavioral system. Perhaps youth who excelled on GBT's token systems and, therefore, reached its pinnacle, were more inclined to join the military because they expected to excel in (or at least tolerate) the military's restrictive discipline.

The issue of military service raises another caveat. Though 21% of 16-year follow-up respondents reported military service, and 36% of those who stayed at GBT for at least 18 months in that cohort reported serving, only 9% of 5-year follow-up respondents fell into this category. This is despite the fact that most of these young adults were in the target age range for military service during the wars in Afghanistan and

Iraq, when recruitment has been particularly intense. (The military was far less active during the prime recruitment years of the 16-year follow-up cohort.) This could be accounted for by the younger age of the 5-year group – they had less time to join – or its greater balance between males and females. However, the proportion of respondents with military experience may be artificially low because the U.S. Defense Department did not allow GBT to look for or attempt to contact active duty soldiers that GBT did not already have contact information for. Based on previous rates of enlistment, GBT staff suspect this may have led to the exclusion of a substantial number of former residents from the follow-up survey (J. C. Huefner, personal communication, August 2006). If there was, in fact, a bias in the sampling, it may also account for the weak predictions of motivation system and Favorable Departure Rating in the outcome measures because many of the youth with good outcomes were systematically excluded. Given that more men enlist in the military than women, this sampling problem may partially account for the systematic differences in survey response rate and long-term outcomes favoring females.

Implications and limitations. Because the outcomes of this study were, at best, tentative, implications are limited. While the general theory described in the introduction is based on a careful exegesis of the literature, the outcomes of this study were not strong enough to support it without extensive further study and replication. Given the results, it is certainly possible that the hypothesis is false. Nevertheless, though the hypothesis was not strongly supported, it was not disproven, either. This was a relatively conservative test of the hypothesis for a number of reasons. First, attempting to predict outcomes 5-16 years *after* treatment is challenging because the follow-up point is so long after treatment.

Second, requiring motivation system to predict unique variance while controlling for Favorable Departure Rating is a conservative test because the two variables share so much variance and because Favorable Departure relies, in large part, on the youth's treatment progress. The motivation system a youth achieves is an important measure of treatment progress, and it influences Favorable Departure. Favorable Departure also has the benefit of taking into account nuanced perceptions of the child's behavior as well as predictions about future success based on the youth's post-treatment plan, supports, and environment. Motivation system, which has just 4 levels, would have to be very robust to overcome the information contained in Favorable Departure, which was a 13-point scale in the 16-year study and a 25-point scale in the 5-year study. Because of this, the liberal set of analyses, which suggested that motivation system predicted three follow-up outcomes significantly and two marginally in the 16-year follow-up should be considered seriously, at least as a counterbalance to the analyses that included Favorable Departure and were, therefore, quite conservative. Third, these were the best samples available within which to test the hypothesis because GBT has unusually long-term follow-up data and because the Achievement System has many of the features proposed in the introduction to promote internalization and generalization. However, as stated in the method section, the Achievement System is far from a perfect test of the model. It contains a token system rather than freeing itself completely from artificial reinforcement, self-regulation is valued but not operationalized clearly and systematically in the manual, and natural and logical consequences are also not defined in the manual. In this author's informal discussions with youth and family-teachers at GBT, there is some variation in how the term "natural and logical consequences" is interpreted. Some

of the examples they offered met the definition contained in this paper, but some did not. It seemed that many defined “natural and logical consequences” as punishments that are not used in GBT’s other motivation systems and that, perhaps, are more similar to the punishments given by typical parents to their teenage children. However, this definition includes punishments that are not logically related to the problem behavior and that do not attempt to mimic the natural consequences in the environment beyond any adult’s intentional involvement. Furthermore, breaches of conduct in the Achievement System can be punished through a return to the Weekly System or placement in a more restrictive environment within GBT. Therefore, there remain strong external contingencies on behavior that do not mimic the contingencies in the natural and social worlds. As far as existing systems go, GBT’s Achievement System is a remarkably good test of the theory, but it was not designed strictly based on the theory, so there are meaningful differences.

Though this was likely a conservative test of the hypothesis, the significant findings should still be interpreted cautiously because of the high likelihood of Type I error. In order to remain sensitive to the inherently weak relationships in such long-term follow-up data, no Type I error corrections were made, and marginal findings were reported. This raised the likelihood of Type I error substantially. Therefore, the findings should be viewed as merely preliminary; they require replication under better controlled or higher powered conditions.

Caution is also warranted in interpreting the findings of which variables, beyond motivation system and Favorable Departure Rating, were significant predictors of various outcomes. Unlike the variables associated with the hypothesis, these variables were not entered into the analyses on theoretical grounds. They were entered based on their

bivariate correlations with the outcome variables of interest with the primary goal of controlling for otherwise unexplained variance in order to test motivation system and Favorable Departure. This approach reduced the number of possible analyses carried out and allowed a test of motivation system and Favorable Departure across different variables and combinations of variables. However, if a future research study has the goal of determining which variables (besides motivation system and Favorable Departure) are predictive, the analyses should be planned *a priori* rather than done in reaction to their correlational values.

Predictors of Post-Treatment Outcomes

Secondary analyses were used to determine which demographic and baseline measures predicted post-treatment, as opposed to follow-up, outcomes. In the 16-year follow-up sample, post-treatment outcomes were negatively predicted by number of presenting problems at admission and positively predicted by age at admission. In addition, boys and members of minority groups had marginally poorer post-treatment outcomes (for a full summary, see Table 51). In the 5-year follow-up sample, number of formal prior placements was a consistent negative predictor of post-treatment outcome, and, once again, girls tended to have better outcomes at the end of their stays at GBT (for a full summary, see Table 52).

Larzelere, Smith, and colleagues (2001) developed a “prognosis index” for youth at admission to GBT based on regressing the Departure Success Scale onto various baseline predictors, some of which were the same as those used in this study. The sample used in that study did not overlap with the samples in this study. Gender and ethnicity were not included in the index, so their significance here cannot be replicated. Instead of

summing all presenting problems, as done in this study, they were factor analyzed into three groups in the index. Two of those factors were significant predictors in some of the models but not in the most complete one; total number of presenting problems was a significant predictor in the 16-year sample. Age at admission, which was a positive predictor of outcome in the 16-year group, was also a significant positive predictor in the index study. Interestingly, though number of formal prior placements was a significant predictor in the 5-year sample, number of prior placements (which includes both formal and informal) was not significant in the index study. On the other hand, ROLES at admission was a significant predictor in the index, but not in the present study (except for predicting ROLES at departure). The numbers of both internalizing and externalizing diagnoses were significant predictors of outcome in the index, but they were not significant predictors in this study.

In a place that was called Boys Town until recently, it is ironic that girls had consistently better outcomes at post-treatment in both samples and in the 5-year follow-up survey. However, these differences did not emerge in the 16-year follow-up survey. This suggests that the poorer outcomes for males at post-treatment and 5-year follow-up were maturational. Perhaps better outcomes for females at younger ages reflect more maturity at that age and a greater willingness of girls to accept social influence. However, by their late 20's and early 30's, the men seemed to have "caught up" to their female counterparts. In terms of minority status, Caucasians had marginally better post-treatment outcomes on two measures in the 16-year follow-up cohort, but this was not replicated in the 5-year. This may have reflected many different things, from socio-

economic differences between white and minority groups to unconscious bias on the part of the raters, so it is difficult to conclude anything from these findings.

Though age at admission negatively predicted length of stay (because younger children can stay longer at GBT before they graduate from high school), it positively predicted outcomes in the 16-year sample. This may suggest that GBT's treatment is better suited to older adolescents. It is also possible that this is an artifact for baseline severity. Youth who needed out-of-home placements at younger ages may have had more severe problems than those referred later. However, the effect of age at admission remained significant within the same model as number of presenting problems, so age and this measure of severity each had unique contributions. It is not surprising that number of presenting problems in the 16-year sample and number of formal prior placements in the 5-year were significant negative predictors of outcome. Having more personal and systemic problems at baseline and having a longer history of out-of-home placements are both markers for greater severity.

The implications of these predictions are not clear. GBT should certainly continue to admit boys, though it may need to examine how boys react uniquely to its treatment. GBT might consider focusing on admitting older adolescents, though that might be presumptuous, given that in the more recent sample, the 5-year follow-up cohort, age was not a factor. GBT might consider granting admission to youth with fewer presenting problems and fewer formal prior placements, but its *raison d'être* is to care for youth with multiple problems and unstable living environments. Besides, these youth may have improved at the same rate as youth from less severe backgrounds, but

they likely started at lower levels of functioning. This analysis does not give a detailed enough picture of this question to glean many implications.

As with the primary analyses, these findings should be interpreted with caution. Predictors were entered into analyses based on bivariate correlations with the dependent variables of interest rather than *a priori*. Furthermore, multicollinearity may be obscuring other variables that relate significantly to these post-treatment outcome measures.

Future Research

As this was just a preliminary test of the hypothesis that a stepped-care approach, in which a stage of token reinforcement is followed by a stage of self-regulation training and natural and logical consequences, would promote long-term internalization and generalization of behavior change better than a token economy alone, much future research remains to be done. The ideal test of this hypothesis, of course, is a prospective, randomized controlled trial in which youth are placed into stepped-care, token-only, and natural-and-logical-reinforcement-only groups. Baseline, in-treatment, post-treatment, and follow-up measures would be compared. However, more small-scale research should be done before reaching this stage. Three types of potential future studies – predictor studies, lab and questionnaire studies, and small clinical studies – are mentioned here.

Despite their limitations, predictor studies are a useful and inexpensive way to test this hypothesis, even if one cannot infer causality from them. Instead of looking at specific outcomes, future studies with the same or similar datasets could develop cross-variable profiles of well- and poorly-functioning adults at follow-up to see whether motivation system uniquely predicts group membership. Of course, in any future follow-

up, GBT would do well to find a way to include as many graduates currently serving in the military as possible. In addition, long-term outcome measures could include measures that have established predictive validity for various outcomes, such as psychopathology and criminality, rather than sociological surveys. Similarly, mediator/moderator studies or path analyses could look at the markers of progress and long-term outcome across multiple time points. For example, while motivation system may not predict a particular outcome independently, it may mediate the effects of a pre-treatment variable on long-term outcome. Alternatively, there may be pre-treatment or demographic variables that moderate the effects of motivation system on outcome. Also, given the limitations of self-reports, future studies could look at multimodal outcomes. For example, rather than just asking youth about arrest histories, it may be possible to compile that sort of data from public records. Information could also be gathered from informants, such as parents/guardians, case workers, and referring agencies. It would certainly be beneficial to replicate this study with different populations and in other agencies and settings that have adopted a similar stepped-care approach.

Research on the topic of promoting generalization in behavioral treatment should certainly not be limited to disruptive adolescents in residential token programs. Similar stepped-care models can be developed and tested with other groups that are frequently treated with behavior modification, from adult inpatients in token economies to behavior disordered children being treated with parent training. However, perhaps before the field will embrace the approach and emphasis on generalization described here, research must directly examine some of the underlying assumptions, such as whether or not logical consequences promote generalization better than artificial reinforcement. Laboratory

studies with both nonclinical and clinical populations may be useful in parallel to studies in less-controlled real-world environments. For example, participants in a laboratory study could be given a contingency based on a logical consequence or based on artificial reinforcement. After the contingency is removed, the target behavior could be observed, and attitudes about the behavior could be measured. A questionnaire describing examples of artificial contingencies and logical contingencies could be given to children of various ages as well as parents, teachers, and other caretakers to measure the acceptability of each contingency and participants' likelihood of either adhering to it or using it. Children could rate the quantity of information and controllingness in each example, as well, and their feelings toward the adult in the example. Laboratory and questionnaire studies cannot replace field-based studies in terms of validity and generalizability, but they can establish some of the basic principles, mechanisms, and boundary conditions that should be taken into account when developing complex interventions for various groups. Because good clinical tests of these principles require long-term follow-up in addition to treatment time, small-scale analog studies may be resource- and time-efficient methods of establishing some of the basic phenomena.

Finally, small clinical studies, using single case-study and small group designs, could be used to test natural and logical consequences and self-regulation training in real-world or simulated contexts. It would be important to do this across different populations, age groups, settings, and diagnoses. Studies on both tangible and verbal rewards suggest that younger and older children perceive and react to them differently. Performance by older children is more likely to be undermined by controlling feedback (Kast & Connor, 1988; Newman, Beauchamp, Latimer, & Kao, 2003). It would be

interesting, therefore, to examine perceptions of controllingness and subsequent performance in reaction to both artificial and logical consequences across the developmental span. In terms of different clinical groups, the concepts about generalization in this paper would seem to apply generally to any population or situation in which generalization is desirable. However, their applicability to individuals with different clinical diagnoses is an empirical question. Do nonclinical groups respond to artificial or natural reinforcement differently than behavior disordered groups? Do children on the autistic spectrum, with their atypical social cognition, differentiate between different classes of reinforcement? Finally, behavioral contingencies and controllingness are culturally mediated concepts. Are various contingencies and methods of reinforcement and punishment perceived differently by members of different ethnic and cultural groups? Does cultural background moderate how these methods influence internalization and generalization? Perhaps reactions to the controllingness of feedback from “elders” differ between Western cultures and more traditional or collective cultures.

Implications for Practice

In this preliminary study on the predictive strength of using specific generalization techniques, the findings were weak. Therefore, there is no more empirical imperative to change practice after this study than there was before it. The practical implications of the theoretical arguments introduced in this paper are wide-ranging. They imply a revamping of behavioral programs to limit the use of artificial reinforcement to situations in which new behaviors are learned or rapid behavior change is needed because of the severity of the behavior. Following initial behavior change, artificial reinforcement should be quickly tapered and replaced with natural, social, and logical

reinforcement as well as training in self-regulation. However well-founded these suggestions are in the literature, they were not unequivocally supported in this study.

In terms of practice implications for GBT and the Teaching Family Model, perhaps it would be useful to operationalize and standardize self-regulation and natural and logical consequences in the Achievement System more clearly. Despite being the least structured motivation system for youth, quality control and strict adherence to the protocol by family teachers may be at least as important in Achievement as they are in the more structured motivation systems because of the incredible challenges when programming for generalization. Based on the predictor findings, perhaps more attention should be paid to how boys react to the Teaching Family Model. Similarly, additional attention might be warranted for youth who enter with arrest histories and large numbers of family problems. These findings, like the motivation system findings, were limited, so no major practical changes appear warranted at this time.

References

- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Behavioral Risk Factor Surveillance System* [Data file] (2003). Retrieved from the Centers for Disease Control Data Archive Web site:
www.cdc.gov/brfss/about.htm
- Brestan, E. V., & Eyberg, S. M. (1998). Effective psychosocial treatment of conduct-disordered children and adolescents: 29 years, 82 studies, and 5,272 kids. *Journal of Clinical Child Psychology*, 27, 180-189.
- Center for Human Resource Research, The Ohio State University. (2005). NLSY97 User's Guide: A Guide to the Rounds 1-7 Data, National Longitudinal Survey of Youth 1997. Retrieved from
<http://www.nlsinfo.org/nlsy97/docs/97HTML00/97guide/toc.htm>
- Chmelka, B., & Noss, B. (1990, August). *Description of the follow-up original, modified, and final questionnaires: Their origins, items, and scales*. Boys Town, NE: Department of Program Planning, Research, and Evaluation, Father Flanagan's Boys' Home, Inc.
- Davis, J. A., Smith, T. W., & Marsden, P. V. (2003). *General Social Survey 1972-2000 Cumulative Codebook* [Data file]. Retrieved from Inter-University Consortium for Political and Social Research, University of Michigan Web site:
<http://webapp.icpsr.umich.edu/GSS/>

- Davis, J., & Daly, D. L. (2003). *Girls and Boys Town long-term residential program training manual* (4th ed.). Boys Town, NE: Father Flanagan's Boys' Home, Inc.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268.
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627-668.
- Dickerson, F. B., Tenhula, W. N., & Green-Paden, L. D. (2005). The token economy for schizophrenia: Review of the literature and recommendations for future research. *Schizophrenia Research*, 75(2-3), 405-416.
- Dreikurs, R., & Loren, G. (1968). *Logical consequences*. New York: Meredith Press.
- Elliott, D. S., Ageton, S. S., Canter, R. J., Huizinga, D. (1979). *The dynamics of delinquent behavior: A national survey*. Boulder, CO: Behavioral Research Institute.
- Farmer, E. M. Z., Compton, S. N., Burns, B. J., & Robertson, E. (2002). Review of the evidence base for treatment of childhood psychopathology: Externalizing disorders. *Journal of Consulting & Clinical Psychology*, 70(6), 1267-1302.
- Field, C. E., Nash, H. M., Handwerk, M., & Friman, P. C. (2004). A modification of the token economy for nonresponsive youth in family-style residential care. *Behavior Modification*, 28(3), 438-457.
- Foxx, R. M. (1998). A comprehensive treatment program for inpatient adolescents. *Behavioral Interventions*, 13(1), 67-77.

- Friman, P. C. (2000). Behavioral, family-style residential care for troubled out-of-home adolescents: Recent findings. In J. Austin & J. E. Carr (Eds.), *Handbook of applied behavior analysis* (pp. 187-209). Reno, NV: Context Press.
- Friman, P. C., Osgood, D. W., Smith, G., Shanahan, D., Thompson, R. W., Larzelere, R., et al. (1996). A longitudinal evaluation of prevalent negative beliefs about residential placement for troubled adolescents. *Journal of Abnormal Child Psychology*, 24(3), 299-324.
- Gottschalk, R., Davidson, W. S., II, Mayer, J., & Gensheimer, L. K. (1987). Behavioral approaches with juvenile offenders: A meta-analysis of long-term treatment efficacy. In E. K. Morris & C. J. Braukmann (Eds.), *Behavioral approaches to crime and delinquency: A handbook of application, research, and concepts* (pp. 399-422). New York: Plenum Press.
- Greene, D., Sternberg, B., & Lepper, M. R. (1976). Overjustification in a token economy. *Journal of Personality and Social Psychology*, 34(6), 1219-1234.
- Greene, R. W., Ablon, J. S., & Goring, J. C. (2003). A transactional model of oppositional behavior: Underpinnings of the Collaborative Problem Solving approach. *Journal of Psychosomatic Research*, 55(1), 67-75.
- Hawkins, R. P., Almeida, M. C., Fabry, B., & Reitz, A. L. (1992). A scale to measure restrictiveness of living environments for troubled children and youths. *Hospital and Community Psychiatry*, 43, 54-58.
- Hersen, M. (1976). Token economies in institutional settings. *Journal of Nervous and Mental Disease*, 162(3), 206-211.

- Hicks, T., & Munger, R. (1990). A school day treatment program using an adaptation of the teaching-family model. *Education & Treatment of Children, 13*(1), 63.
- Hobbs, T. R., & Holt, M. M. (1976). The effects of token reinforcement on the behavior of delinquents in cottage settings. *Journal of Applied Behavior Analysis, 9*(2), 189-198.
- Horner, R. D. (1971). Establishing use of crutches by a mentally retarded spina bifida child. *Journal of Applied Behavior Analysis, 4*(3), 183-189.
- Huefner, J. C., Ringle, J. L., Chmelka, M. B., & Ingram, S. D. (2007). Breaking the cycle of intergenerational abuse: The long-term impact of a residential care program. *Child Abuse & Neglect, 31*(2), 187-199.
- Jewell, J., Handwerk, M., Almquist, J., & Lucas, C. (2004). Comparing the validity of clinician-generated diagnosis of conduct disorder to the Diagnostic Interview Schedule for Children. *Journal of Clinical Child and Adolescent Psychology, 33*(3), 536-546.
- Jones, R. R., Weinrott, M. R., & Howard, J. R. (1981). *Impact of the Teaching Family Model on troublesome youth* (Final Report MH 31018). Washington, DC: Center for the Studies of Crime and Delinquency, National Institutes of Mental Health.
- Kast, A., & Connor, K. (1988). Sex and age differences in response to informational and controlling feedback. *Personality & Social Psychology Bulletin, 14*(3), 514-523.
- Kazdin, A. E. (1977). *The token economy: A review and evaluation*. New York: Plenum Press.
- Kazdin, A. E. (1982). The token economy: A decade later. *Journal of Applied Behavior Analysis, 15*(3), 431-445.

Kazdin, A. E., & Bootzin, R. R. (1972). The token economy: An evaluative review.

Journal of Applied Behavior Analysis, 5(3), 343-372.

Kazdin, A. E., Siegel, T. C., & Bass, D. (1992). Cognitive problem-solving skills training and parent management training in the treatment of antisocial behavior in children. *Journal of Consulting & Clinical Psychology*, 60(5), 733-747.

Kendall, P. C. (1989). The generalization and maintenance of behavior change:

Comments, considerations, and the "no-cure" criticism. *Behavior Therapy*, 20(3), 357-364.

Kirigin, K. A., Braukmann, C. J., Atwater, J. D., & Wolf, M. M. (1982). An evaluation of teaching-family (Achievement Place) group homes for juvenile offenders. *Journal of Applied Behavior Analysis*, 15(1), 1-16.

Larzelere, R. E., Daly, D. L., Davis, J. L., Chmelka, M. B., & Handwerk, M. L. (2004).

Outcome evaluation of Girls and Boys Town's Family Home Program. *Education & Treatment of Children*, 27(2), 130-149.

Larzelere, R. E., Dinges, K., Schmidt, M. D., Spellman, D. F., Criste, T. R., & Connell, P. (2001). Outcomes of residential treatment: A study of the adolescent clients of Girls and Boys Town. *Child & Youth Care Forum*, 30(3), 175-185.

Larzelere, R., Smith, G. L., Jorgensen, D. D., Daly, D. L., & Handwerk, M. L. (2001). A prognosis index for group home residential programs. In C. C. Newman, C. J. Liberton, K. Kutash & R. M. Friedman (Eds.), *The 13th Annual Research Conference Proceedings: A System of Care for Children's Mental Health: Expanding the Research Base* (pp. 359-362). Tampa, FL: University of South

Florida, Louis de la Parte Florida Mental Health Institute, Research and Training Center for Children's Mental Health.

Leichtman, M. (2006). Residential treatment of children and adolescents: Past, present, and future. *American Journal of Orthopsychiatry*, 76(3), 285-294.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality & Social Psychology*, 28(1), 129-137.

Levine, F. M., & Fasnacht, G. (1974). Token rewards may lead to token learning. *American Psychologist*, 29(11), 816-820.

National Center for Health Statistics, Centers for Disease Control and Prevention. (2007, January). *National Health and Nutrition Examination Survey*. Retrieved from <http://www.cdc.gov/nchs/nhanes.htm>

National Household Survey on Drug Abuse [Data file] (1985). Available from Substance Abuse and Mental Health Data Archive, University of Michigan Web site: <http://www.icpsr.umich.edu/SAMHDA/index.html>

Neilans, T. H., & Israel, A. C. (1981). Towards maintenance and generalization of behavior change: Teaching children self-regulation and self-instructional skills. *Cognitive Therapy and Research*, 5(2), 189-195.

Newman, J., Beauchamp, H. M., Latimer, B. C., & Kao, C. C. (2003). Developmental understanding of means-end contingencies: Effect of familiarity of contingency content. *British Journal of Developmental Psychology*, 21, 527-542.

Nowicki, J., & Strickland, B. R. (1973). A locus of control scale for children. *Journal of Consulting and Clinical Psychology*, 40(1), 148-154.

- O'Brien, C., Ringle, J., & Larzelere, R. (2001, August). Serving youths by modifying treatment: Girls and Boys Town approach uses ongoing outcome measures. *Behavioral Healthcare Tomorrow*, 10, 19-21.
- Phillips, E. L. (1968). Achievement Place: Token reinforcement procedures in a homestyle rehabilitation setting for "pre-delinquent" boys. *Journal of Applied Behavior Analysis*, 1, 213-223.
- Phillips, E. L., Phillips, E. A., Fixsen, D. L., & Wolf, M. M. (1974). *The teaching-family handbook*. Lawrence, KS: University of Kansas Press.
- Reid, M. J., Webster-Stratton, C., & Hammond, M. (2003). Follow-up of children who received the Incredible Years intervention for Oppositional-Defiant Disorder: Maintenance and prediction of 2-year outcome. *Behavior Therapy*, 34(4), 471-491.
- Research Triangle Institute. (2004). 2005 *National Survey on Drug Use and Health: CAI Specs for Programming, English Version*. Retrieved from <http://oas.samhsa.gov/nsduh/2k5MRB/2k5CAISpecs.pdf>
- Schuhmann, E. M., Foote, R. C., Eyberg, S. M., Boggs, S. R., & Algina, J. (1998). Efficacy of Parent-Child Interaction Therapy: Interim report of a randomized trial with short-term maintenance. *Journal of Clinical Child Psychology*, 27(1), 34-45.
- Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39(1), 28-38.

- Social Capital Community Benchmark Survey* [Data file]. (2000). Retrieved from The Roper Center, University of Connecticut Web site:
http://ropercenter.uconn.edu/scc_bench.html
- Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, 10(2), 349-367.
- Stokes, T. F., & Osnes, P. G. (1989). An operant pursuit of generalization. *Behavior Therapy*, 20(3), 337-355.
- Sullivan, M. A., & O'Leary, S. G. (1990). Maintenance following reward and cost token programs. *Behavior Therapy*, 21(1), 139-149.
- Sweet, J. A., & Bumpass, L. L. (1996). *The National Survey of Families and Households – Waves 1 and 2: Data description and documentation* [Data file]. Retrieved from the University of Wisconsin-Madison, Center for Demography and Ecology Web site: www.ssc.wisc.edu/nsfh/home.htm
- Thompson, R. W., Huefner, J. C., Ringle, J. L., & Daly, D. L. (2005). Adult outcomes of Girls and Boys Town youth: A follow-up report. In C. C. Newman, C. J. Liberton, K. Kutash & R. M. Friedman (Eds.), *The 17th Annual Research Conference Proceedings: A System of Care for Children's Mental Health: Expanding the Research Base* (pp. 529-534). Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, Research and Training Center for Children's Mental Health.
- Thompson, R. W., Smith, G. L., Osgood, D. W., Dowd, T. P., Friman, P. C., & Daly, D. L. (1996). Residential care: A study of short- and long-term educational effects. *Children and Youth Services Review*, 18(3), 221-242.

- Turkewitz, H., O'Leary, K. D., & Ironsmith, M. (1975). Generalization and maintenance of appropriate behavior through self-control. *Journal of Consulting and Clinical Psychology, 43*(4), 577-583.
- United States Census Bureau. (2003, July 1). *Long form questionnaire*. Retrieved from <http://www.census.gov/dmd/www/2000quest.html>
- Wang, K., Cantor, D., & Vaden-Kiernan, N. (2000, February). *1999 NSAF questionnaire* (National Survey of America's Families Round Two Methodology Series Report No. 1). Retrieved from <http://www.urban.org/center/anf/questionnaire.cfm>
- Wolf, M. M., Phillips, E. L., Fixsen, D. L., Braukmann, C. J., Kirigin, K. A., Willner, A. G., et al. (1976). Achievement Place: The teaching-family model. *Child Care Quarterly, 5*(2), 92-103.
- Wolfe, B. D., Dattilo, J., & Gast, D. L. (2003). Effects of a token economy system within the context of cooperative games on social behaviors of adolescents with emotional and behavioral disorders. *Therapeutic Recreation Journal, 37*(2), 124-141.
- Wood, R., & Flynn, J. M. (1978). A self-evaluation token system versus an external evaluation token system alone in a residential setting with predelinquent youth. *Journal of Applied Behavior Analysis, 11*(4), 503-512.
- Woodlock, D., Juliano, N., & Ringle, J. (2002, August). Giving hope to troubled adolescents: Diverse treatment approach emphasizes self-control skills. *Behavioral Healthcare Tomorrow, 11*, 8-11.

Table 1

16-Year Follow-Up Study Variables: Demographic and Baseline Variables

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Demographics				
Gender	0 = female; 1 = male	GBT admissions records	N/A	N/A
Ethnicity	Dummy variables for African American, Hispanic, and Other	GBT admissions records	N/A	N/A
Minority status	0 = minority; 1 = Caucasian	GBT admissions records	N/A	N/A
Age at admission	Age in years when first arrived at GBT	GBT admissions records	N/A	N/A
Age at follow-up	Age in years when completed survey	Follow-up survey	N/A	N/A
History				
ROLES at admission	30-point scale of living environment immediately before admission. Ranges from 1 = "Jail" to 25 = "Independent living by self."	GBT admissions records	Hawkins et al. (1992)	N/A
Number of prior placements	Total number of formal (e.g., foster home) and informal (e.g., staying with a relative) prior placements.	GBT admissions records	N/A	N/A
Number of presenting problems	Total number of problems at admission out of a list of 16.	GBT admissions records	GBT	N/A

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Estimated GPA	Typical grades received by youth prior to admission. Ranges from 4 = "Mostly A's" to 0 = "Mostly F's."	Pre-treatment questionnaire completed by youth	National Youth Survey (Elliott et al., 1979)	N/A
Self-reported delinquency				
Aggressive Offense Scale	Mean of 6 yes-no questions about physically aggressive behavior. Ranges from 0 (all no's) to 1 (all yes's).	Pre-treatment questionnaire completed by youth	National Youth Survey (Elliott et al., 1979)	.42
Property Offense Scale	Mean of 11 yes-no questions about vandalism and theft. Ranges from 0 (all no's) to 1 (all yes's).	Pre-treatment questionnaire completed by youth	National Youth Survey (Elliott et al., 1979)	.76
Drug Scale	Mean of 6 yes-no questions about using and selling alcohol and illicit drugs. Ranges from 0 (all no's) to 1 (all yes's).	Pre-treatment questionnaire completed by youth	National Youth Survey (Elliott et al., 1979)	.70
Psychological scales				
Consumer Scale	Mean of 4 7-point items about youth satisfaction with the adults he or she was living with before admission. Ranges from 1 = "completely dissatisfied" to 7 = "completely satisfied."	Pre-treatment questionnaire completed by youth	<i>The Teaching-Family Handbook</i> (Phillips et al., 1974)	.83
Locus of Control Scale at admission	Mean of 8 dichotomous (agree/disagree) items about locus of control (LoC). Ranges from 1 = high external LoC to 2 = high internal LoC.	Pre-treatment questionnaire completed by youth	Powerlessness Scale (Nowicki & Strickland, 1973)	.66

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Happiness Rating	Single item about overall happiness, where 0 = “not too happy,” 1 = “pretty happy,” and 2 = “very happy.”	Pre-treatment questionnaire completed by youth	GBT	N/A

^aInternal reliability (Cronbach’s alpha) for multi-item scales as measured in the present sample.

Table 2

16-Year Follow-Up Study Variables: Treatment and Post-Treatment

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Treatment				
Length of stay	Total length of stay at GBT in days. If there were multiple stays, days are totaled.	GBT records	N/A	N/A
Motivation system at departure	Last motivation system the youth was on before last departure, where 0 = Subsystem, 1 = Daily, 2 = Weekly, and 3 = Achievement / Natural & Logical. Treated as ordinal.	GBT records	N/A	N/A
Post-treatment				
Favorable Departure Rating	Mean of 3 5-point ratings given at departure: Favorable Conditions of Leaving, Goal Achievement, and Prediction of Future Success. Ranges from 0 (least favorable) to 4 (most favorable).	Clinical Specialist	GBT	.93
Prediction of Future Success	5-point rating of staff's predictions about a child's future, ranging from 0 = "very unsuccessful" to 5 = "very successful." Component of Favorable Departure Rating.	Clinical Specialist	GBT	N/A

^aInternal reliability (Cronbach's alpha) for multi-item scales as measured in the present sample.

Table 3

16-Year Follow-Up Study Variables: Follow-Up Survey

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Psychological outcomes				
Mental health index	Mean of 3 5-point items about frequency of internalizing symptoms. Ranges from 1 (frequent internalizing symptoms) to 5 (rare internalizing symptoms).	16-Year Follow-Up Survey	General Social Survey (J. A. Davis et al., 2003)	.77
Recent substance abuse	1 = had 5 or more drinks on 6 or more occasions in the past 30 days OR used illegal drugs in the past 12 months; 0 = neither of these.	16-Year Follow-Up Survey	Variable created from items taken from multiple surveys.	N/A
Locus of Control at follow-up	Mean of 7 5-point items about LoC. Ranges from 1 (high external LoC) to 5 (high internal LoC).	16-Year Follow-Up Survey	National Survey of Families and Households (Sweet & Bumpass, 1996)	.80
Satisfaction Index	Mean of 7 7-point items about satisfaction in various domains. Ranges from 1 (least satisfied) to 7 (most satisfied).	16-Year Follow-Up Survey	National Survey of Families and Households (Sweet & Bumpass, 1996)	.77
Happiness at follow-up	4-point item about overall happiness, ranging from 0 = "Not at all happy" to 3 = "Very happy."	16-Year Follow-Up Survey	Social Capital Community Benchmark Survey (2000)	N/A

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Deceased	1 = Deceased by 2002; 0 = Not deceased or unknown.	16-Year Follow-Up Survey Search	N/A	N/A
Criminality				
Arrested in past 12 months	1 = Arrested in past 12 months; 0 = Not arrested in past 12 months.	16-Year Follow-Up Survey	National Household Survey on Drug Abuse (1985)	N/A
Perpetration Index	Total number of crimes perpetrated in the past 12 months, based on 6 items that asked for frequencies of perpetrating specific types of crimes.	16-Year Follow-Up Survey	National Youth Survey (Elliott et al., 1979)	.56
Education				
Highest grade completed	Highest grade of school completed. Ranges from 7 (7 th grade) to 20 (8 years of college or greater).	16-Year Follow-Up Survey	Multiple surveys have this item.	N/A
In school past 12 months	1 = enrolled in academic, technical, or military school or training in past 12 months; 0 = not enrolled in past 12 months.	16-Year Follow-Up Survey	2000 Census (United States Census Bureau, 2003)	N/A
Employment and earnings				
Currently working	1 = currently working at a paying job; 0 = not working.	16-Year Follow-Up Survey	2000 Census (United States Census Bureau, 2003)	N/A

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Job Satisfaction	5-point scale of satisfaction with current job or, if no current job, job done most recently or most often. Ranges from 0 = “very dissatisfied” to 4 = “very satisfied.”	16-Year Follow-Up Survey	National Youth Survey (Elliott et al., 1979)	N/A
Household income	Household income in 2001. Ranges from 1 = “less than \$20,000” to 7 = “greater than \$100,000.”	16-Year Follow-Up Survey	Social Capital Community Benchmark Survey (2000)	N/A
Could not pay bills	1 = there was a time in past 12 months when unable to pay mortgage, rent, or utilities; 0 = always able to pay.	16-Year Follow-Up Survey	National Survey of America's Families (Wang et al., 2000)	N/A
Military service	1 = been in military for at least 2 consecutive months; 0 = never in military.	16-Year Follow-Up Survey	General Social Survey (J. A. Davis et al., 2003)	N/A
Social systems				
Living with spouse/partner	1 = married OR living with partner in committed relationship; 0 = not living with partner.	16-Year Follow-Up Survey	Created from 2 items; marital question on multiple surveys.	N/A
Child Involvement Index	Sum of 3 6-point items about frequency of spending time with children. Ranges from 3 (“never or rarely” on all items) to 18 (“almost every day” on all items)	16-Year Follow-Up Survey	National Survey of Families and Households (Sweet & Bumpass, 1996)	.87

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Social Involvement Count	Number of domains in which socialize at least once per month. Ranges from 0 (no domains) to 4 (all 4 domains).	16-Year Follow-Up Survey	National Survey of Families and Households (Sweet & Bumpass, 1996)	N/A
Spirituality	4-point item on the extent to which the respondent considers self spiritual. Ranges from 0 = “not spiritual at all” to 3 = “very spiritual.”	16-Year Follow-Up Survey	General Social Survey (J. A. Davis et al., 2003)	N/A
Evaluation of GBT				
GBT Index	Mean of 5 5-point items about helpfulness of treatment at GBT. Ranges from 0 (least helpful) to 4 (most helpful).	16-Year Follow-Up Survey	GBT	.84

^aInternal reliability (Cronbach’s alpha) for multi-item scales as measured in the present sample.

Table 4

5-Year Follow-Up Study Variables: Demographic and Baseline Variables

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Demographics				
Gender	0 = female; 1 = male	GBT admissions records	N/A	N/A
Ethnicity	Dummy variables for African American, Hispanic, and Other	GBT admissions records	N/A	N/A
Minority status	0 = minority; 1 = Caucasian	GBT admissions records	N/A	N/A
Age at admission	Age in years when arrived at GBT	GBT admissions records	N/A	N/A
Age at departure	Age in years when departed from GBT. (Add 5 years for age at long-term follow-up.)	GBT records	N/A	N/A
IQ	Full Scale IQ	File Review	Mostly WISC-III	N/A
History				
ROLES at admission	30-point scale of living environment immediately before admission. Ranges from 1 = "Jail" to 25 = "Independent living by self."	File Review	Hawkins et al. (1992)	N/A
Number of formal prior placements	Total number of formal (e.g., foster home, detention center) prior placements.	File Review	N/A	N/A

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Number of presenting problems	Total number of personal problems at admission out of a list of 38. Completed by admissions counselor.	GBT admissions records	GBT	.80
Number of family problems	Total number of family problems at admission out of a list of 34. Completed by admissions counselor.	GBT admissions records	GBT	.73
Ever arrested	1 = arrest history prior to admission; 0 = never arrested	File Review	N/A	N/A
Psychopathology at admission (DISC)				
Total number of diagnoses	Total number of DSM diagnoses given by the DISC	GBT admissions records	DISC (Shaffer et al., 2000)	N/A
Number of internalizing diagnoses	Total number of DSM internalizing (e.g., depression, social anxiety disorder) diagnoses given by the DISC	GBT admissions records	DISC (Shaffer et al., 2000)	N/A
Number of externalizing diagnoses	Total number of DSM externalizing (e.g., conduct disorder, oppositional defiant disorder) diagnoses given by the DISC	GBT admissions records	DISC (Shaffer et al., 2000)	N/A
Any substance abuse diagnosis	1 = at least one substance abuse diagnosis given by the DISC; 0 = no substance abuse diagnosis	GBT admissions records	DISC (Shaffer et al., 2000)	N/A

^aInternal reliability (Cronbach's alpha) for multi-item scales as measured in the present sample.

Table 5

5-Year Follow-Up Study Variables: Treatment and Post-Treatment

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Treatment				
Length of stay	Length of stay at GBT in days	GBT records	N/A	N/A
Motivation system at departure	Last motivation system the youth was on before departure, where 0 = Subsystem, 1 = Daily, 2 = Weekly, and 3 = Achievement / Natural & Logical. Treated as ordinal.	GBT records	N/A	N/A
Post-treatment				
Presenting problems percentage improved upon	Percentage of presenting problems at admission that were improved upon during treatment	GBT discharge records	GBT	N/A
ROLES at departure	30-point scale of living environment immediately after departure. Ranges from 1 = "Jail" to 25 = "Independent living by self."	GBT discharge records	Hawkins et al. (1992)	N/A
Favorable Departure Rating	Total of 4 7-point ratings given at departure: Overall Behavior, Departure Conditions, Goal Achievement, and Prediction of Future Success. Ranges from 4 (least favorable) to 28 (most favorable).	Clinical Specialist	GBT	.94

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Departure Success Scale	Favorable Departure Rating + ROLES at departure (collapsed into 6-point scale, range 1-6) + Program Completion Indicator (1 = complete or graduation; 0 = not complete). Ranges from 5 (least favorable) to 35 (most favorable).	Clinical Specialist and discharge data	GBT	.45
Prediction of Future Success	7-point scale of staff's predictions about a child's future, ranging from 1 = "very unsuccessful" to 7 = "very successful." Component of Favorable Departure Rating and Departure Success Scale.	GBT Clinical Specialist	GBT	N/A

^aInternal reliability (Cronbach's alpha) for multi-item scales as measured in the present sample.

Table 6

5-Year Follow-Up Study Variables: Follow-Up Survey

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Psychological outcomes				
Mental health index	Mean of 4 4-point items about frequency of internalizing symptoms in the past month. Ranges from 1 (frequent internalizing symptoms) to 4 (rare internalizing symptoms).	5-Year Follow-Up Survey	NLSY97 (Center for Human Resource Research, OSU, 2005)	.69
Recent substance abuse	1 = Avg. 5 or more drinks per outing in past month OR used marijuana in past month OR used hard drugs in past 12 months; 0 = none of these.	5-Year Follow-Up Survey	NLSY97	N/A
Friends' substance use	Mean of 3 4-point scales about proportion of friends who use marijuana, get drunk weekly, and use hard drugs. Ranges from 0 (no friends do these) to 3 (all friends do all of these).	5-Year Follow-Up Survey	National Health and Nutrition Examination Survey (National Center for Health Statistics, CDC, 2007)	.66
Criminality				
Perpetrated any crime in past 12 months	1 = perpetrated a crime in any of 6 categories in past 12 months; 0 = did not perpetrate	5-Year Follow-Up Survey	NLSY97	N/A
Arrested in past 12 months	1 = arrested in past 12 months; 0 = not arrested in past 12 months.	5-Year Follow-Up Survey	NLSY97	N/A

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Currently incarcerated	1 = currently incarcerated; 0 = not incarcerated	5-Year Follow-Up Survey	NLSY97	N/A
Education				
Highest grade completed	Highest grade of school completed. Ranges from 9 (9 th grade) to 17 (5 years of college).	5-Year Follow-Up Survey	NLSY97	N/A
In school past 12 months	1 = enrolled in any school or training program in past 12 months; 0 = not enrolled in past 12 months.	5-Year Follow-Up Survey	NLSY97	N/A
Employment and earnings				
Worked last week	1 = worked full- or part-time last week; 0 = did not work	5-Year Follow-Up Survey	NLSY97	N/A
Household income	Household income in 2004. Ranges from 1 = "\$1 - \$5,000" to 7 = "greater than \$250,001."	5-Year Follow-Up Survey	NLSY97	N/A
Military service	1 = been in military for at least 2 consecutive months; 0 = not in military.	5-Year Follow-Up Survey	NLSY97	N/A
Social systems				
Living with spouse/partner	1 = currently living with spouse or partner; 0 = not living with partner	5-Year Follow-Up Survey	National Survey on Drug Use and Health (Research Triangle Institute, 2004)	N/A
Closeness to spouse/partner	Single 11-point item on closeness to spouse/partner. Ranges from 0 = "not at all close" to 10 = "very close."	5-Year Follow-Up Survey	NLSY97	N/A

Variable Name	Description and Coding	Data Source	Questionnaire Source	α^a
Conflict with spouse/partner	Single 11-point item on conflict with spouse/partner. Ranges from 0 = “a lot of conflict” to 10 = “no conflict.”	5-Year Follow-Up Survey	NLSY97	N/A
Average child involvement	Mean of 4 5-point items about frequency of interacting with one’s children. Ranges from 0 (“not at all” on all items) to 4 (“once a day” on all items)	5-Year Follow-Up Survey	NLSY97	.97
Support from at least one parent	Mother/mother figure is at least somewhat supportive OR father/father figure is at least somewhat supportive	5-Year Follow-Up Survey	NLSY97	N/A
Religion important	Singe 4-point item on the importance of religious beliefs in one’s life. Ranges from 0 = “strongly disagree” to 3 = “strongly agree.”	5-Year Follow-Up Survey	National Health and Nutrition Examination Survey (National Center for Health Statistics, CDC, 2007)	N/A

Note. NLSY97 = National Longitudinal Survey of Youth 1997 (Center for Human

Resource Research, The Ohio State University, 2005).

^aInternal reliability (Cronbach’s alpha) for multi-item scales as measured in the present sample.

Table 7

Descriptive Statistics for the 16-Year Follow-Up Sample

Variable Name	Scale Range	Mean (SD)	N
Pre-treatment			
History			
ROLES at admission	1-25	17.02 (6.94)	397
Number of prior placements	0-? ^a	5.42 (3.31)	493
Number of presenting problems	0-16	4.41 (1.43)	497
Estimated GPA	0-4	3.02 (1.05)	456
Self-reported delinquency			
Aggressive Offense Scale	0-1	.12 (.15)	497
Property Offense Scale	0-1	.09 (.16)	497
Drug Scale	0-1	.14 (.21)	497
Psychological Scales			
Consumer Scale	1-7	5.38 (1.28)	466
Locus of Control Scale at admission	1-2	1.55 (.26)	462
Happiness Rating	0-2	.99 (.61)	497
Treatment			
Length of stay (days)	1-? ^a	755.08 (532.37)	497
Motivation system at departure	0-3	1.45 (1.24)	446
Post-treatment			
Favorable Departure Rating	0-4	2.27 (1.17)	473
Prediction of Future Success	0-4	2.24 (1.11)	469
Follow-up			
Psychological outcomes			
Mental health index	1-5	4.12 (.86)	207

Variable Name	Scale Range	Mean (<i>SD</i>)	<i>N</i>
Recent substance abuse	0-1	.42 (.49)	207
Locus of Control at follow-up	1-5	4.10 (.62)	206
Satisfaction Index	1-7	4.92 (1.17)	206
Happiness at follow-up	0-3	1.98 (.65)	207
Deceased	0-1	.05 (.21)	410
Criminality			
Arrested in past 12 months	0-1	.19 (.40)	207
Perpetration Index	0-? ^a	.81 (2.40)	206
Education			
Highest grade completed	7-20	12.46 (2.04)	206
In school past 12 months	0-1	.30 (.46)	206
Employment and earnings			
Currently working	0-1	.75 (.43)	205
Job Satisfaction	0-4	3.00 (1.05)	201
Household income	1-7	2.95 (1.85)	197
Could not pay bills	0-1	.31 (.46)	194
Military service	0-1	.21 (.41)	204
Social Systems			
Living with spouse/partner	0-1	.67 (.47)	200
Child Involvement Index	3-18	14.11 (4.34)	122
Social Involvement Count	0-4	2.56 (1.14)	200
Spirituality	0-3	1.83 (.86)	205
Evaluation of GBT			
GBT Index	0-4	3.60 (.66)	205

^aRanges that include a question mark (?) have no specified limit because they are counts or normalized scores.

Table 8

Logistic Regression Analyses of Variables Predicting Recent Substance Abuse (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	Wald's χ^2	<i>p</i>	<i>OR</i> ^a
Liberal model: $\chi^2 (3, n = 177) = 19.09, p < .001$					
Gender	2.35	1.05	5.02	.025	10.47
Property Offense Scale	3.18	1.22	6.81	.009	24.08
Motivation system	.028	.13	.048	.826	1.03
Conservative model: $\chi^2 (4, n = 177) = 19.16, p < .001$					
Gender	2.36	1.05	5.05	.025	10.56
Property Offense Scale	3.18	1.22	6.77	.009	23.94
Motivation system	-.01	.19	.002	.962	.99
Favorable Departure Rating	.05	.20	.07	.792	1.05

^a*OR* = odds ratio.

Table 9

Regression Analyses of Variables Predicting Locus of Control at Follow-Up (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .064$, $F(3, 172) = 3.93$, $p = .01$						
Minority status	-0.17	0.10	-0.13	0.090	0.016	0.96
Aggressive Offense Scale	-0.79	0.31	-0.19	0.013	0.035	1.00
Motivation system	0.06	0.04	0.13	0.094	0.015	0.96
Conservative model: $R^2 = .064$, $F(4, 171) = 2.93$, $p < .05$						
Minority status	-0.17	0.10	-0.13	0.091	0.016	0.96
Aggressive Offense Scale	-0.79	0.32	-0.19	0.013	0.034	0.99
Motivation system	0.06	0.05	0.12	0.279	0.006	0.42
Favorable Departure Rating	0.00	0.06	0.00	0.971	0.000	0.43

Table 10

Regression Analyses of Variables Predicting Satisfaction Index at Follow-Up (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .024$, $F(2, 173) = 2.15$, $p = .12$						
Drug Scale	-0.75	0.43	-0.13	0.084	0.017	0.99
Motivation system	0.09	0.07	0.10	0.208	0.009	0.99
Conservative model: $R^2 = .034$, $F(3, 172) = 2.01$, $p = .12$						
Drug Scale	-0.75	0.43	-0.13	0.084	0.017	0.99
Motivation system	-0.01	0.10	-0.02	0.888	0.000	0.43
Favorable Departure Rating	0.14	0.11	0.15	0.195	0.010	0.43

Table 11

Regression Analyses of Variables Predicting Happiness at Follow-Up (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .053$, $F(3, 173) = 3.23$, $p < .05$						
Number of presenting problems	-0.06	0.03	-0.14	0.076	0.017	0.94
Length of stay	+0.00	0.00	0.01	0.856	0.000	0.84
Motivation system	0.07	0.04	0.15	0.077	0.017	0.80
Conservative model: $R^2 = .074$, $F(4, 172) = 3.45$, $p < .01$						
Number of presenting problems	-0.05	0.03	-0.12	0.120	0.013	0.92
Length of stay	+0.00	0.00	0.02	0.800	0.000	0.83
Motivation system	-0.01	0.06	-0.02	0.872	0.000	0.39
Favorable Departure Rating	0.12	0.06	0.22	0.049	0.021	0.43

Table 12

*Logistic Regression Analyses of Variables Predicting whether Arrested in Past 12**Months (16-Year Follow-Up)*

Variable	<i>b</i>	<i>SE(b)</i>	Wald's χ^2	<i>p</i>	<i>OR</i> ^a
Liberal model: $\chi^2 (5, n = 177) = 17.91, p < .01$					
Minority status	-.68	.41	2.78	.096	.51
Age at follow-up	-.11	.19	1.05	.306	.90
Drug Scale	-1.59	1.24	1.66	.198	.20
Number of presenting problems	.20	.15	1.83	.176	1.22
Motivation system	-.27	.17	2.65	.104	.77
Conservative model: $\chi^2 (6, n = 177) = 18.57, p < .01$					
Minority status	-.69	.41	2.82	.093	.50
Age at follow-up	-.13	.11	1.33	.249	.88
Drug Scale	-1.57	1.24	1.59	.207	.21
Number of presenting problems	.21	.15	2.06	.152	1.23
Motivation system	-.42	.25	2.80	.094	.66
Favorable Departure Rating	.21	.26	.65	.420	1.24

^a*OR* = odds ratio.

Table 13

Regression Analyses of Variables Predicting Perpetration Index (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .115$, $F(5, 154) = 4.01$, $p < .01$						
Age at admission	-0.20	0.13	-0.13	0.125	0.014	0.76
ROLES at admission	-0.05	0.03	-0.14	0.067	0.020	0.96
Locus of Control Scale at admission	-1.39	0.73	-0.15	0.061	0.021	0.93
Length of stay	-0.00	0.00	-0.12	0.194	0.010	0.70
Motivation system	-0.18	0.18	-0.09	0.313	0.006	0.70
Conservative model: $R^2 = .115$, $F(6, 153) = 3.33$, $p < .01$						
Age at admission	-0.20	0.13	-0.13	0.130	0.013	0.75
ROLES at admission	-0.05	0.03	-0.14	0.067	0.020	0.96
Locus of Control Scale at admission	-1.38	0.74	-0.15	0.066	0.020	0.91
Length of stay	-0.00	0.00	-0.12	0.198	0.010	0.70
Motivation system	-0.16	0.25	-0.08	0.515	0.002	0.37
Favorable Departure Rating	-0.03	0.24	-0.01	0.907	0.000	0.40

Table 14

Regression Analyses of Variables Predicting Highest Grade Completed (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .103$, $F(3, 160) = 6.10$, $p < .001$						
Locus of Control Scale at admission	0.99	0.56	0.13	0.081	0.017	0.97
Length of stay	+0.00	0.00	0.08	0.324	0.005	0.83
Motivation system	0.37	0.13	0.23	0.006	0.043	0.81
Conservative model: $R^2 = .136$, $F(4, 159) = 6.27$, $p < .001$						
Locus of Control Scale at admission	0.77	0.56	0.10	0.170	0.010	0.95
Length of stay	+0.00	0.00	0.08	0.317	0.005	0.83
Motivation system	0.03	0.19	0.02	0.875	0.000	0.39
Favorable Departure Rating	0.47	0.19	0.29	0.014	0.034	0.41

Table 15

Logistic Regression Analyses of Variables Predicting whether In School in Past 12

Months (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	Wald's χ^2	<i>p</i>	<i>OR</i> ^a
Liberal model: $\chi^2 (2, n = 173) = 5.68, p = .059$					
ROLES at admission	0.06	0.03	4.88	0.027	1.06
Motivation system	0.02	0.14	0.02	0.879	1.02
Conservative model: $\chi^2 (3, n = 173) = 5.68, p = .13$					
ROLES at admission	0.06	0.03	4.87	0.027	1.06
Motivation system	0.02	0.21	0.01	0.912	1.02
Favorable Departure Rating	0.00	0.22	0.00	0.989	1.00

^a*OR* = odds ratio.

Table 16

Logistic Regression Analyses of Variables Predicting whether Currently Working (16-Year Follow-Up)

Variable	b	$SE(b)$	Wald's χ^2	p	OR^a
Liberal model: $\chi^2 (4, n = 173) = 14.84, p < .01$					
Minority status	0.83	0.39	4.44	0.035	2.28
ROLES at admission	0.05	0.02	4.91	0.027	1.05
Length of stay ^b	+0.00	0.00	3.00	0.083	> 1.00
Motivation system	0.03	0.16	0.03	0.865	1.03
Conservative model: $\chi^2 (5, n = 173) = 15.27, p < .01$					
Minority status	0.83	0.39	4.44	0.035	2.28
ROLES at admission	0.05	0.02	5.04	0.025	1.05
Length of stay ^b	+0.00	0.00	3.06	0.080	> 1.00
Motivation system	-0.08	0.23	0.12	0.725	0.92
Favorable Departure Rating	0.15	0.23	0.43	0.511	1.16

^a OR = odds ratio. ^bBecause length of stay is measured in days and the standard deviation

is so large (755 days), b is still very close to 0 and OR is close to 1 even when length of stay is significant.

Table 17

Regression Analyses of Variables Predicting Household Income (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .096$, $F(3, 164) = 5.80$, $p < .001$						
Number of presenting problems	-0.25	0.10	-0.19	0.016	0.033	0.93
Length of stay	+0.00	0.00	0.02	0.845	0.000	0.83
Motivation system	0.29	0.12	0.20	0.019	0.031	0.80
Conservative model: $R^2 = .097$, $F(4, 163) = 4.39$, $p < .01$						
Number of presenting problems	-0.24	0.10	-0.18	0.019	0.031	0.92
Length of stay	+0.00	0.00	0.02	0.829	0.000	0.83
Motivation system	0.23	0.18	0.15	0.202	0.009	0.38
Favorable Departure Rating	0.09	0.18	0.06	0.624	0.001	0.42

Table 18

Logistic Regression Analyses of Variables Predicting Military Service (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	Wald's χ^2	<i>p</i>	<i>OR</i> ^a
Liberal model: $\chi^2 (3, n = 175) = 26.52, p < .001$					
Happiness Rating	-0.72	0.35	4.13	0.042	0.49
Length of stay	+0.00	0.00	1.09	0.296	> 1.00
Motivation system	0.65	0.20	10.46	0.001	1.91
Conservative model: $\chi^2 (4, n = 175) = 27.07, p < .001$					
Happiness Rating	-0.72	0.35	4.21	0.040	0.48
Length of stay	+0.00	0.00	1.17	0.279	> 1.00
Motivation system	0.51	0.27	3.61	0.057	1.67
Favorable Departure Rating	0.19	0.26	0.54	0.462	1.21

^a*OR* = odds ratio.

Table 19

Logistic Regression Analyses of Variables Predicting whether Currently Living with a Spouse or Partner (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	Wald's χ^2	<i>p</i>	<i>OR</i> ^a
Liberal model: $\chi^2 (3, n = 171) = 9.46, p < .05$					
Age at follow-up	0.23	0.09	6.02	0.014	1.26
Happiness Rating	0.30	0.28	1.10	0.293	1.35
Motivation system	0.07	0.14	0.25	0.619	1.07
Conservative model: $\chi^2 (4, n = 171) = 9.95, p < .05$					
Age at follow-up	0.24	0.10	6.33	0.012	1.27
Happiness Rating	0.30	0.28	1.12	0.290	1.35
Motivation system	0.18	0.21	0.73	0.394	1.20
Favorable Departure Rating	-0.16	0.22	0.49	0.484	0.85

^a*OR* = odds ratio.

Table 20

Regression Analyses of Variables Predicting Spirituality (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .023$, $F(2, 172) = 2.05$, $p = .13$						
Aggressive Offense Scale	0.83	0.46	0.13	0.076	0.018	1.00
Motivation system	0.05	0.05	0.08	0.312	0.006	1.00
Conservative model: $R^2 = .028$, $F(3, 171) = 1.66$, $p = .18$						
Aggressive Offense Scale	0.86	0.46	0.14	0.065	0.020	0.99
Motivation system	-0.00	0.08	-0.01	0.960	0.000	0.43
Favorable Departure Rating	0.08	0.08	0.11	0.346	0.005	0.43

Table 21

Regression Analyses of Variables Predicting GBT Index (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .064$, $F(3, 171) = 3.88$, $p < .05$						
Minority status	-0.17	0.10	-0.12	0.106	0.014	0.95
Length of stay ^a	+0.00	0.00	0.18	0.026	0.028	0.82
Motivation system	0.03	0.04	0.07	0.414	0.004	0.80
Conservative model: $R^2 = .079$, $F(4, 170) = 3.66$, $p < .01$						
Minority status	-0.17	0.10	-0.13	0.093	0.015	0.95
Length of stay ^a	+0.00	0.00	0.19	0.023	0.029	0.82
Motivation system	-0.04	0.06	-0.08	0.520	0.002	0.39
Favorable Departure Rating	0.10	0.06	0.19	0.092	0.016	0.43

^aBecause length of stay is measured in days and the standard deviation is so large (755

days), *b* is still very close to 0 even when length of stay is significant. β is a better

indicator of strength of association.

Table 22

Regression Analysis of Variables Predicting Motivation System (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Model: $R^2 = .100$, $F(8, 344) = 4.78$, $p < .001$						
Gender	-0.46	0.25	-0.10	0.071	0.009	0.91
Minority status	0.25	0.14	0.09	0.084	0.008	0.96
Age at Admission	0.09	0.04	0.12	0.021	0.014	0.98
ROLES at Admission	0.01	0.01	0.08	0.166	0.005	0.89
Number of prior placements	-0.02	0.02	-0.05	0.351	0.002	0.92
Aggressive Offense Scale	-0.55	0.48	-0.06	0.249	0.003	0.83
Property Offense Scale	-0.44	0.44	-0.06	0.323	0.003	0.81
Number of presenting problems	-0.14	0.05	-0.16	0.003	0.024	0.91

Table 23

Regression Analysis of Variables Predicting Favorable Departure Rating (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>P</i>	<i>sr</i> ²	Tolerance
Model: $R^2 = .110$, $F(7, 369) = 6.49$, $p < .001$						
Gender	-0.58	0.23	-0.13	0.012	0.015	0.92
Minority status	0.21	0.13	0.08	0.097	0.007	0.97
Age at Admission	0.09	0.04	0.13	0.012	0.015	0.96
ROLES at Admission	0.01	0.01	0.07	0.204	0.004	0.87
Number of prior placements	-0.03	0.02	-0.09	0.096	0.007	0.91
Aggressive Offense Scale	-0.62	0.39	-0.08	0.116	0.006	0.99
Number of presenting problems	-0.13	0.04	-0.15	0.004	0.021	0.91

Table 24

Regression Analysis of Variables Predicting Length of Stay (16-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Model: $R^2 = .146$, $F(8, 420) = 9.00$, $p < .001$						
Gender	117.45	91.33	0.06	0.199	0.003	0.97
African American	109.35	63.45	0.08	0.086	0.006	0.94
Age at admission	-103.05	16.62	-0.30	0.000	0.078	0.85
Property Offense Scale	-153.21	184.09	-0.04	0.406	0.001	0.72
Drug Scale	-137.37	147.39	-0.05	0.352	0.002	0.67
Estimated GPA	1.87	24.30	0.00	0.939	0.000	0.90
Consumer Scale	16.51	20.97	0.04	0.432	0.001	0.83
Happiness Rating	-55.92	43.36	-0.06	0.198	0.003	0.84
Analysis with gender, the Property Offense Scale, estimated GPA, and Happiness Rating removed						
Model: $R^2 = .137$, $F(4, 461) = 18.258$, $p < .001$						
African American	141.33	59.65	0.10	0.018	0.011	0.98
Age at admission	-95.61	15.17	-0.28	0.000	0.074	0.92
Drug Scale	-233.86	118.44	-0.09	0.049	0.007	0.89
Consumer Scale	36.70	18.62	0.09	0.049	0.007	0.95

Table 25

Descriptive Statistics for the 5-Year Follow-Up Sample

Variable Name	Scale Range	Mean (<i>SD</i>)	<i>n</i>
Demographics			
IQ	? ^a	94.62 (11.85)	151
Pre-treatment			
History			
ROLES at admission	1-25	15.71 (7.01)	178
Number of formal prior placements	0-? ^a	2.86 (3.35)	179
Number of presenting problems	0-38	9.98 (4.73)	212
Number of family problems	0-35	6.79 (3.93)	203
Ever arrested	0-1	.62 (.49)	179
Psychopathology at admission (DISC)			
Total number of diagnoses	0-? ^a	1.61 (1.89)	205
Number of internalizing diagnoses	0-? ^a	.45 (1.11)	205
Number of externalizing diagnoses	0-? ^a	.53 (.72)	205
Any substance abuse diagnosis	0-1	.36 (.48)	204
Treatment			
Length of stay (days)	1-? ^a	584.04 (507.22)	214
Motivation system at departure	0-3	2.07 (1.09)	212
Post-treatment			
Presenting problems percentage improved upon	0-100	89.26 (19.41)	204
ROLES at departure	1-25	18.75 (6.22)	213
Favorable Departure Rating	4-28	19.98 (5.84)	214
Departure Success Scale	5-35	24.97 (6.97)	213

Variable Name	Scale Range	Mean (<i>SD</i>)	<i>n</i>
Prediction of Future Success	1-7	4.74 (1.39)	214
Follow-up			
Psychological outcomes			
Mental health index	1-4	3.10 (.50)	131
Recent substance abuse	0-1	.30 (.46)	131
Friends' substance use	0-3	.84 (.61)	126
Criminality			
Perpetrated any crime in past 12 months	0-1	.28 (.45)	130
Arrested in past 12 months	0-1	.20 (.40)	131
Currently incarcerated	0-1	.11 (.31)	131
Education			
Highest grade completed	9-17	12.25 (1.48)	129
In school past 12 months	0-1	.57 (.50)	131
Employment and earnings			
Worked last week	0-1	.59 (.49)	131
Household income	1-7	3.08 (1.40)	123
Military service	0-1	.09 (.29)	130
Social Systems			
Living with spouse/partner	0-1	.39 (.49)	130
Closeness to spouse/partner	0-10	8.95 (1.94)	57
Conflict with spouse/partner	0-10	6.67 (2.59)	58
Average child involvement	0-4	2.95 (1.48)	62
Support from at least one parent	0-1	.89 (.32)	131
Religion important	1-4	3.11 (.77)	122

^aRanges that include a question mark (?) have no specified limits.

Table 26

Chi-Square Table Comparing Responders to Nonresponders on Gender (5-Year Follow-Up)

Gender	Responder status		Total
	Nonresponder	Responder	
Female	17	65	82
Male	66	66	132
Total	83	131	214

Note. $\chi^2(1, n = 214) = 18.25, p < .001$.

Table 27

Chi-Square Table Comparing Responders to Nonresponders on Ethnicity (5-Year Follow-Up)

Ethnicity	Responder status		Total
	Nonresponder	Responder	
African American	19	30	49
Caucasian	42	83	125
Hispanic	16	6	22
Other	6	12	18
Total	83	131	214

Note. $\chi^2(3, n = 214) = 12.32, p < .01$.

Table 28

Regression Analyses of Variables Predicting Mental Health Index (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .121$, $F(2, 121) = 8.32$, $p < .001$						
Number of internalizing diagnoses	-0.14	0.04	-0.32	0.000	0.102	1.00
Motivation system	0.06	0.04	0.13	0.131	0.017	1.00
Conservative model: $R^2 = .125$, $F(3, 120) = 5.74$, $p < .01$						
Number of internalizing diagnoses	-0.13	0.04	-0.31	0.000	0.094	0.98
Motivation system	0.03	0.06	0.05	0.678	0.001	0.44
Favorable Departure Rating	0.01	0.01	0.10	0.433	0.005	0.43

Table 29

Logistic Regression Analyses of Variables Predicting Recent Substance Abuse (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	Wald's χ^2	<i>p</i>	<i>OR</i> ^a
Liberal model: $\chi^2 (3, n = 131) = 9.00, p < .05$					
Age at departure	-0.28	0.16	3.17	0.075	0.76
Length of stay ^b	-0.00	0.00	2.50	0.114	< 1.00
Motivation system	0.21	0.22	0.87	0.351	1.23
Conservative model: $\chi^2 (4, n = 131) = 9.67, p < .05$					
Age at departure	-0.28	0.16	3.10	0.078	0.76
Length of stay ^b	-0.00	0.00	2.57	0.109	< 1.00
Motivation system	0.02	0.32	0.00	0.951	1.02
Favorable Departure Rating	0.05	0.06	0.66	0.415	1.05
Analyses with outlier removed					
Liberal model: $\chi^2 (3, n = 130) = 11.57, p < .01$					
Age at departure	-0.19	0.16	1.32	0.251	0.83
Length of stay ^b	-0.00	0.00	4.87	0.027	< 1.00
Motivation system	0.26	0.23	1.29	0.257	1.30
Conservative model: $\chi^2 (4, n = 130) = 12.29, p < .05$					
Age at departure	-0.18	0.16	1.24	0.265	0.84
Length of stay ^b	-0.00	0.00	4.95	0.026	< 1.00
Motivation system	0.06	0.32	0.04	0.852	1.06
Favorable Departure Rating	0.05	0.06	0.71	0.401	1.05

^a*OR* = odds ratio. ^bBecause length of stay is measured in days and the standard deviation is so large (755 days), *b* is still very close to 0 and *OR* is close to 1 even when length of stay is significant.

Table 30

Regression Analyses of Variables Predicting Friends' Substance Use (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .200$, $F(6, 97) = 4.05$, $p < .01$						
Gender	0.19	0.11	0.16	0.095	0.023	0.91
Hispanic	0.22	0.30	0.07	0.459	0.005	0.90
Ever arrested	0.25	0.12	0.20	0.042	0.035	0.88
Any substance abuse diagnosis	0.25	0.12	0.20	0.044	0.034	0.85
Presenting problems percent improved upon	-0.01	0.00	-0.16	0.190	0.014	0.55
Motivation system	0.00	0.07	0.01	0.963	0.000	0.57
Conservative model: $R^2 = .200$, $F(7, 96) = 3.44$, $p < .01$						
Gender	0.20	0.12	0.16	0.095	0.024	0.90
Hispanic	0.22	0.30	0.07	0.458	0.005	0.90
Ever arrested	0.25	0.12	0.20	0.043	0.035	0.88
Any substance abuse diagnosis	0.25	0.12	0.20	0.045	0.034	0.84
Presenting problems percent improved upon	-0.01	0.00	-0.17	0.212	0.013	0.47
Motivation system	0.00	0.09	-0.01	0.970	0.000	0.37
Favorable Departure Rating	0.00	0.02	0.02	0.899	0.000	0.33

Table 31

Logistic Regression Analyses of Variables Predicting whether Perpetrated Any Crime in the Past 12 Months (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	Wald's χ^2	<i>p</i>	<i>OR</i> ^a
Liberal model: $\chi^2 (4, n = 109) = 20.08, p < .001$					
Age at departure	-0.43	0.18	5.77	0.016	0.65
Ever arrested	1.06	0.54	3.90	0.048	2.90
ROLES at departure	-0.07	0.04	2.75	0.097	0.93
Motivation system	-0.05	0.26	0.03	0.853	0.95
Conservative model: $\chi^2 (5, n = 109) = 20.10, p < .01$					
Age at departure	-0.44	0.18	5.80	0.016	0.65
Ever arrested	1.07	0.54	3.91	0.048	2.93
ROLES at departure	-0.07	0.04	2.76	0.096	0.93
Motivation system	-0.08	0.35	0.06	0.811	0.92
Favorable Departure Rating	0.01	0.06	0.02	0.880	1.01

^a*OR* = odds ratio.

Table 32

Logistic Regression Analyses of Variables Predicting whether Arrested in the Past 12 Months (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	Wald's χ^2	<i>p</i>	<i>OR</i> ^a
Liberal model: $\chi^2 (4, n = 128) = 12.59, p < .05$					
Age at departure	-0.37	0.18	4.54	0.033	0.69
Presenting problems percent improved	-0.02	0.01	2.06	0.151	0.98
ROLES at departure	-0.06	0.05	1.69	0.194	0.94
Motivation system	0.20	0.33	0.37	0.543	1.22
Conservative model: $\chi^2 (5, n = 128) = 12.62, p < .05$					
Age at departure	-0.37	0.18	4.53	0.033	0.69
Presenting problems percent improved	-0.02	0.01	1.93	0.165	0.98
ROLES at departure	-0.06	0.05	1.70	0.192	0.94
Motivation system	0.17	0.39	0.18	0.669	1.18
Favorable Departure Rating	0.01	0.07	0.02	0.876	1.01

^a*OR* = odds ratio.

Table 33

Logistic Regression Analyses of Variables Predicting Currently Incarcerated (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	Wald's χ^2	<i>p</i>	<i>OR</i> ^a
Liberal model: $\chi^2 (4, n = 128) = 21.13, p < .001$					
Gender	2.74	1.08	6.45	0.011	15.51
Presenting problems percent Improved	-0.02	0.02	1.40	0.237	0.98
ROLES at departure	-0.10	0.05	3.25	0.071	0.91
Motivation system	0.11	0.37	0.08	0.774	1.11
Conservative model: $\chi^2 (5, n = 128) = 22.47, p < .001$					
Gender	2.85	1.09	6.80	0.009	17.26
Presenting problems percent Improved	-0.03	0.02	2.27	0.131	0.97
ROLES at departure	-0.11	0.06	3.71	0.054	0.89
Motivation system	-0.23	0.47	0.24	0.626	0.79
Favorable Departure Rating	0.11	0.10	1.20	0.273	1.12

^a*OR* = odds ratio.

Table 34

Regression Analyses of Variables Predicting Highest Grade Completed (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .171$, $F(5, 82) = 3.39$, $p < .01$						
Age at departure	0.28	0.11	0.30	0.009	0.072	0.82
IQ	0.02	0.01	0.16	0.127	0.024	0.95
Length of stay	0.00	0.00	0.00	0.997	0.000	0.78
Presenting problems percent improved upon	0.01	0.01	0.11	0.362	0.008	0.73
Motivation system	0.06	0.17	0.04	0.733	0.001	0.63
Conservative model: $R^2 = .172$, $F(6, 81) = 2.80$, $p < .05$						
Age at departure	0.28	0.11	0.30	0.009	0.072	0.82
IQ	0.02	0.01	0.16	0.127	0.024	0.95
Length of stay	0.00	0.00	0.00	0.973	0.000	0.77
Presenting problems percent improved upon	0.01	0.01	0.12	0.360	0.009	0.58
Motivation system	0.09	0.21	0.07	0.678	0.002	0.40
Favorable Departure Rating	-0.01	0.04	-0.04	0.810	0.001	0.32

Table 35

Logistic Regression Analyses of Variables Predicting whether In School in the Past 12 Months (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	Wald's χ^2	<i>p</i>	<i>OR</i> ^a
Liberal model: $\chi^2 (2, n = 110) = 4.75, p = .093$					
Ever arrested	-0.80	0.43	3.41	0.065	0.45
Motivation system	0.15	0.19	0.62	0.432	1.16
Conservative model: $\chi^2 (3, n = 110) = 5.00, p = .17$					
Ever arrested	-0.82	0.43	3.52	0.061	0.44
Motivation system	0.26	0.29	0.78	0.378	1.29
Favorable Departure Rating	-0.03	0.05	0.25	0.619	0.97

^a*OR* = odds ratio.

Table 36

Regression Analyses of Variables Predicting Household Income (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .074$, $F(3, 107) = 2.83$, $p < .05$						
Number of family problems	-0.06	0.03	-0.17	0.071	0.029	1.00
Number of internalizing diagnoses	-0.21	0.10	-0.19	0.042	0.037	1.00
Motivation system	0.09	0.12	0.07	0.441	0.005	1.00
Conservative model: $R^2 = .075$, $F(4, 106) = 2.13$, $p = .082$						
Number of family problems	-0.06	0.03	-0.17	0.068	0.030	0.97
Number of internalizing diagnoses	-0.21	0.10	-0.20	0.041	0.038	0.98
Motivation system	0.14	0.19	0.11	0.447	0.005	0.43
Favorable Departure Rating	-0.01	0.04	-0.05	0.734	0.001	0.42

Table 37

Logistic Regression Analyses of Variables Predicting Military Service (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	Wald's χ^2	<i>p</i>	<i>OR</i> ^a
Liberal model: $\chi^2 (6, n = 107) = 37.89, p < .001$					
Gender	3.90	2.08	3.52	0.060	49.27
Age at departure	0.65	0.68	0.92	0.336	1.92
ROLES at admission	0.26	0.16	2.61	0.106	1.29
Ever arrested	-3.67	1.50	6.00	0.014	0.03
Number of externalizing diagnoses	2.87	1.37	4.39	0.036	17.64
Motivation system	4.69	3.05	2.36	0.125	108.39
Conservative model: $\chi^2 (7, n = 107) = 37.93, p < .001$					
Gender	3.97	2.13	3.48	0.062	53.21
Age at departure	0.66	0.69	0.91	0.340	1.94
ROLES at admission	0.26	0.17	2.46	0.117	1.30
Ever arrested	-3.64	1.47	6.10	0.014	0.03
Number of externalizing diagnoses	2.85	1.37	4.37	0.037	17.35
Motivation system	4.93	3.29	2.25	0.134	138.89
Favorable Departure Rating	-0.05	0.24	0.04	0.838	0.95

^a*OR* = odds ratio.

Table 38

Logistic Regression Analyses of Variables Predicting Currently Living with a Spouse or Partner (5-Year Follow-Up)

Variable	b	$SE(b)$	Wald's χ^2	p	OR^a
Liberal model: $\chi^2 (4, n = 107) = 13.94, p < .01$					
Minority status	0.69	0.47	2.13	0.144	1.99
Ever arrested	0.74	0.49	2.32	0.128	2.10
Any substance abuse diagnosis	1.01	0.44	5.22	0.022	2.74
Motivation system	-0.11	0.20	0.32	0.571	0.89
Conservative model: $\chi^2 (5, n = 107) = 14.43, p < .05$					
Minority status	0.64	0.48	1.79	0.181	1.89
Ever arrested	0.76	0.49	2.41	0.121	2.14
Any substance abuse diagnosis	1.03	0.44	5.42	0.020	2.81
Motivation system	-0.29	0.32	0.80	0.370	0.75
Favorable Departure Rating	0.04	0.06	0.48	0.487	1.04

^a OR = odds ratio.

Table 39

Regression Analyses of Variables Predicting Closeness to Spouse or Partner (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .248$, $F(3, 53) = 5.82$, $p < .01$						
“Other” ethnicity	-3.18	1.03	-0.37	0.003	0.135	0.99
Age at departure	0.43	0.18	0.31	0.021	0.081	0.85
Motivation system	-0.07	0.23	-0.04	0.772	0.001	0.86
Conservative model: $R^2 = .249$, $F(4, 52) = 4.31$, $p < .01$						
“Other” ethnicity	-3.20	1.04	-0.37	0.003	0.136	0.98
Age at departure	0.42	0.18	0.30	0.026	0.076	0.83
Motivation system	-0.11	0.28	-0.06	0.693	0.002	0.58
Favorable Departure Rating	0.02	0.06	0.04	0.781	0.001	0.60
Analyses with 2 outliers removed						
Liberal model: $R^2 = .052$, $F(3, 51) = .94$, <i>n.s.</i>						
“Other” ethnicity	-2.31	1.57	-0.20	0.148	0.040	0.98
Age at departure	-0.01	0.20	-0.01	0.940	0.000	0.78
Motivation system	0.20	0.22	0.14	0.375	0.015	0.78
Conservative model: $R^2 = .053$, $F(4, 50) = .69$, <i>n.s.</i>						
“Other” ethnicity	-2.30	1.59	-0.20	0.154	0.040	0.98
Age at departure	-0.02	0.20	-0.01	0.929	0.000	0.77
Motivation system	0.17	0.27	0.12	0.519	0.008	0.53
Favorable Departure Rating	0.01	0.05	0.03	0.885	0.000	0.59

Table 40

Regression Analyses of Variables Predicting Conflict with Spouse or Partner (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .164$, $F(3, 48) = 3.15$, $p < .05$						
Number of family problems	-0.15	0.08	-0.24	0.085	0.054	0.96
Total number of diagnoses	-0.49	0.25	-0.27	0.053	0.068	0.91
Motivation system	0.11	0.34	0.04	0.752	0.002	0.95
Conservative model: $R^2 = .172$, $F(4, 47) = 2.45$, $p = .059$						
Number of family problems	-0.15	0.08	-0.23	0.092	0.052	0.96
Total number of diagnoses	-0.48	0.25	-0.27	0.061	0.065	0.90
Motivation system	-0.07	0.43	-0.03	0.872	0.000	0.59
Favorable Departure Rating	0.06	0.09	0.11	0.508	0.008	0.60

Table 41

Regression Analyses of Variables Predicting Average Child Involvement (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .149$, $F(3, 57) = 3.33$, $p < .05$						
Gender	-0.88	0.37	-0.29	0.022	0.083	0.98
Number of family problems	0.08	0.04	0.23	0.072	0.050	0.97
Motivation system	-0.04	0.18	-0.03	0.834	0.001	0.98
Conservative model: $R^2 = .157$, $F(4, 56) = 2.62$, $p < .05$						
Gender	-0.87	0.38	-0.29	0.024	0.081	0.98
Number of family problems	0.08	0.04	0.24	0.065	0.053	0.97
Motivation system	-0.15	0.24	-0.10	0.524	0.006	0.58
Favorable Departure Rating	0.03	0.04	0.12	0.454	0.009	0.59

Table 42

*Logistic Regression Analyses of Variables Predicting Support from at Least One Parent
(5-Year Follow-Up)*

Variable	<i>b</i>	<i>SE(b)</i>	Wald's χ^2	<i>p</i>	<i>OR</i> ^a
Liberal model: $\chi^2 (3, n = 106) = 10.63, p < .05$					
Number of family problems	-0.16	0.07	5.16	0.023	0.85
Ever arrested	1.46	0.71	4.22	0.040	4.32
Motivation system	0.09	0.35	0.07	0.798	1.09
Conservative model: $\chi^2 (4, n = 106) = 10.80, p < .05$					
Number of family problems	-0.16	0.07	5.09	0.024	0.85
Ever arrested	1.49	0.72	4.29	0.038	4.44
Motivation system	-0.06	0.51	0.01	0.908	0.94
Favorable Departure Rating	0.04	0.09	0.16	0.685	1.04
Analyses with 1 outlier removed					
Liberal model: $\chi^2 (3, n = 105) = 9.02, p < .05$					
Number of family problems	-0.11	0.08	2.11	0.146	0.89
Ever arrested	1.65	0.74	4.92	0.027	5.20
Motivation system	-0.01	0.37	0.00	0.970	0.99
Conservative model: $\chi^2 (4, n = 105) = 10.10, p < .05$					
Number of family problems	-0.10	0.08	1.82	0.178	0.90
Ever arrested	1.77	0.77	5.31	0.021	5.85
Motivation system	-0.45	0.57	0.62	0.432	0.64
Favorable Departure Rating	0.10	0.10	1.10	0.295	1.11

^a*OR* = odds ratio.

Table 43

Regression Analyses of Variables Predicting the Importance of Religion (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Liberal model: $R^2 = .149$, $F(2, 78) = 6.49$, $p < .01$						
IQ	-0.03	0.01	-0.38	0.001	0.136	0.94
Motivation system	0.00	0.08	0.01	0.957	0.000	0.94
Conservative model: $R^2 = .171$, $F(3, 77) = 5.28$, $p < .01$						
IQ	-0.03	0.01	-0.40	0.000	0.148	0.93
Motivation system	-0.16	0.13	-0.21	0.227	0.016	0.38
Favorable Departure Rating	0.04	0.02	0.27	0.110	0.028	0.37

Table 44

Hierarchical Regression Analysis of Variables Predicting Motivation System (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Model: $R^2 = .052$, $F(1, 137) = 7.58$, $p < .01$						
Gender	-0.53	0.19	-0.23	0.007	0.052	1.00
Step 2: $\Delta R^2 = .025$, $F(1, 136) = 3.75$, $p = .055$						
Gender	-0.55	0.19	-0.24	0.005	0.056	1.00
IQ	0.02	0.01	0.16	0.055	0.025	1.00
Step 3: $\Delta R^2 = .057$, $F(8, 128) = 1.06$, <i>n.s.</i>						
Gender	-0.60	0.20	-0.26	0.003	0.060	0.89
IQ	0.01	0.01	0.15	0.125	0.016	0.76
Minority status	0.20	0.22	0.09	0.358	0.006	0.77
Age at admission	0.10	0.07	0.12	0.163	0.013	0.85
ROLES at admission	0.01	0.01	0.08	0.384	0.005	0.81
Number of formal prior placements	-0.03	0.03	-0.11	0.243	0.009	0.80
Number of presenting problems	-0.02	0.02	-0.11	0.302	0.007	0.66
Number of family problems	0.03	0.03	0.10	0.321	0.007	0.72
Ever arrested	-0.14	0.20	-0.06	0.486	0.003	0.87
Total number of diagnoses	-0.03	0.05	-0.05	0.558	0.002	0.87

Table 45

*Hierarchical Regression Analysis of Variables Predicting Favorable Departure Rating
(5-Year Follow-Up)*

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Model: $R^2 = .121$, $F(2, 138) = 9.46$, $p < .001$						
Gender	-3.90	1.02	-0.31	0.000	0.093	0.99
Number of formal prior placements	-0.33	0.14	-0.19	0.019	0.036	0.99
Step 2: $\Delta R^2 = .050$, $F(8, 130) = .986$, <i>n.s.</i>						
Gender	-3.93	1.08	-0.31	0.000	0.084	0.89
Number of formal prior placements	-0.39	0.16	-0.22	0.014	0.040	0.80
Minority status	1.64	1.18	0.13	0.169	0.012	0.76
Age at admission	0.51	0.37	0.12	0.177	0.012	0.85
IQ	0.04	0.05	0.08	0.414	0.004	0.76
ROLES at admission	0.05	0.08	0.05	0.538	0.002	0.81
Number of presenting problems	0.00	0.13	0.00	0.986	0.000	0.66
Number of family problems	0.05	0.15	0.03	0.732	0.001	0.72
Ever arrested	-0.09	1.09	-0.01	0.932	0.000	0.86
Total number of diagnoses	-0.17	0.29	-0.05	0.550	0.002	0.87

Table 46

Hierarchical Regression Analysis of Variables Predicting Departure Success Scale (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Model: $R^2 = .124$, $F(2, 138) = 9.74$, $p < .001$						
Gender	-4.59	1.21	-0.30	0.000	0.091	0.99
Number of formal prior placements	-0.43	0.17	-0.20	0.012	0.042	0.99
Step 2: $\Delta R^2 = .055$, $F(8, 130) = 1.09$, <i>n.s.</i>						
Gender	-4.67	1.28	-0.31	0.000	0.084	0.89
Number of formal prior placements	-0.48	0.19	-0.23	0.011	0.042	0.80
Minority status	1.96	1.40	0.13	0.165	0.012	0.76
Age at admission	0.58	0.44	0.11	0.194	0.011	0.85
IQ	0.08	0.09	0.07	0.413	0.004	0.81
ROLES at admission	0.05	0.06	0.09	0.341	0.006	0.76
Number of presenting problems	-0.03	0.15	-0.02	0.860	0.000	0.66
Number of family problems	0.06	0.17	0.03	0.737	0.001	0.72
Ever arrested	-0.44	1.29	-0.03	0.736	0.001	0.86
Total number of diagnoses	-0.17	0.34	-0.04	0.619	0.002	0.87

Table 47

Regression Analysis of Variables Predicting Length of Stay (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Model: $R^2 = .075$, $F(5, 169) = 2.73$, $p < .05$						
Hispanic	60.13	76.94	0.06	0.436	0.003	0.97
Age at admission	-41.45	16.90	-0.19	0.015	0.033	0.92
Number of formal prior placements	-12.02	7.50	-0.13	0.111	0.014	0.88
Number of presenting problems	-4.24	5.33	-0.06	0.428	0.003	0.88
Any substance abuse diagnosis	18.08	49.07	0.03	0.713	0.001	0.93

Table 48

Regression Analysis of Variables Predicting ROLES at departure (5-Year Follow-Up)

Variable	<i>b</i>	<i>SE(b)</i>	β	<i>p</i>	<i>sr</i> ²	Tolerance
Model: $R^2 = .077$, $F(3, 137) = 3.80$, $p < .05$						
Number of presenting problems	-0.02	0.12	-0.02	0.858	0.000	0.87
ROLES at admission	0.16	0.08	0.17	0.056	0.025	0.90
Number of formal prior placements	-0.33	0.17	-0.17	0.054	0.025	0.85
Step 2: $\Delta R^2 = .084$, $F(7, 130) = 1.85$, $p = .083$						
Number of presenting problems	-0.15	0.14	-0.10	0.299	0.007	0.66
ROLES at admission	0.17	0.09	0.18	0.046	0.026	0.81
Number of formal prior placements	-0.38	0.17	-0.20	0.025	0.033	0.80
Gender	-2.68	1.17	-0.20	0.023	0.034	0.89
Minority status	0.54	1.28	0.04	0.675	0.001	0.76
Age at admission	0.23	0.40	0.05	0.562	0.002	0.85
IQ	0.06	0.05	0.11	0.239	0.009	0.76
Ever arrested	-1.50	1.18	-0.11	0.205	0.010	0.86
Number of family problems	0.13	0.16	0.08	0.416	0.004	0.72
Total number of diagnoses	0.22	0.31	0.06	0.477	0.003	0.87

Table 49

Summary of Primary Analyses for the 16-Year Follow-Up Sample

Variable Name	Significant Predictors	MS & FDR
Psychological outcomes		
Mental health index	None	Neither
Recent substance abuse	Property Offense Scale** Gender (males = higher)*	Neither
Locus of Control at follow-up	Aggressive Offense Scale* Minority Status (minorities = more internal) ⁺	MS ⁺ in L only
Satisfaction Index	None	Neither
Happiness at follow-up	None	MS ⁺ in L only FDR* in C
Deceased	None	Neither
Criminality		
Arrested in past 12 months	Minority status (minorities = more likely) ⁺	MS ⁺ in C only
Perpetration Index	Locus of Control Scale at admission ⁺ ROLES at admission ⁺	Neither
Education		
Highest grade completed	None	MS** in L only FDR* in C
In school past 12 months	ROLES at admission*	Neither
Employment and earnings		
Currently working	ROLES at admission* Minority status (minorities = less likely)* Length of stay (Caucasians = more likely) ⁺	Neither

Variable Name	Significant Predictors	MS & FDR
Job Satisfaction	None	Neither
Household income	Number of presenting problems*	MS* in L only
Could not pay bills	None	Neither
Military service	Happiness Rating (happier = less likely)*	MS** in L MS ⁺ in C
Social Systems		
Living with spouse/partner	Age at follow-up (older = more likely)*	Neither
Child Involvement Index	None	Neither
Social Involvement Count	None	Neither
Spirituality	None	Neither
Evaluation of GBT		
GBT Index	Length of stay* Minority status (minorities = higher) ⁺	FDR ⁺ in C

Note. Only predictors from the liberal model for each dependent variable are reported unless Favorable Departure Rating changed the model substantially in the conservative analysis. Outcomes from outlier analyses, when different, are not included. Significant predictors were related to the outcome variable in the expected direction (i.e., better pre-treatment functioning predicted better follow-up outcomes, and the inverse) unless otherwise noted. MS = Motivation system at departure. FDR = Favorable Departure Rating. L = liberal analysis. C = conservative analysis.

** $p < .01$ * $p < .05$ ⁺ $p < .10$

Table 50

Summary of Primary Analyses for the 5-Year Follow-Up Sample

Variable Name	Significant Predictors	MS & FDR
Psychological outcomes		
Mental health index	Number of internalizing diagnoses**	Neither
Recent substance abuse	Age at departure (older = less likely) ⁺	Neither
Friends' substance use	Ever arrested* Any substance abuse diagnosis* Gender (males = higher) ⁺	Neither
Criminality		
Perpetrated any crime in past 12 months	Age at departure (older = lower)* Ever arrested* ROLES at departure ⁺	Neither
Arrested in past 12 months	Age at departure (older = less likely)*	Neither
Currently incarcerated	Gender (males = more likely)* ROLES at departure ⁺	Neither
Education		
Highest grade completed	Age at departure (older = higher)**	Neither
In school past 12 months	Ever arrested ⁺	Neither
Employment and earnings		
Worked last week	None	Neither
Household income	Number of internalizing diagnoses* Number of family problems ⁺	Neither
Military service	Ever arrested* Number of externalizing diagnoses (higher = more likely)* Gender (males = more likely) ⁺	MS* (even when controlling for FDR)

Variable Name	Significant Predictors	MS & FDR
Social Systems		
Living with spouse/partner	Any substance abuse diagnosis (SA diagnosis = more likely)*	Neither
Closeness to spouse/partner	“Other” ethnicity (other = lower)** Age at departure (older = higher)*	Neither
Conflict with spouse/partner	Total number of diagnoses ⁺ Number of family problems ⁺	Neither
Average child involvement	Gender (females = higher)* Number of family problems (more FP’s = higher) ⁺	Neither
Support from at least one parent	Number of family problems* Ever arrested (arrest history = more likely)*	Neither
Religion important	IQ (higher IQ = less important)**	Neither

Note. Only predictors from the liberal model for each dependent variable are reported

unless Favorable Departure Rating changed the model substantially in the conservative analysis. Outcomes from outlier analyses, when different, are not included. Significant predictors were related to the outcome variable in the expected direction (i.e., better pre-treatment functioning predicted better follow-up outcomes, and the inverse) unless otherwise noted. MS = Motivation system at departure. FDR = Favorable Departure Rating. L = liberal analysis. C = conservative analysis.

** $p < .01$ * $p < .05$ ⁺ $p < .10$

Table 51

Predictors of Post-Treatment Outcome Variables in 16-Year Follow-Up Sample

Outcome Variable Name	Significant Baseline Predictors
Motivation system at departure	Number of presenting problems** Age at admission (older = higher)* Gender (females = higher) ⁺ Minority status (Caucasians = higher) ⁺
Favorable Departure Rating	Number of presenting problems* Gender (females = higher)* Age at admission (older = higher)* Minority status (Caucasians = higher) ⁺ Number of prior placements ⁺
Length of stay	Age at admission (older = shorter stay)* African American status (African Americans = longer stay) ⁺

Note. Only predictors from the highest significant or marginally significant model for each dependent variable are reported. Outcomes from additional analyses are not included. Significant predictors were related to the outcome variable in the expected direction (i.e., better pre-treatment functioning predicted better post-treatment outcomes, and the inverse) unless otherwise noted.

** $p < .01$ * $p < .05$ ⁺ $p < .10$

Table 52

Predictors of Post-Treatment Outcome Variables in 5-Year Follow-Up Sample

Outcome Variable Name	Significant Baseline Predictors
Motivation system at departure	Gender (females = higher)** IQ (higher IQ = higher) ⁺
Favorable Departure Rating	Gender (females = higher)** Number of formal prior placements*
Departure Success Scale	Gender (females = higher)** Number of formal prior placements*
Length of stay	Age at admission (older = shorter stay)*
ROLES at departure	Gender (females = higher)* Number of formal prior placements* ROLES at admission*

Note. Only predictors from the highest significant or marginally significant model for each dependent variable are reported. Outcomes from additional analyses are not included. Significant predictors were related to the outcome variable in the expected direction (i.e., better pre-treatment functioning predicted better post-treatment outcomes, and the inverse) unless otherwise noted.

** $p < .01$ * $p < .05$ + $p < .10$

Curriculum Vita

Avi-Yonah M. Schwab

Education

- May 2008 Rutgers, The State University of New Jersey (Piscataway, NJ)
Ph.D., Psychology (Clinical Psychology Ph.D. Program)
- October 2005 Rutgers, The State University of New Jersey (Piscataway, NJ)
M.S., Psychology (Clinical Psychology Ph.D. Program)
- May 2000 Columbia University (New York, NY)
B.A., Psychology
- May 2000 List College, Jewish Theological Seminary (New York, NY)
B.A., Talmud and Rabbinics

Employment

- 2004-2007 Teaching Assistant, Rutgers University (Piscataway, NJ)
- 2000-2003 Account Executive, Perry Davis Associates, Inc. (New York, NY)

Publications

- Schwab, Y., & Cedeno, L. (2007, November). *The Effects of Cognitive-Behavioral Techniques on Math Test Performance: Exploring Outcomes, Implementation, and Context in the 'Raising Test Scores' Study*. Paper presented at the annual meeting of the Association of Behavioral and Cognitive Therapies, Philadelphia.
- Schwab, Y., & Elias, M. J. (2006, November). *Raising Test Scores: Using social-emotional learning and cognitive-behavioral techniques to increase math test performance*. Poster presented at the annual meeting of the Association of Behavioral and Cognitive Therapies, Chicago.
- Elias, M. J., & Schwab, Y. (2006). From compliance to responsibility: Social and emotional learning and classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 309-341). Mahwah, NJ: Lawrence Erlbaum Associates.
- Elias, M. J., & Schwab, Y. (2004, October 20). What about parental involvement in parenting? The case for home-focused school-parent partnerships. *Education Week*, 24(8), 39, 41.