TEXAS TECH UNIVERSITY
College *of* Education

Good Behavior Game Training Manual

Including:
Good Behavior Game Rules
Benefits of the Good Behavior Game
Medical Model vs. Ecological Model
Misconceptions
Further Readings

Good Behavior Game Rules

Good Behavior Game Rules

Behaviors

- We will work quietly (unless otherwise instructed e.g. in groups or class discussion)
- We will actively participate in classroom assignments (completing seat work, group work, or participating in classroom discussions)
- We will follow directions
- We will listen carefully when instructions/directions are being given, and while lessons are being taught.
- We will not distract other students from working or listening to the teacher by talking, whistling, humming, playing with items, throwing items, physically violating another student, acting silly, or any other distracting behaviors
- We will stay in our seats, unless otherwise instructed
- We will begin our work immediately after instructions/directions have been given (no gazing, delaying work, putting head down before completed, etc.)
- We will raise our hands if we need help with assignments and instructions/directions (we will not talk out without permission)

**Each of these will be listed on a poster to be hung in the classroom in a central location*

Steps for implementation:

1. Students are divided into equal teams (if possible).
2. Each team is assigned a team leader (Danielle should be a leader on the first day)
3. Teams are listed on the board in a location for all students to see
4. Prizes are distributed following the end of the game (reinforcers are listed below)
5. The game should last 10 to 20 minutes and is recorded with a timer

Procedures

1. Turn timer on
    a. "Ok class, we are going to play a game during (insert activity or lesson) for 10 minutes"
2. Review rules with the class
    a. "Each time a student in your group breaks one of the rules on the board, your whole team earns a tally. The team with 4 tallies or less at the end of the game is the winner! Both (or all) teams can be winners as long as they have 4 tallies or less."
3. Identify child and specific problem behavior in a neutral tone of voice
    a. "Ok Johnny, you got out of your chair without permission. That's one tally for your team."
4. Give a check mark to child's team and immediately recognize other team(s) for correctly following the game rules
    a. "Hey Sally! Great job making eye contact with me while I am giving instructions!"
5. Identify winning teams on the dry erase board.
6. Distribute or promise to distribute rewards
7. The team with the most wins at the end of each week is the "Weekly Winner" and is eligible to receive special privileges identified by the teacher at their discretion.

Reinforcers (prizes)

Tangible items

- Stickers
- Erasers
- Certificates
- Letters of commendation to parents
- Homework/assignment passes

(The teacher may use whatever reinforcers are appropriate and at their discretion)

Intangible items and special privileges

- group leaders
- line leaders
- pencil sharpener (sharpens classmates pencils)
- plant caretaker
- paper collector
- flag holders
- message carrier
- board washer
- book distributor
- 5 minutes to write on a chalkboard
- 15 minutes to play their favorite game
- name placed on Principals Outstanding Student List (displayed in the hallway)
- wears "Outstanding Student" badge
- teacher hugs the members of the winning team
- Students get a special hand shake from peers, other teachers, support teacher or principal
- Students get a congratulatory phone call from the teacher at home

First Day of the Good Behavior Game

On the first day, the teacher announces that the class will play a "game" for 10 minutes during reading (or any other subject), and announces the members of each team. The teacher will read the rules from the GBG booklet (i.e., this document), and review definitions of disruptive behaviors. The teacher will then explain that each rule violation (that is, occurrence of a disruptive behavior) will result in writing a check mark in the blackboard next to the team to which the offending child belongs. The teacher will verbally identify the misbehaving student and the behavior which earned the check mark.

The class will be told that any team with 4 marks or fewer at the end of 10 minutes wins the game, and that all teams can win if they all earn 4 marks or fewer. The class will then be told that the winning team(s) will get a happy face or star placed on the Good Behavior Game Magnetic Scoreboard (or regular dry erase board) at the end of the Game. Additionally, the class will be told that the winning team(s) will get a prize, immediately following the game. The teacher sets the timer for 10 minutes and announces the beginning of the game.

During the Game, the teacher should drop whatever they are saying or doing with the regular lesson and put a check mark on the board as soon as a disruptive behavior occurs; during this time the teacher should:

a) State what the wrong behavior was in a normal tone of voice
b) Identify the child who did it
c) Praise the other teams for behaving well

It is critical to stick to the Good Behavior Game rules when giving check marks. For example, getting arithmetic problems right is not one of the Game conduct rules, and teams should not earn check marks for poor academic performance.

At the end of 10 minutes when the timer goes off, the teacher should review with the class the number of check marks per team, repeat the 4-point or less criterion for winning the Game, and should announce the winning team (or teams).

The Team Leader should put up a star (drawn or pasted) on the GBG Scoreboard. Accurate start and stop times are important. The Team Leader should then hand out prizes to the children in the winning teams, which should be tangible rewards such as stickers or candy in the first few weeks. Children on the losing team(s) should do quiet seat-work with no special attention from the teacher. However, the teacher may decide what to do with the losing team(s). It is important not to punish teams that had more than 4 points.

Second Day of the Good Behavior Game

The teacher should announce that the class will again play the Good Behavior Game for 10 minutes, repeat that the teams are the same as the day before, review the classroom conduct rules, and review the four points or less rule for winning the game. The blackboard should be erased of the previous day's check marks before the children come in. On Day 2, the teacher should announce that they will be playing the Game on some days for the rest of the week, and that the team(s) that won the Game most often will be the Weekly Winner(s). The Weekly Winners will be awarded a special privilege. Otherwise, the Game will be conducted exactly as the last time.

Weekly Winners

At the end of the week, the Weekly Winner Team(s) will get a star with a happy face (or another distinguishable sticker) at the right-hand side end of the GBG Score Board. Non-winners should engage in quiet seat-work with no special attention from the teacher if the Weekly Winner gets to participate in special privileges. The use of "Weekly Winners" may be left up to the teacher's discretion as to not disrupt the flow of their regular class. However, this may motivate students to participate.

Good Behavior Game (GBG) Procedures throughout the Year

Continuing the Good Behavior Game after the First Week

On the first day of Week 2, the teacher together with the class should review their progress, and solicit their opinions about the Game. At this point, the children in the class should choose a special privilege for Weekly Winners that week. This gives all the children a long-term goal to work toward.

In the early weeks of the game the teacher should begin with tangible rewards that are distributed immediately following the game. These tangible rewards should be chosen from the list provided (or new ones may be added by the teacher). When all 3 teams are winning consistently, then the teacher should begin to include intangible rewards.

If all 3 teams continue to win consistently with the tangible and intangible rewards, then the teacher should start lengthening the time the game is played.

When all 3 teams are winning consistently with a longer game time, the teacher should begin altering the reward delivery time. For example, if the game is played from 9:00 am to 10:00 am, the children should be told at the end of the game that rewards will be distributed after lunch. If the teams continue to win consistently, the teacher may play the game early in the morning and distribute rewards at the end of the school day. Rewards should now be chosen from the list provided (or added by the teacher).

What to Do if All 3 Teams are Not Winning Consistently

If all 3 teams are not winning consistently, then composition of the teams should be altered, as well as rewards, reward delivery time, game time, or form a fourth team.

If too many disruptive children are in one team, these students may be put into a fourth team. If all 3 teams are losing, give tangible rewards immediately following the game and/or shorten the game time. If several children are consistently responsible for a team losing, these students may be placed into a fourth team.

When you split into the 4th team - do not make it appear to be a good thing or a privilege to be moved to the 4th team. Explain calmly to the children that they are being moved so that their team can win the GBG. Explain that they must remain on the 4th team until they win the GBG for 3 consecutive sessions.

Generalization Procedures

Generalization may begin when all 3 teams are winning consistently with the increased use of intangible rewards, varied reward delivery time, and increased game time. To generalize good behavior, the Game should be played at different times of day, during different activities, and even in different locations, (such as in the hallway walking to the cafeteria, or in the auditorium if possible). The idea is that Good Behavior is expected at all times, everywhere.

Benefits of the Good Behavior Game

Benefits of the Good Behavior Game

In GBG classrooms there is less off-task and aggressive behavior, which means that teachers dedicate more time to instruction.

Students who play the GBG are less likely to need behavioral services, less likely to abuse drugs and alcohol, and have lower suicide and depression.

It does not compete with instructional time and seamlessly integrates into the school day.

The Good Behavior Game's Strong Evidence Base

The Good Behavior Game was tested with 1st and 2nd grade classrooms in Baltimore City, MD beginning in the 1985-1986 school year.

The trial was implemented in 41 classrooms in 19 elementary schools with two groups of first graders.

Experts followed up with students in these classrooms periodically to study the immediate, mid- and long-term effects of the game.

Positive Outcomes

Elementary Schools - male students who entered the first grade displaying aggressive behavior had reduction in: aggressive and disruptive behavior and off-task behavior.

Middle Schools - male students who entered the first grade displaying aggressive behavior had reduction in: aggressive and disruptive behavior, off-task behavior, and delay in age of first smoking.

Young Adulthood - males at ages 19-21 had a reduction in: use of school based services for problems with mental health or use of tobacco/alcohol, illicit drug use/dependence disorder, alcohol use/dependence disorder, tobacco use, and antisocial personality disorder

GBG has been studied using randomized trials in multiple locations in the U.S. and abroad. Results from these trials support the findings from the first Baltimore City trial:

Outcomes	Baltimore City, MD (1990s)	Eugene/Springfield, OR (1990s)	Netherlands (2000s)	Belgium (2000s)
Short-term (Elementary)	Reduction in aggressive and disruptive behavior [2]	Reduction in aggression [4]	Reduction in attention deficit hyperactivity [7]	Reduction in aggressive and disruptive behavior, increase in on-task behavior, decrease in development of oppositional behavior [9]
Mid-term (Middle)	Reduction in conduct disorder, school suspensions, use of mental health services, and smoking [3]	Delayed first police arrest, reduction in association with misbehaving peers, and delayed time to first patterned alcohol and marihuana use [5]	Reduction in physical and relational victimization, major depressive disorder, generalized anxiety and panic disorder, and use of tobacco [8]	NA
Long-term	NA	Reduction in use of tobacco, alcohol, and illicit drugs [6]	NA	NA

The Good Behavior Game has been implemented in settings that included a significant population of ethnic/racial minorities as well as populations from urban areas in the U.S. and abroad. In the Baltimore trial, 65% of the sample identified as African American and 31% who identified as White. GBG has been most effective for boys with higher levels of aggressive, disruptive behavior.

**If you would like more information on the studies and benefits of the GBG, visit goodbehaviorgame.air.org*

Medical Model vs. Ecological Model

Why Are Models Important?

Different models of disability imply different intervention approaches.

The model we prescribe to affects our thinking and the way we approach problems in the classroom.

Medical Model Vs. Ecological Model

(Bricout et al . 2004)

Medical

"The disabled person is the problem; the child is faulty."

Diagnosis-Cure

The power to change a disabled person lies within a medical or associated professional.

The impairment is the focus.

"They need to be adapted to fit into the world as it is."

Ecological

Broader view of the "problem."

Individual characteristics are attributed to the joint effect of personal traits and environment.

Child is viewed through the context of his/her environment.

Teacher and parents have power to change child behavior.

The environment is changed.

Misconceptions

Misconceptions

Misconception #1

Positive Behavior Intervention Support such as the GBG is an intervention or practice.

However, it is an approach that provides the means of selecting, organizing and implementing evidence-practices by giving equal attention to:

-clearly defined and meaningful student outcomes
-data-driven decision making and problem solving processes
-preparing and supporting implementers to use these practices with high fidelity and durability

Misconception #2

The GBG emphasizes the use of tangible rewards that can negatively effect the development of intrinsic motivation.

However,
The GBG framework includes practices that provide students with feedback on the accuracy and use of their social skills and behaviors.
-When new and difficult social skills are being acquired, more external feedback systems might be used to give students information about their social behavior.
-As students become more fluent in their use of social skills, external feedback systems are reduced and replaced by more natural environmental and/or self-managed feedback.

Misconception #3

Positive behavioral supports like the GBG were designed for students with disabilities.

However,
Interventions like the GBG have been developed to improve how schools select, organize, implement, and evaluate behavioral practices in meeting the needs of all students.

Other Myths about Applied Behavior Analysis
(The theory underlying the Good Behavior Game)

Myth 1

Use of rewards by behavior modifiers to change behavior is bribery.

However,

-Reinforcement is one of the cornerstones of ABA therapy. In the beginning of an ABA therapy program reinforcement is used frequently "in order to shape appropriate behavior and motivate the child to learn."

-You are providing an item that has reinforcing value after a desired behavior, however the difference is when the possibility of that reward is offered. If the child's behavior is good or neutral and the offer for reinforcement is presented contingent upon behavior, then it is considered a promise. If the child is exhibiting non-compliant behavior and the offer of reinforcement is given as a means to gain compliance, then that would be considered a bribe.

-It is important for ABA treatment providers to not only find items that are reinforcing, but to also know when to offer those reinforcers. Secondary reinforcers will be key to the intervention by fading the external reinforcer by pairing the new item (e.g., preferred toy, song, etc.) or activity (e.g. social approval, hug, smile, etc.) with the original reinforcer.

Myth 2

Behavior modifiers are cold and unfeeling and don't develop empathy with the students.

Myth 3

Behavior modifiers deal only with observable behavior; they don't deal with thoughts and feelings of students.

Myth 4

Behavior modification only changes symptoms, not the underlying problem (Martin & Pear, 2015).

Myth 5

ABA practices promote robotic language/ behavior.

However,

-This myth stems from the misconception that ABA therapy is nothing but a series of drills and rewards, which causes children to display "appropriate" but robotic behaviors

-People who mistakenly connect ABA treatment to robotic behavior often fail to realize that these initial responses are just one part of an intervention focused on teaching new skills and transferring those skills into a variety of new situations, until a child learns how to learn in the natural environment

Myth 6

ABA therapy is only for children with autism.

However, ABA treatment programs have been utilized to correct all sorts of socially significant behaviors—communication, social skills, academics, reading and adaptive living skills such as gross and fine motor skills, toileting, dressing, eating, personal self-care, domestic skills, and work skills.

References

Martin, G., & Pear, J. J. (2015). *Behavior modification: What it is and how to do it*. Psychology Press.

The National Autism Network. Retrieved from http://nationalautismnetwork.com/about-autism/autism-treatments/top-10-myths-about-applied-behavior-analysis.html

Sugai, G., & Simonsen, B. (2012). Positive behavioral interventions and supports: History, defining features, and misconceptions. *Center for PBIS & Center for Positive Behavioral Interventions and Supports, University of Connecticut*.

Clinical Child and Family Psychology Review, Vol. 5, No. 4, December 2002 (© 2002)

The Good Behavior Game: A Best Practice Candidate as a Universal Behavioral Vaccine

Dennis D. Embry[1]

A "behavioral vaccine" provides an inoculation against morbidity or mortality, impacting physical, mental, or behavior disorders. An historical example of a behavioral vaccine is antiseptic hand washing to reduce childbed fever. In current society, issues with high levels of morbidity, such as substance abuse, delinquency, youth violence, and other behavioral disorders (multiproblems), cry out for a low-cost, widespread strategy as simple as antiseptic hand washing. Congruent research findings from longitudinal studies, twin studies, and other investigations suggest that a possibility might exist for a behavioral vaccine for multiproblem behavior. A simple behavioral strategy called the Good Behavior Game (GBG), which reinforces inhibition in a group context of elementary school, has substantial previous research to consider its use as a behavioral vaccine. The GBG is not a curriculum but rather a simple behavioral procedure from applied behavior analysis. Approximately 20 independent replications of the GBG across different grade levels, different types of students, different settings, and some with long-term follow-up show strong, consistent impact on impulsive, disruptive behaviors of children and teens as well as reductions in substance use or serious antisocial behaviors. The GBG, named as a "best practice" for the prevention of substance abuse or violent behavior by a number of federal agencies, is unique because it is the only practice implemented by individual teachers that is documented to have long-term effects. Presently, the GBG is only used in a small number of settings. However, near universal use of the GBG, in major political jurisdictions during the elementary years, could substantially reduce the incidence of substance use, antisocial behavior, and other adverse developmental or social consequences at a very modest cost, with very positive cost-effectiveness ratios.

KEY WORDS: substance abuse prevention; violence prevention; public policy; best practice.

INTRODUCTION

A behavioral vaccine is a simple, scientifically proven routine or practice put into widespread daily use that reduces morbidity and mortality. A powerful example comes from an epidemic that occurred 150 years ago.

During the nineteenth century, women died in childbirth at alarming rates in Europe and the United States. Up to 25% of women who delivered their babies in hospitals died from childbed fever (puerperal sepsis), discovered later to be caused by *Streptococcus pyogenes* bacteria.

In the late 1840s, Dr Ignaz Semmelweis worked in the maternity wards of a Vienna hospital. By meticulous observation, he discovered that the mortality rate in a delivery room staffed by medical students was up to three times higher than in a second delivery room staffed by midwives. Semmelweis postulated that the students might be carrying the infection from their dissections to mothers giving birth. He tested the hypothesis by having doctors and medical students wash their hands with a chlorinated solution before examining women in labor. The mortality rate in his maternity wards eventually dropped to less than 1%. Washing of hands with antiseptic solution—a behavioral vaccine—now saves millions of lives every year. Today, the Centers of Disease Control and Prevention (CDC) web site states, "[Antiseptic] hand washing is

[1]PAXIS Institute, PO Box 68494, Tucson, Arizona 85737; e-mail: dde@paxis.org.

1096-4037/02/1200-0273/0 © 2002 Plenum Publishing Corporation

the single most important means of preventing the spread of infection."

Other behavioral vaccines have been promoted on the premise of reduced mortality or morbidity. In the 1960s and 1970s, seat-belt use for adults and car seats for children were examples from the injury control literature.

In contemporary society, an epidemic affecting young people has waxed and waned since the 1960s. Substance abuse; delinquency; school failure; psychiatric disorders such as ADHD, oppositional defiance, and depression; teen suicide; teen pregnancy; and youth violence have adversely affected the lives of America's adolescents (see the various Surgeon General's reports on these topics). These problems often co-occur in what Biglan (2001) describes as multiproblem youth. Could there be a behavioral vaccine, nearly as simple as antiseptic hand washing, which might significantly reduce the mortality and morbidity of multiproblem behavior? Yes, there could be. This paper details what one might be and how it might become as common as a doctor or nurse washing hands with antiseptic solution.

Behavioral Vaccine Defined

A behavioral vaccine is a simple procedure that can dramatically change an adverse outcome. Unlike prevention programs, which are typically described as a collection of procedures delivered over a set time such as 8–12 weeks, a behavioral vaccine is given only once or used as a simple *routine* of daily life. Traffic safety research offers some useful examples. A drivers' education program is a prevention program. Both optional air bags and seat belts are behavioral vaccines. In the case of optional air bags when they were introduced, a person only needed to make a decision to buy a new car *with air bags* to enjoy the benefit of increased safety. In the case of seat belts, one must buckle up each time to maximize safety from harm. Both types of behaviors are relatively easy, unlike the complexity of a drivers' training course on accident avoidance. A hallmark of a behavioral vaccine is that a simple action yields large results. Per se, behavioral vaccines do not preclude other strategies, and may even work synergistically with or be combined with more strategies to leverage effects. Behavioral vaccines are typically very inexpensive, and work for a broad population, with few adverse effects.

The Logic Model of Behavioral Vaccines

Like any public-health measure, behavioral vaccines must be able to be used across the whole population to achieve the full societal as well as individual prevention effect. The need for universality can be modeled mathematically and logically (e.g., Embry & Flannery, 1999). Mathematically, prevalence rates of multiproblem behaviors (e.g., substance abuse, misuse, juvenile crime) typically range from 1 to 15% for the purposes of illustration here. Prediction of who will develop these problems over life span is quite clearly *not* 100%, and ranges vary dramatically depending upon the complexity, comprehensiveness, and sensitivity of the prediction tools. Issues of false negative and false positive identification loom large, however (Embry & Flannery, 1999). If one presumes that certain problem behaviors happen in around 5% of the population, then 500 out of 10,000 people ought to be afflicted. An 85% accurate prediction model (which would be the envy of most behavioral epidemiologists) could correctly classify 425 people, missing 75. How many might be false positives though? In a rough way, that can be calculated by subtracting the 425 correct positive predictions from 10,000. Then, multiply that number by the prediction error term (15%). The result is that public-health practitioners, school, law-enforcement, and/or parents might have erroneously identified 1,436 people as at-risk when they are not. In times of scarce resources, implementing selected or targeted interventions for those 1,436 people makes little economic, logical, political, social, or health sense—especially when the delivery of targeted or selected prevention strategies may run thousands of dollars per person, be very difficult to deploy, or both.

The logic of a behavioral vaccine improves the power, prediction, and cost-effectiveness of targeted and selected prevention strategies. This can be illustrated by problems posed by bioterrorism prevention and early detection. Chills, fevers, vomiting, and other such symptoms are common early signals of some agents suitable for bioterrorism, such as anthrax—necessitating an elaborate screening and detection (U.S. Surgeon General, 2002). These signals are also the early symptoms of the flu and colds, which are perversely common, resulting in false negatives or positives. Thus, an epidemic of flu-like symptoms could precipitate a dramatic overresponse from authorities (false positive)—which uses up valuable social, economic and political capital making, paradoxically making the society more vulnerable.

Or, the authorities might underrespond, dismissing the events as simply colds or flu (false negative). In the case of multiproblem behavior such as substance abuse and juvenile crime, behaviors such as early impulsivity, inattention, and disruptiveness among children—nearly as common as flu-like symptoms metaphorically—predict serious problems a decade or so later (e.g., Tremblay, Masse, Perron, and Leblanc, 1992); even though, a good half or so of the children will desist a decade later in these behaviors (e.g., Walker, Colvin, & Ramsey, 1995). If every young child who exhibits these behaviors receives medication, behavioral interventions at home, and behavioral interventions at school, the personnel and economic cost would be substantial. And, substantial numbers of children or families would be subjected to medication or services simply not needed and possibly iatrogenic. Now, imagine that a universal precaution can cut the incidence rate of the key manifestations of a behavioral or a public-health problem from 20 to 50%. Such prevention effects dramatically improve the sensitivity, power, and and cost-effectiveness of selected or targeted interventions—which can be modeled mathematically. This efficiency effect can be exemplified by the harried school counselor or psychologist who now has 20–50% *fewer* referrals for evaluation for conduct or attention problems, and who now has more time for more accurate screening and treatment.

The logic of a behavioral vaccine has even more potency if there are suspected contagion effects. Contagion can be real or via social learning in multiproblem behaviors. For example, placing a child with risk factors among other children who manifest those same symptoms for an intervention or prevention can dramatically escalate the expression of the rate and severity of symptoms, causing more harm to the individuals, peers, and society (e.g., Dishion, McCord, & Poulin, 1999). Even the simple random assignment of impulsive or disruptive 1st graders to classrooms with high, medium, or low levels of peer aggression can dramatically escalate or mitigate serious behavioral problems a decade later (e.g., Kellam, Ling, Merisca, Brown, & Ialongo, 1998). These adverse contagion effects could be the result of imitation, peer reinforcement of antisocial behavior, or escape conditioning from aversive behaviors by adults, or some combination. One might usefully think about contagion effects as "tipping points," which could be altered by classroom management, school climate, or community-wide behavioral vaccines.

A final set of issues exists in the logical model of behavioral vaccines: ability to scale to nearly universal coverage, low adverse reactions, and robustness to be used in combinations with other strategies.

Logically, a behavioral vaccine must be easily scaled to cover large areas of social geography and its attendant population to achieve protective effects. Logically, the behavioral vaccine would have to work with very diverse ages and work across different ethnic or cultural groups. Mathematically, it is virtually impossible to affect community-level outcomes (e.g., crime rates, drug use) without near universal coverage of a primary prevention strategy. A behavioral vaccine must also have low negative side effects, if used at scale. Why so? Lipsey (1992) reports that approximately 29% of the interventions to prevent delinquency actually make young people worse, and this may be a significant underestimate because efforts with adverse results are less likely to be published for many reasons. Thus, a behavioral vaccine with significant adverse effects for a subset could actually make community-level results worse, instead of better. An extension of the logic of reducing adverse reactions would extend to how the vaccine interacted with other prevention or intervention efforts, as a behavioral vaccine could be like certain drug interactions. To the dismay of most program developers, users of prevention protocols often do not implement them with fidelity or may mix them with home grown strategies. A potential behavioral vaccine could have robust internal validity in carefully randomized control-group studies, yet fail miserably in the field. Thus, a behavioral vaccine would need to have evidence of impact and utility in sloppy, naturalistic conditions.

The logic model for a behavioral vaccine shares some elements of the risk and protective factor literature currently driving much of the prevention policy in the United States (e.g., Catalano, 2001), yet is quite different in other ways more akin to large public-health campaigns. Both models rely on empirical data. In the risk and protective factor model (Catalano, 2001), small units of government (e.g., schools, school districts, or communities) attempt to create a plan presumptively based on their *unique* data. The behavioral vaccine model holds that certain risks or protective factors must be considered at a population or near universal level. The nature of the data construct (normative based) of the risk and protective factor model makes it very difficult to detect general population factors adversely affecting child development vis-a-vis small unit prediction. Further, the risk and protective factor model does not take into account the time sequence of prediction, only the current prediction in a cross-sectional mode. The behavioral vaccine model presumes a developmental sequence or vector, which if interrupted, has long-lasting effects.

If the logic model is true for behavioral vaccines, then great benefits could accrue for individuals, families, schools, and communities from a powerful prevention strategy that could be used in large-scale public health models. The question begs: does prevention science suggest any strategies as potentially appropriate as a behavioral vaccine for multiproblem behavior?

A Candidate Behavioral Vaccine

A bit over 30 years ago, two graduate students, Harriet Barrish and Muriel Saunders, and one of the founders of behavior analysis, Montrose Wolf, published a study on the effects of something called the Good Behavior Game (hereinafter, the Game; Barrish, Saunders, & Wolf, 1969). It worked pretty well, and became a behavior-modification "trick" most graduate students in behavior analysis or special education learned during the heyday of behavioral psychology. Neither Barrish, Saunders, or Wolf, nor the graduate students who learned to use the Game as a classroom strategy, had the slightest idea then how powerful the strategy might be for changing the future of children destined for lifetime multiproblems of substance abuse, violence, and school failure (Kellam & Anthony, 1998; Reid, Eddy, Fetrow, & Stoolmiller, 1999; White, Loeber, Stouthamer-Loeber, & Farrington, 1999).

Even with the spread of "best practice" guides, very few policymakers, government agencies, educators, prevention specialists, mental-health providers, or even research scientists know about the Good Behavior Game. Very few people know about the potential for the Game to prevent multiproblem behavior that gobbles up special education, juvenile delinquency, and treatment dollars.

The Game is the simplest of behavioral strategies, which has been described in detail in a manual (Embry & Straatemeier, 2001). First, the adult inducts children's definitions of the rules of the setting, specifically what would make the classroom or nonacademic setting a good place to learn, more enjoyable, pleasant, etc., all labeled as the "good things we all want." Second, the adult inducts children's descriptions of behaviors that would interfere with desirable outcomes and labels these generically as "fouls." Third, examples of both are presented physically and in words for the children to form a generalized concept. Fourth, the adult explains that the Game is played at intervals, like innings, but never for the whole day. Fifth, the adult divides the group into teams and explains that a team may win the Game by having the fewest fouls (or below a criterion in later research, enabling multiple winners), because that means more good has happened. Every team can win some brief activity prize if they have less than a predetermined number of fouls during an interval. Sixth, the adult makes sure a daily scoreboard is highly visible, just like the scoreboard of baseball or football, with fouls much smaller than wins. The Game has procedures for how to play in certain circumstances, how to keep it exciting, how to improve generalization, and how to solve problems for players who cheat or flout the conventions.

In this paper, I outline why and how the widespread application of the Game might be one of the most cost-beneficial prevention strategies available for schools and other settings. The paper will also map out the scientific and practical ways that the Game might become a universal public-health measure or vaccine for the prevention of multiproblem behavior. The rationale for the idea of a universal behavioral vaccine can be advanced on the basis of epidemiological research, findings from the neurochemistry of behavior, evolutionary psychology, replicated behavioral studies, and simple mathematics. This paper also discusses research and practical issues related to a "behavioral vaccine" for prevention.

EPIDEMIOLOGY OF MULTIPROBLEM BEHAVIOR AS FOUNDATION FOR A BEHAVIORAL VACCINE

The foundation for a behavioral vaccine would, of necessity, make sense only if there were evidence of a behavioral trajectory that predicted adverse outcomes. That evidence would be even stronger for the vaccine if the behavioral trajectory were measurable, meaningful, and malleable. Such a foundation is becoming much stronger because of the quality and quantity of scientific research on multiproblem behavior of substance abuse, delinquency, violence, school failure, and related mental-health disorders.

Just a few years ago, practitioners and scientists built program and scientific castles about the causes, prevention, and treatment of substance abuse, delinquency, violence, various mental-disorders, and school failure. Champions argued that each problem was caused by very unique factors, necessitating a tobacco prevention program, a marijuana prevention program, a violence prevention program, etc. These prevention castles have been defended to the death, even when they are expensive and show weak or no effects. Typically, the prevention models emerged largely as a result of simple cross-sectional studies or incomplete epidemiological information. It was and is a classic case of inadequate experimental design on *developmental* issues, leading to erroneous conclusions—just as Schaie and Baltes (1975) warned.

Over time, well-controlled multiple longitudinal and twin-studies stormed and demolished the castles, though defenders of the rubble still continue. Consider some examples of the castle sieges.

In 1990, Shedler and Block published landmark results on substance abuse from a long-standing longitudinal study. They reported that substance abuse (vs. substance experimentation) at age 18 could be predicted by simple measures of coercive parent–child interactions at age 8. Shedler and Block's findings mirrored the more fine-grained longitudinal studies on the role of parent–child coercive interactions in the cause of antisocial behavior by Patterson and Stouthamer-Loeber (1984), by Patterson, De Baryshe, and Ramsey (1989), and more recently by Ary et al. (1999). Other longitudinal studies, such as by Walker, Stieber, Ramsey, and O'Neill (1993), followed, showing the links between early aggression in boys and lifetime problem behavior. Tremblay et al. (1992) observed these connections in boys in Montreal. Consistent reports emerged from researchers in other locations. Raine, Venables, and Mednick (1997) found similar relationships in a long-term study in Mauritius. In the long-standing Child Development Study in New Zealand, Moffitt (1990, 1993) provided strong evidence for life-course continuity of early problem behaviors and adverse adolescent outcomes. Swedish studies showed long-term relationships between aggression, alcohol use, and criminals behaviors (Andersson, Mahoney, Wennberg, Kuehlhorn, & Magnusson, 1999). In the United Kingdom, Champion, Goodall, and Rutter (1995) have shown the connections between various adverse developmental outcomes in a decade-long study. Recently, more complex longitudinal studies have revealed similar data (Loeber, Stouthammer-Loeber, & White, 1999), yet expand on how depression and internalizing symptoms affect the outcomes along with early aggression. What do all these longitudinal data tell us? In general, the data suggest that many serious behavioral problems of adolescence and young adulthood emerge from similar behavioral pathways. These studies clearly suggest that the behavioral trajectory is measurable and meaningful. Are the trajectories malleable?

Some of the longitudinal studies, by happy circumstance, indicate that environmental or social events alter the apparent trajectory of multiproblem behavior. Consider just a few examples from the longitudinal literature. Patterson and colleagues have had the opportunity to study behavioral interactions (interval-by-interval coding) in the context of longitudinal study of antisocial children. What did they find? Patterson, Dishion, and Yoerger (2000) reported that more than 50% of the outcome of substance use, health-risking sexual behavior, and police arrests can be predicted by how much reinforcement of deviant behavior children receive. In a 1998 study, Patterson, Forgatch, Yoerger, and Stoolmiller argued that the prediction of lifetime deviancy had stable behavioral roots at least as early as the 4th grade, based on their data. One of Patterson's key colleagues has further documented that deviance reinforcement and delinquent behavior follow the matching law (Dishion, Spracklen, Andrews, & Patterson, 1996). The pattern of reinforcement delivered by parents and the reciprocal interactions between parent and child have been well documented to be malleable in high-quality, thorough behavior analysis or in other studies (e.g., Kosterman, Hawkins, Spoth, Haggerty, et al., 1997; Tremblay, Pagani-Kurtz, Masse, Vitaro, et al., 1995; Webster-Stratton & Hammond, 1997).

Most of the above work focuses on the family context, and other researchers have examined school or community contexts in terms of behavioral trajectory. Rutter, Maughan, Mortimore, Ouston, and Smith (1979) and Rutter (1985) show powerful effects of school organization on delinquency, behavior problems, and other outcomes. Rutter proposes that the structure and organization of school may differentially reinforce resilient behavior versus antisocial behavior. One of the original descriptive studies of the Baltimore Prevention Project (Kellam, Mayer, Rebok, & Hawkins, 1998) showed that classroom context had a 6-year impact on developmental outcomes for children with elevated developmental risk. Specifically, Kellam, Ling, et al. (1998) report that high-risk children who were randomly assigned to classrooms with naturally occurring low or high levels of aggression by other children had very adverse impact on the randomly assigned longitudinally studied boys but not girls. Collectively, Kellam's work suggests that the boys in his research settings might have been reinforced for aggressive behavior by peers (both negatively and positively), in much the same way as Patterson's cycle of coercion was observed in a family context. School context, at least, offers evidence of a behavioral trajectory that is measurable and meaningful.

Some evidence suggests that the behaviors might not be easily malleable, perhaps reducing the likelihood of a behavioral vaccine. It appears, from several types of inquiry, that some children have an innate vulnerability to the cycle of family or peer coercion, and possibly, the reinforcement of aggressive behavior. Some of the longitudinal studies strongly suggest a genetic modulation of outcome, as well as leverage points for intervention or prevention. In a report from their Montreal study, Tremblay, Pihl, Vitaro, and Dobkin (1994) obtained teacher ratings on 1,161 kindergarten boys from 53 schools with the lowest socioeconomic status, on the dimensions proposed by Cloninger, Sigvardson, and Bohman (1988). Tremblay et al. (1992) correlated the teacher survey results with the presence of self-reported delinquent behavior at age 13. Scores for *high impulsivity* and *hyperactivity* were the strongest predictors of delinquency ($p <$ .0001), whereas scores for low anxiety ($p < .016$) and low reward dependence ($p < .029$) provided a lower level of prediction (see Fig. 1). The results confirmed the prediction of Cloninger's neurotransmitter model that high impulsivity and novelty seeking predict high risk for antisocial behaviors, which are behaviors modulated by serotonin, dopamine, and norepinephrine (e.g., Cloninger, 1994).

If the longitudinal studies are correct, then the need for a strong behavioral vaccine might be even greater for individuals who have a genetic risk for multiproblem behaviors. The question is whether such genetic vulnerability exists. The answer is yes. Studies of twins amplify and refine the general longitudinal studies on multiproblem behavior, suggesting strong genetic linkages. Slutske et al. (1997) utilized the Australian Twin Registry for the largest twin study of conduct disorder ever reported. They examined 2,682 adult twins, and concluded that genetic factors contributed to at least 71% of the disorder. A related publication from the Australian Twin Registry (Slutske et al., 1995) showed that girls with conduct disorder had a 10-fold greater risk of having problems with alcoholism than girls without conduct disorder. The Minnesota Twin Study shows a strong association for alcoholism, ADHD, and other behavioral problems among 1,200 twins (Disney, Elkins, McGue, & Iacono, 1999). Most of the twin studies suggest a strong linkage between problems of attention, hyperactivity, and aggression as key underlying factors predicting multiproblem behavior in boys. Reduction in rate, intensity, and duration of these behaviors might be the logical target of a behavioral vaccine—unless such behaviors were so profoundly genetically driven as to be immutable. The research on genetic mechanisms of these findings has considerable implications for prevention.

Adapted from Trembley et al., 1994. p <.0001

Fig. 1. Longitudinal prediction from Montreal Study.

Genetic studies of multiproblem behavior have advanced significantly in the last decade, and these advances suggest that genetic vulnerability is not static but sensitive to social events—potentially making the need for behavioral vaccine higher, which might prevent the disturbing problems from unfolding. Few social scientists realize the significance of advances in genetics research, which regulate some of the neurotransmitter candidates identified by Cloninger (e.g., Cloninger, Adolfsson, & Svrakic, 1996) as implicated in multiproblem behavior (e.g., Comings, 1995; Comings et al. 2000; Comings, Gade, Muhleman, & MacMurray, 1996; Comings, Gade, Wu, et al., 1996). Importantly, candidate polygenic alleles for multiproblem behaviors have strong evidence for being turned on by exposure to perceived human stress (e.g., Madrid, Anderson, Lee, MacMurray, & Comings, 2001), and the neurotransmitters implicated in multiproblem behavior are clearly related to social interactions (e.g., Quist & Kennedy, 2001). Because the evolutionary psychologists and other scientists have convincingly documented that individuals who likely carry these genes (and behaviors) do not randomly mate (e.g., Buss, 1984; Krueger, Moffitt, Caspi, Bleske, & Silva, 1998), a behavioral vaccine for multiproblem behavior in children might have to operate in schools or community. The advances in genetics research help resolve the tension between nature versus nurture debate (see Embry, 2002, for a complete discussion), and a behavioral vaccine might mitigate against the interactions between genetic vulnerability and common social risk factors articulated by numerous investigators found in schools, communities, peers, and even homes.

What are the implications of all of these diverse epidemiological findings? First, reductions in early inattention, disruptiveness, and related behaviors ought to decrease long-term adverse socially undesirable outcomes—nothing particularly new but worth restating. Second, the *biological* processes of multiproblem behaviors are clearly affected by social events, and scientific advances now make it possible to understand how the social environment might affect the expression of genes related to the biology of multiproblem behavior. Third, the epidemiological data suggest that effective behavioral procedures, universally promoted and used, might well be powerfully effective environmental or behavioral "vaccines" to prevent the occurrence of multiproblem behavior.

A BEHAVIORAL VACCINE

Presently, society has two current operative definitions or venues of the vaccine concept. In medicine, a vaccine is a preparation containing weakened or dead microbes of the kind that cause a particular disease administered to stimulate the immune system, protecting the individual from future exposure. In computer science, it is a software program that protects a computer from a virus or worm infection. Both of these concepts can be extended to the behavioral realm.

With a behavioral vaccine, a person might be exposed to a weakened behavioral risk, which could stimulate a protective response to a more full-blown exposure to the social, emotional, or psychological risk. Or, a person might learn a protective program of behavior that attacks, dislodges, or protects against any exposure to a dangerous behavioral assault in the future.

Vaccines are most effective when everyone who has a risk receives a critical dose. Under such circumstances, the virus has no host population to infect. Childhood immunizations are classic cases of vaccines for a vulnerable population, with few children in developed countries now dying from scourges of the past.

A vaccine is not like treatment, the latter of which is typically given after the onset of the disease or disorder. Vaccines are typically given universally before onset.

Could certain simple-to-apply, universal behavioral interventions confer some sort of "immunity" against multiproblem behaviors such as substance abuse, juvenile delinquency, and other problems? The answer appears to be "yes." The Good Behavior Game is a good candidate to consider as a potential behavioral vaccine, and the next sections of this paper present the evidence and logic for the possibility.

The Good Behavior Game: General Theory and History

Some 100 years of solid psychological research shows that behavior varies as a function of its consequences (e.g., Catania, 1992; Malott, Whaley, & Malott, 1997). Thorndike first labeled this as the "Law of Effect" back in the early 1900s. Since that time, the observations have been codified into the most robust replicated general principles of the science of behavior such as the "Matching Law" (e.g. Herrnstein, 1970). There is a profound reason that scientists refer to this principle as a "law." It is universal, highly replicated, easily demonstrated, and parsimonious. Against this backdrop, graduate students like Harriet Barrish and Muriel Saunders and scientists like Montrose Wolf thought disruptive, disagreeable behaviors by students might happen because peers and others somehow reinforced them in school settings. Perhaps, the smiles, giggles, laughs, and even pointed taunting from other students were *reinforcing* the high rate of the behaviors that teachers found so difficult to handle or harmful to the learning process. In this context and time, the graduate students and senior scientists reasoned that some kind of group-based reward for *inhibiting* negative behavior might be a boon for classrooms. Already, there were powerful precedents for such an idea. The idea for the Good Behavior Game was born.

Behavior Analysis Studies of Good Behavior Game Demonstrate Efficacy

Applied behavior analysis (Baer, Wolf, & Risley, 1968) posits careful testing of strategies to change human behavior in context, most frequently using time-series methodologies such as reversal or multiple-baseline evaluations, which have powerful advantages in applied research (e.g., Barlow & Hersen, 1973). The initial efficacy evaluations of the Game occur in this context.

First Test of Efficacy

In 1969, Barrish et al. published the first study on the Good Behavior Game using a multiple-baseline design in a very difficult classroom. It was this class that became the first to try the Game in a controlled study. The 4th-grade children were observed during maths and reading. Trained observers coded student behavior every minute for an hour, 3 days a week for several weeks. The children were out-of-seat or talking-out for about 80–96% of each class period, making instruction nearly impossible. Bedlam would have described the class.

The Game was played everyday during maths, with the class divided down the middle row into two teams. One or both teams could win privileges (e.g. wear victory tags, be first in lunch line, get a star on a winners' chart, earn free time) by having the lowest number of marks tallied on the board for disruptive behaviors. Teams with under 20 marks for the week earned special privileges at the end of the week.

The rate of disruptions fell immediately from about 90% to 10% of the intervals during the math hour, a great improvement. Meanwhile, the disruptions during reading time stayed pretty much the same.

After a few weeks, the teacher stopped playing the Game during maths but started playing it in reading. The results immediately showed the efficacy of the Game. Behavior during maths looked pretty bad again, just like the "baseline." Behavior during reading was greatly improved. After a week, the teacher played the Game during both times, and the rate of problem behavior fell quite low.

Efficacy Test of Game Components

The Good Behavior Game actually has several potentially "active ingredients" that might account for its efficacy. In 1972, Medland and Stachnik tested the good-behavior Game in a 5th-grade reading class consisting of two groups of 14 students each in a reversal design, using the class as its own control. They tested the whole game and different components to see how they worked. Game components included rules, red or green lights (response feedback using nonemotional cueing), and group consequences of extra recess and extra free time. Two observers counted talking-out, disruptive, and out-of-seat behaviors. The graphs from the study show that the total Game package reduced all the disruptive behaviors from their baseline rate by almost 99% for one group and 97% for the other. The component analysis revealed that after association in the Game, the nonemotional cueing stimuli of rules and lights were moderately effective in reducing the problem behaviors; the whole Game package was, however, most efficacious. What was particularly noteworthy was the fact that the students and teacher were able to cover 25% more academic material during the Game. This study revealed that the use of the signal light decreased bad behavior, underscoring the importance of a consistent, unemotional response or cue about bad behavior. The study also revealed that enunciation of the rules by the teacher each day had a small effect, which could explain the often reported comment by teachers that the children "need to be nagged" about the rules. Figure 2 summarizes Medland and Stachnik results.

Efficacy Test With Higher Risk Population

Children who ultimately develop multiproblem outcomes often have a special-education history (e.g.,

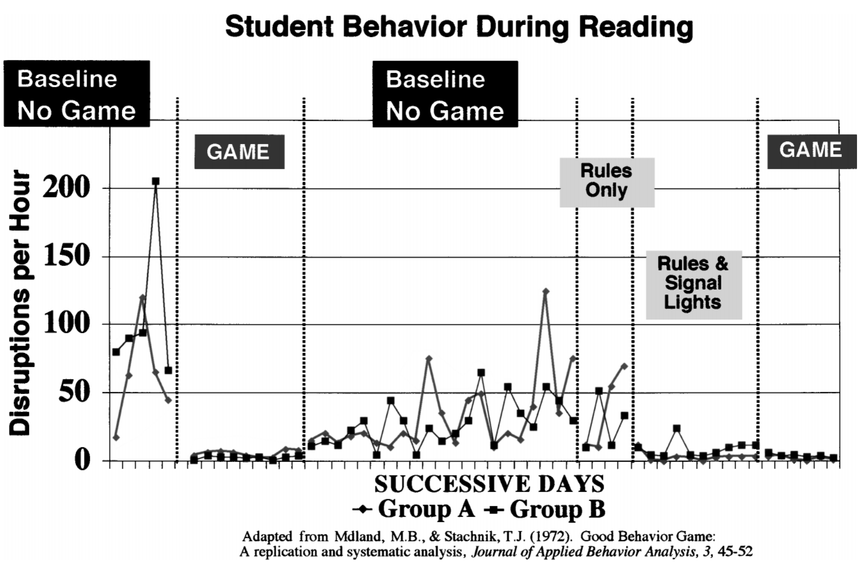

Fig. 2. Medland and Stachnik (1972) results.

Walker et al., 1995), and it would be important to demonstrate that a potential behavioral vaccine could be efficacious with such higher risk populations. Grandy, Madsen, and De Mersseman tried the Game with elementary-age special-education students in 1973 in a behavior analysis design. Again, the disruptive behaviors went way down. This study showed that the Game could generalize to a higher risk population.

Refinement of the Efficacious Components

Medland and Stachnik (1972) did not test all the salient components of the Game package, which might be crucial in understanding the active ingredients of this potential behavioral vaccine. Harris and Sherman tested the Game components in 1973, and they too found that disruptive talking and out-of-seat behavior fell dramatically in 5th- and 6th-grade students. By testing the Game in multicomponent reversal design, they allowed for a better understanding of key, effective components of the Game. Key ingredients turned out to be the division of the class into teams, positive consequences for a team winning the Game, and a low number of marks set as criteria for winning the Game. Harris and Sherman did find that reductions in negative behavior only slightly affected academic achievement, which flags the need for other research to determine whether the Game could be combined with explicit academic improvement strategies without adverse effects as teachers and schools would be likely to pursue additional components.

Efficacy Test With Young Primary School Children

In the chronology of efficacy studies, all had been focused on intermediate-level students in elementary schools. No evidence existed that it could be efficacious with younger students, which would naturally boost its potential as behavioral vaccine. Bostow and Geiger evaluated the Game's effects using a behavior analysis design on 2nd graders in 1976. Here again, it was effective, expanding the generalizability to younger ages.

Comparative Efficacy Trials for Rival Strategies

The Game is not the only school-based strategy that could be used to decrease the impulsive, disruptive, and inattentive behaviors that predict multiproblem behavior. A good candidate for a behavioral vaccine is likely to have a family of related interventions, and finding the most efficacious alternative would be logical. One of the most obvious alternative strategies is teacher attention, that is, training a teacher to pay more attention to a child's good behavior. In 1977, Warner, Miller, and Cohen compared the effects of the Game against simple teacher attention for being good among 4th and 5th graders. The Game was much more effective and simpler to use, which was important for building a case for it as a potential behavioral vaccine. Warner and colleagues also provided a key finding for social validity of the Game as a potential behavioral vaccine. As teachers often complain that they cannot praise for a variety of reasons, the differential effects of the Game met a key objection to a common recommendation of increasing praise.

The Role of Peer Pressure as Key Component

Deviant peers are clearly a risk factor in the epidemiology of multiproblem behavior (e.g., Dishion et al., 1999), and the Game historically made explicit use of peer-related variables: peer pressure, peer competition, and peer recognition via teams. Was this an important element for the Game achieving its therapeutic effects, which is important to understand for the use of the Game as a behavioral vaccine. Hegerle, Kesecker, and Couch directly replicated the Game again in 1979, but examined the efficacy of these peer-related components. They found that peer pressure, competition, and social recognition were all important components. This added to the understanding of why the Game might work. These components fit well into the notion of the matching law with peers and school systems (e.g., Dishion et al., 1996; Embry & Flannery, 1999). The matching law (Herrnstein, 1970) can be expressed as

$$B = kr/(r + re)$$

B is the behavior in question. k is a asymptotic constant and r is the rate of reinforcement of the B; this is divided by the same r plus re (the rate of reinforcement of all other behaviors. Peer pressure and competition reduce the re term, thereby making the r (social recognition) more potent for positive actions in the classroom. This author believes this matching law effect helps explain why just putting check marks up by individual children's names is far less effective than the strategy of a mark for a child's team. The competition diminishes the re (e.g., peer attention to negative behavior), making the rewards controlled by the teacher for wining the game (e.g., the *r*) more potent.

Efficacy of the Game After Initial Training

How long might the effects of the Game last after being played briefly with no coaching from anyone outside the classroom? Johnson, Turner, and Konarski answered that question in 1978. The answer helps shape how an effective behavioral vaccine might be delivered. Among highly disruptive intermediate classrooms, they found that the effects of the Game did last but started to decay after 2 months when the "coach" stopped coming to the classroom to encourage the use of the Game. This particular study suggests, not surprisingly, that a diffusion model of the Game as behavioral vaccine might require some attention to produce longer term effects.

Efficacy of the Game Across Cultures

If the Game worked across different cultures then it might mean that the processes were very strong, profound, and universal. Such a finding would boost confidence that the Game could be a viable candidate as a behavioral vaccine. Huber reported positive results in Germany in 1979 in a behavior analysis efficacy study. Saigh and Umar (1983) found strong effects for Sudanese 2nd graders whose parents could not read or write, in a reversal design. Saigh and Umar were among the first investigators to report that the Game reduced aggression. It is interesting to note that younger children vis-a-vis older children seem to show reversal effects rather quickly, suggesting that young children will require more consistent, lengthy use of the Game. These published studies suggest that the Game can be effective in culturally diverse contexts.

Generalized Efficacy of the Game to Non-classroom Settings

Previously, all published studies had focused on the efficacy of the Game in classrooms. From a behavioral vaccine perspective, the odds for success would be strengthened if the "vaccine" could be administered in other settings where the epidemiologically relevant behaviors are manifest. In 1981, Fishbein and Wasik showed that the Game could be played in the school library and bridge to the classroom at the same time. Their study also illuminated a variable that could improve the social validity of the Game, its widespread use: A delightful twist involved having the students help set and define the rules, with no loss of effects. As almost any classroom teacher could articulate, students are more likely to "buy in" and not resist the Game, if they can help set the rules. Although the efficacy of the Game in the library is nice, bad behavior in the library is not a huge known predictor of substance abuse, violence, and other ills. In 1998, Patrick, Ward, and Crouch found that the Game could be powerfully adapted to physical education or play-type activities outside. This suggested that the Game could also be used to solve playground or recess problems—which is an epidemiologically relevant risk predictor (e.g., Walker et al., 1995).

Efficacy of the Game for Special Education Students in Regular Classrooms

A behavioral vaccine would have limited value if it could not buffer or protect a vulnerable child in a high-risk setting. Children with special education designation in regular classrooms are an example of such a risk. Did the Game work for really serious behavior-problem children who were "mainstreamed" in a regular classroom when the whole class played the Game? Yes, discovered Darveaux in 1984. She had the Game played in a classroom while observing two targeted children on each team. The two target behavior-problem children did improve when the whole class played the Game. This suggested that classroom teachers would be able to use the Game as an effective behavior management strategy for children at-risk for placement in special services.

Impact of Different Kinds of Rewards on Efficacy

Teachers typically select and apply rewards for behavior quite idiosyncratically, which could seriously impair the efficacy of the Game if significant fidelity of implementation were required for rewards for the behavioral vaccine to work. What kind of rewards work for the Game? Kosiec, Czernicki, and McLaughlin found in 1986, that students did equally well when they played the Game for activity rewards versus candy. The children did like the candy as a reward, but it was useful to discover that activity rewards were powerful. The fact that activity rewards appear to be as powerful as material or edible rewards helps with the acceptability of the Game by teachers and school administrators, who often express dislike for material rewards.

Efficacy of the Game With Adolescents

Previous prevention research has suggested that boosters, rather like vaccine boosters, improve long-term results. Thus, it is reasonable to ask if the Game might work with adolescents. In 1986, Phillips and Christie found the Game worked quite well for intellectually impaired students whose ages ranged from 12 to 23 years. In 1989, Salend, Reynolds, and Coyle proved that the Game worked for emotionally disturbed adolescents. The older students liked the Game and stopped doing inappropriate verbalizations, inappropriate touching, negative comments, cursing, and drumming. These findings suggest that the Game could be played, possibly as a booster, with older youth.

Efficacy With Very Young Children

People often apply medications for other uses or for different age groups. It is natural to wonder if the Game might be used with very young children, which would broaden the basis for the Game as a behavioral vaccine. A special puppet helped the preschoolers learn the Game in the study by Swiezy, Matson, and Box in 1992. Some other adaptations were required, however. Special colored badges were needed by the teacher to track the preschoolers as they moved from place to place in the room.

Summary of Efficacy Studies

The early phases of science are best served by repeated measure studies such as those used in applied behavior analysis. Such studies provide a powerful, simple way of determining if the procedure has any probability of effect and helps identify how it varies based on different conditions, something not easy to do in randomized control group studies or is very very expensive. The early studies on the Good Behavior Game show it to be a very promising, robust procedure.

Social Validity Studies

A potential behavioral vaccine might be efficacious, but highly disliked by its putative users. Consumer liking of a product can obviously affect word-of-mouth, fidelity of use, and other factors that would be relevant to long-term prevention. Social validity is an important concept in large-scale behavior change, which measures (1) the social significance or importance of the goals, (2) the social appropriateness of the procedures, and (3) the social importance of the effects (Sulzer-Azaroff & Mayer, 1991). These questions are pivotal in the diffusion of any science-based practice. How does the Game measure in the field of consumer satisfaction? In 1994, Tingstrom found out that over 200 teachers did like the Game and would use it. An important signal came from that study in that teachers who did not "believe in positive reinforcement" were not as likely to adopt it, however.

Randomized Control Studies for Effectiveness of a Potential Behavioral Vaccine

The efficacy studies discussed certainly point to the utility of the Good Behavior Game in changing modifiable, meaningful, and measurable risk factors of multiproblem behavior. However, the "Gold Standard" of science is the use of random assignment to condition, especially large numbers of participants. By the late 1980s, it was apparent that the Game had strong effects and could be something to try in a large randomized trial, which happened with the Baltimore Prevention Project.

A total of 864 1st-grade students from 19 Baltimore public schools participated in the study during the 1985–86 academic year. Short-term results relied on assessments of all students in the fall and spring of 1st grade using three tools:

- The Teacher Observation of Classroom Adaptation Revised (TOCA-R)—measuring a variety of childhood developmental psychopathologies,
- The Peer Assessment Inventory (PAI)—measuring peer social networks, and
- Direct observations of student behavior by classroom observers.

The study had both control classrooms within (internal controls) and across schools (external controls), making for a more powerful but complicated study.

In Baltimore, as in the earliest versions of the Game, classes were divided into teams, which were rewarded when members behaved appropriately and participated in classroom activities rather than broke rules and fought. Three teams were created per class, with equal distributions of aggressive and shy children per team. During the first weeks of the intervention, the Good Behavior Game was played three times each week, for a period of 10 min. Over successive weeks, duration per Game period was increased by 10 min, up to a maximum of 3 hr.

What were the early results? Dolan and the other Johns Hopkins scientists made an initial report in 1993. First, both teachers and peers rated boys as more aggressive. Second, boys were seen as more shy by teachers, but not by peers. Third, the Good Behavior Game had a significant short-term impact on teacher ratings of aggressive and shy behavior for both males and females. There were some useful subfindings:

- The intervention had greater impact in reducing aggressive behavior in students who began the year with high aggressive ratings compared with students who began with low aggressive ratings—an important finding if the Game were to be viable as a potential behavioral vaccine.
- Peer nominations of aggressive behavior among boys by their classmates were also significantly reduced. Only one of the three peer nominations of shy behavior showed significant impact ("has few friends") and that was only in the case of females.
- Finally, the Good Behavior Game increased students' on task performance in the classroom as assessed through direct observations.

What were the longer term results? These are exceptionally important from a developmental perspective, because the real problems, related to early predictors such as aggression, do not show up until the adolescent years. In Baltimore, the longitudinal results were collected 6 years later. Kellam, Mayer, et al. (1998) reported that although the positive effects reported by teachers during intervention years in 1st and 2nd grades waned somewhat in the 3rd and 4th years, they reappeared in 5th grade and strengthened in 6th grade. More aggressive 1st-grade males benefited the most from the Game, with the aggression rating of over 30% significantly dropping by 6th grade. It appears, then, that the Game might function as a behavioral vaccine in a long-term study.

There were other long-term effects, not wholly predicted when the study started, strengthening the potential of the Game as a behavior vaccine for multiproblem behavior. For example, males were significantly less likely to initiate smoking (a 50% reduction in initiation rate) in the early teens (Kellam & Anthony, 1998). Teacher ratings and self-reported age at first use of tobacco showed that (1) boys who had received the Good Behavior Game intervention were rated as better-behaved than their counterparts in the other study conditions ($p < .05$), and (2) the risk of starting to smoke tobacco by age 13–14 years was substantially greater for boys in the "standard setting" control classrooms as compared to those who had spent 1st and 2nd grades in the Good Behavior Game classrooms ($p < .05$). Kellam and Anthony (1998) concluded from the long-term follow-up that targeting early risk of aggressive behavior is an important smoking prevention strategy, something that longitudinal tracking studies with no intervention had suggested but not proved. To this author's knowledge, the result published by Kellam and Anthony is the first inkling that a single classroom teaching strategy by an individual teacher might substantially reduce substance abuse, misuse, or initiation (see Fig. 3).

A whole array of publications exist on the Baltimore project, noting its theory, design, and results (e.g., Ialongo et al.,1999; Kellam et al., 2000; Kellam, Ling, et al., 1998; Kellam, Mayer, et al., 1998; Kellam & Rebok, 1992; Kellam, Rebok, Ialongo, & Mayer, 1994). Kellam and associates are continuing longitudinal follow-ups of the original cohorts, which will likely reveal more information about the life-course effect of the Game on such issues as arrest, educational attainment, and other milestones. When new

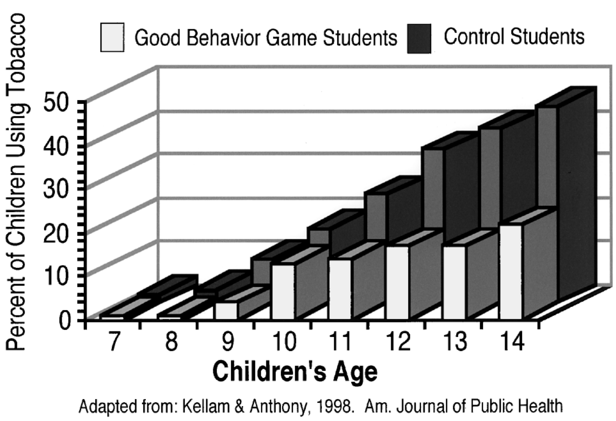

Fig. 3. Good Behavior Game impact on tobacco initiation.

medicines are introduced and approved by the Federal Drug Administration, it is rare for the approvals to cite ongoing inquiries with a decade or more long-term follow-up. Game is similarly rare in the prevention science literature, and the long-term follow-up strengthens the case for the use of the Game as a potential behavioral vaccine.

Not all reviewers concur about the value of the Game for prevention. Greenberg, Domitrovich, and Bumbarger (1999) offer a critique of Kellam's studies, observing that the intervention did not include family or the larger school ecology (which this author views as a strength, in terms of the utility of the Game as a behavior vaccine). The 1999 critique did not have the benefit of the Ialongo, Poduska, Werthamer, and Kellam (2001) study comparing the impact of combined classroom intervention (both the Game and Mastery Learning) against a Family Program, which showed that the combined classroom approach was superior to the family-only program. Greenberg and colleagues also argue that two of the primary sources of data (teachers and peers) were aware of the treatment condition and in some ways had a stake in the outcome, which may have affected internal validity. Again, the fact that these two sources of data did show change is a source of strength, considering that both peer nominations and teacher ratings are extremely resistant to *any* intervention, yet are highly predictive of serious antisocial behavior many years later (e.g., Embry & Flannery, 1999; Embry, Flannery, Vazsonyi, Powell, & Atha, 1996; Walker et al., 1995). To provide a comparison in top-rated prevention programs, Second Step (a violence prevention curriculum for elementary students) shows no impact on teacher ratings or parent ratings after a considerably more intensive classroom intervention in a randomized control group study (e.g., Grossman et al., 1997). Greenberg and colleagues review (Greenberg et al., 1999) of the Good Behavior Game erroneously reported that there had been no independent replications of the intervention, failing to cite the extensive, prior, peer reviewed studies mentioned herein while also observing that the Linking the Interests of Families and Teachers (LIFT) project incorporated the Game as part of its overall strategy.

Linking the Interests of Families and Teachers (LIFT), a prevention program designed for delivery to children and parents within the elementary school setting (e.g., Eddy, Reid, & Fetrow, 2000), worked in 12 public elementary schools with about 700 students in higher risk neighborhoods. The LIFT targets child oppositional, defiant, and socially inept behavior and parent discipline and monitoring—many of the variables targeted by Kellam and colleagues. The LIFT is (a) classroom-based child social and problem skills training, (b) playground-based behavior modification using an adaptation of the Good Behavior Game, and (c) group-delivered parent training. The results of a randomized controlled evaluation of the LIFT are reviewed. To date, during the 3 years following the program, the LIFT delayed the time that participants first became involved with antisocial peers during middle school, as well as the time to first patterned alcohol use, to first marijuana use, and to first police arrest. Reid et al. (1999) report reductions in playground aggression, with the largest effect size among the most aggressive children, as well as improvements in family problem-solving actions. At 30-month posttest, children from the treatment group were also significantly less likely to have been arrested. Microcoding of real-time playground aggression showed that intervention benefited the most aggressive children at recess with substantially high effect sizes (Stoolmiller, Eddy, & Reid, 2000).

The LIFT effort by Reid and his colleagues is noteworthy, because it is a systematic rather than direct replication of the Game, which was imbedded in a larger effort. This means that the Game can be incorporated with family and social skills interventions with no apparent adverse effects. From the perspective of a behavioral vaccine, it is vital that a strategy be able to work in combination with other strategies and still show benefit.

Awards and Recognition for the Good Behavior Game

The positive effects of the Game have been recognized by a number of sources. The Game is one of the few "universal," simple strategies identified by the Colorado Violence Prevention Blueprints Project, funded by the U.S. Centers for Disease Control, as meeting the scientific standards for a truly promising violence prevention practice. The Substance Abuse and Mental Health Administration has also identified the Game as a research-based promising practice. The Surgeon General's Report on Youth Violence (U.S. Surgeon General, 2001) lists the Good Behavior Game as a desirable practice.

These awards and recognition are all the more remarkable, because the Game is the only such intervention in the public domain, and something that an individual teacher or staff member can implement versus a comprehensive school-wide program. The breadth of replications of the Game by so many different investigators across time only strengthens the accolades.

Support From Current Field Trials and Other Studies for Potential Behavioral Vaccine

As established in the early parts of this paper, a behavioral vaccine envisions widespread use of a procedure. The Game needs to have some evidence of real-world diffusability.

Presently, my colleagues and I are engaged in a number of trials of the Game in a larger context. These community trials are described below.

Approximately 15 schools in the Greater Cleveland area are involved in an open field trial of the Game to determine if the game can be simply packaged and trained in the course of 4–6 h. The Game is referred to here as the PAX Game to denote the inclusion of some ancillary components documented to improve compliance and classroom management such as "beat the timer," nonverbal cues for stop (see Medland & Stachnik, 1972) and transition cues for walking in hallways. Early data show that schools can implement the game, and have impact on such variables as student referrals and suspensions.

Several years ago, my colleagues and I helped Cook County Health Department in Cook County, Illinois, design a protocol to have paraprofessionals visit classrooms and teach the Game to the students and their teachers. To date, Cook County Health Department has taught numerous classrooms the Game and collected simple observational data on those classrooms. The iteration of the Game designed by the author and colleagues incorporates the identified active ingredients from the efficacy and effectiveness studies, and it has been put together in such a way to encourage the use of other research-based protocols that might round out the effectiveness of the Game.

Besides the components of teams, peer pressure, competition, activity rewards, nonemotional cues, enunciation of the rules, and group-based rewards, the iteration includes some simple procedures to help improve the social acceptability, participant buy-in, facilitate generalization, and assist the tracking of the game. Here are a few examples. The students induct the rules and vision of the class using some special lessons. They pursue productivity, peace, health, and happiness by creating PAXIS™. Things that get in the way of PAXIS, a made-up word, are called spleems™, also a made-up word. The word for the goal helps foster positive debriefs (e.g., "What did you do to create PAXIS today?"), which has been shown to assist in the generalization of self-management and is a substitute behavior for teachers to avoid negative attention. Spleems are a word designed to reduce the verbally inflected emotionality attached to noticing a rule-breaking event, a key ingredient. Conversationally, it is much less explosive to say "that was a spleem" than "you broke the rule." The PAXIS version includes many small but useful stratagems needed to package a research-based practice for diffusion—a critical factor in a bringing a potential behavioral vaccine to scale.

The new words like PAXIS and spleems help track the behavioral contagion effects of the Game, as the words are completely novel. The words are what some cultural anthropologists define as "memes"—a sort of potentially self-replicating cultural concept, again to a gene. Lynch (2001) describes a meme (pronounced "meem") as a self-spreading thought, idea, attitude, belief, or other brain-stored item of learned culture. The idea of memes are frequently used in marketing as a way to track name recognition and build up brand recognition.

The use of words for the Game such as PAXIS and spleems create a "meme" in a school setting, providing a way to assess the frequency of the use of the Game. For example, children who have played the Game in the last week are able to explain in great detail if their team received any "spleems" that week. Children who do not know what the Game is, will look quite blankly at a visitor if you ask what "spleems" their team committed yesterday. Thus, prevention specialists such as the ones in Cook County Health Department can quickly assess whether staff are really following through with the daily repetitions—a necessary element of a putative behavior vaccine. It is rather like the question, "did you floss your teeth this morning" versus "do you practice good dental hygiene?"

The teaching of the Game by Cook County paraprofessionals is the first attempt to move the Game to a behavioral vaccine model, capable of being taught outside the context of graduate students and research personnel. A sample of data from one school and classrooms in Cook County appears in Fig. 4, showing observed "spleems" over time before and after the teaching of the Game, which are very encouraging. Not all schools and classrooms have the same results.

The current effort in Cleveland and the past effort in Cook County suggest that the Game can be

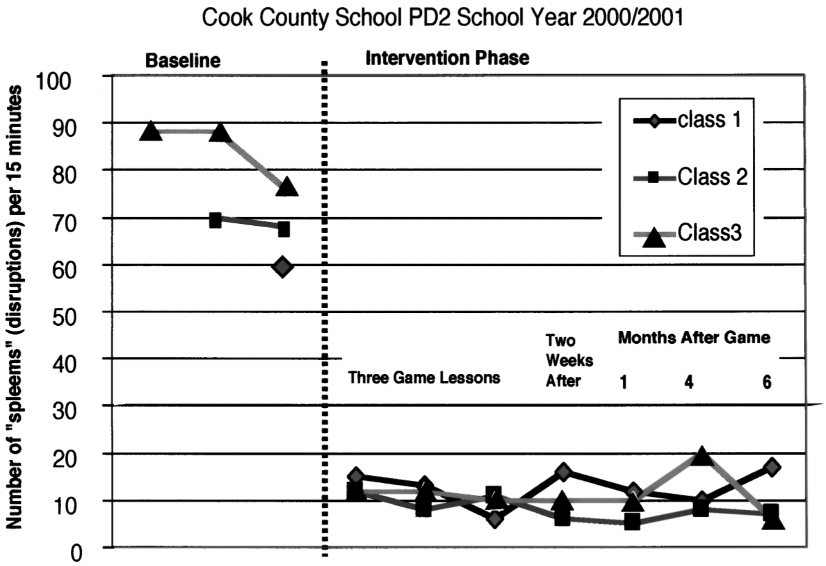

Fig. 4. Impact of Game taught by paraprofessionals.

practically disseminated in a real-world context. In the case of greater Cleveland, the Game was trained on a school site or across school sites in a brief training for teachers. In the case of the effort by Cook County, paraprofessionals learned how to implement and teach the game to many different schools in the actual classrooms. Other field trials are in place by the first author in Wyoming (a rural area with extremely high rates of substance abuse), in Tucson, AZ, with very high rates of Hispanic and Native American populations, in the multicultural context of some schools undergoing comprehensive school reform, and even in Singapore and Malaysia to assess the acceptability in very different systems and cultures.

MEDICAL RESEARCH ON INHIBITION RELATED TO A BEHAVIORAL VACCINE

Various studies implicate the problems of inhibition in the etiology of multiproblem behavior (e.g., Frick, Kuper, Silverhorn, & Cotter, 1995). For some time, it has been evident that medications, such as methylphenidate, increase inhibition and improve the kinds of behaviors studied in all of the studies on the Good Behavior Game (see Gadow, Nolan, Sverd, Sprafkin, & Paolicelli, 1990). In the United States, the daily use of such stimulant medication is extremely widespread—representing a rival treatment for the risk factors that might be addressed by a behavioral vaccine.

It is documented that an effective behavior management protocol will reduce the dose need of medication (i.e., Carlson, Pelham, Milich, & Dixon, 1992). Recent reviews suggest that behavioral protocols ought to be the first line of defense for the treatment of such conditions as ADHD (e.g., Pelham & Fabiano, 2000), for a variety of legal, ethical, and practical considerations. The issue here is not whether behavioral interventions or medical interventions are better.

The fact that both medication and a powerful strategy like the Game result in inhibition of negative behavior suggests that the two techniques probably operate in similar ways in the brain. In science, this is called the Law of Parsimony or Occam's Razor. It typically means that if two things have similar effects they most likely have common causal mechanisms. In the beginning of this paper, I have hypothesized that the common factor is the inhibition circuitry of the brain, which may have been altered as a result of genetic expression, gene–environment interaction, exposure to traumatic events, coercive parenting practices, deviant peer reinforcement, or even exposure to environmental toxins such as lead. The potential mechanisms for this are becoming more apparent with various scanning technologies and reaction-time studies (e.g., Lazzaro, Gordon, Whitmont, Meares, & Clark, 2001). Reaction times can be measured in two ways: go reaction and stop reaction.

Hyperactive children and children with oppositional defiant disorder compared to "normal" children have similar "go" reaction times, but have longer stop times (e.g., Oosterlaan, Logan, & Sergeant, 1998). Methylphenidate improves children's stop times (Tannock, Schachar, Car, Chajczyk, & Logan, 1989). A study by Tannock, Schachar, and Logan shows various dose effects for stimulant medication. Pharmacologically, methylphenidate stimulates the inhibition circuitry of the brain via dopaminergic and serotonergic mechanisms. The Game creates social, activity, and primary reward for inhibition as well as a sense of belonging for inhibition—which appear to be dopamninergic and serotonergic respectively. The Game clearly and rapidly increases "stop" behavior, by rewarding it. The Game is not like most behavior programs (e.g., Kolko, Bukstein, & Barron, 1999) that reward positive behavior (e.g., social skills or attention to task); the Game rewards *not* doing things such as blurting, interrupting, getting out of seat, etc. All behavior modification is not the same in effectiveness on children with these attention or behavior problems, even with or without the use of medication (e.g., Baldwin, 1999; Northup et al., 1999). The Game is different from most behavioral protocols in that it is group based, decreases peer reinforcement for antisocial behaviors, and provides yoked individual and group rewards. The use of rewards for attention or positive behavior for individual behavior does not seem to have the same power compared with medication (e.g., Solanto, Wender, & Bartell, 1997). The fact that this simple Game can have profound long-term effects on the "stop circuitry" is very promising from a putative medical explanation of it as a potential behavioral vaccine.

MAKING THE GAME INTO A UNIVERSAL BEHAVIORAL VACCINE

Good research and best practices do not necessarily translate into public benefit. An effective behavioral vaccine must overcome a number of barriers. First, policymakers must be sold on the idea. Second, the vaccine must be appropriately packaged for delivery. Third, the vaccine must have appropriate infrastructure to support diffusion and practice. Fourth, regulations, policies, and even laws may need to change to support the distribution of a behavioral vaccine. Fifth, current practitioners may need enticement to change. It is wise to note that it took some 80 years to make the practice of antiseptic hand washing common practice. Figure 5 summarizes what is required to create a system for a universal behavioral vaccine.

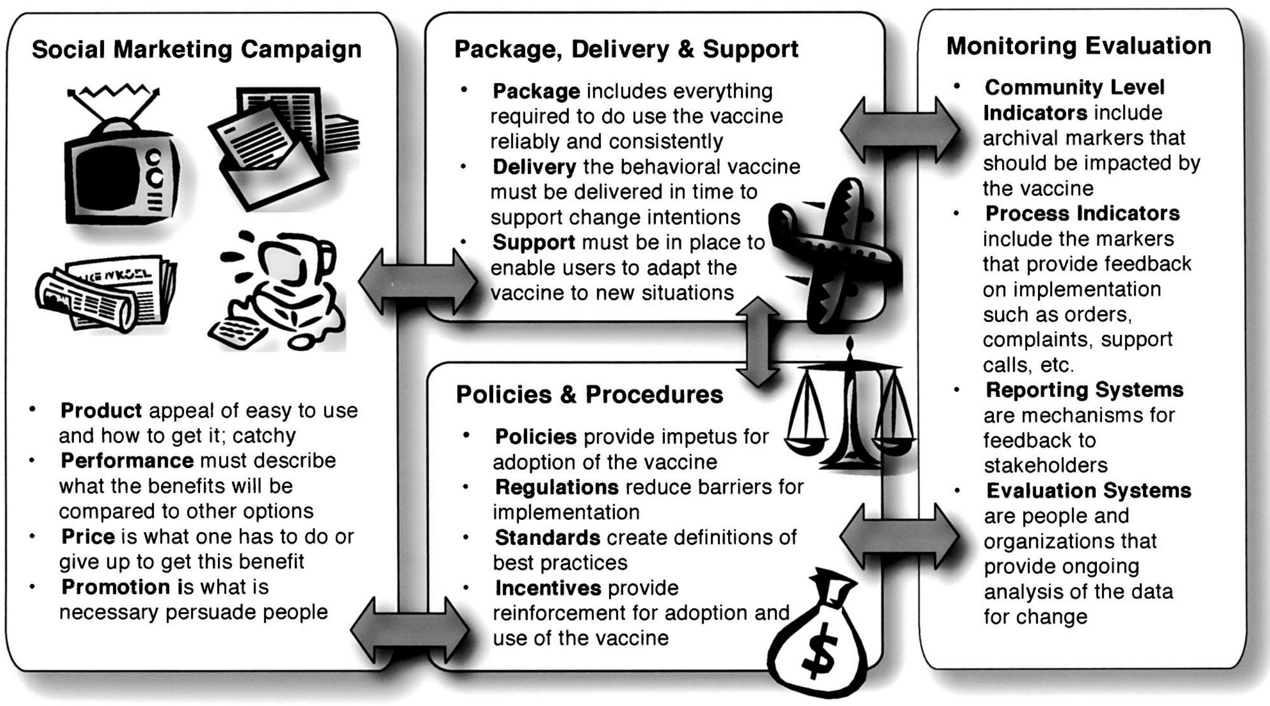

Fig. 5. System diagram for behavioral vaccine.

$$\frac{\text{Resource Cost (e.g., people, time and/or money)}}{\left(\text{Likely Prevention Effect} = \text{Power of Prevention Tools (effect size)} \times \text{Percentage of population reached} \times \text{Effects across time, people \& places} - \text{Negative Side Effects}\right)}$$

Fig. 6. Cost-effectiveness formula.

"Selling" the Game to Policymakers

Proven practices can take decades to become common practice, with many lives in forfeit as a consequence. A public-health model of prevention envisions that most effective practices must be universal for positive effect. This can be seen in a formula shown in Fig. 6. For a behavioral vaccine to work, the formula requires that the resource cost per participant be low, the effects potent, and the reach of the strategy be wide and long-lasting with few adverse side effects.

The Game works potentially well in this formula. The costs of implementation are low compared to other alternatives. The comparison between several alternatives illustrates the point. These types of data are crucial for selling state policymakers on the benefits of a behavioral vaccine.

Medication costs about $70 per child per month, plus medical supervision. Just 10 children in a school will cost at least $7,000 per year. Long-term positive results of medication are not well documented by comparison. Trademarked interventions such as Second Step, which are highly rated or extolled, have little or no impact on aggression in the classroom (e.g., Grossman et al., 1997), yet may cost at least $10,000 per school to use. (Note: this is not an exhaustive analysis of all the rival strategies).

Measures of lifetime prevention benefits are microscopic from a mathematical perspective at the time of this writing. Favorite strategies such as character education, peer mediation, conflict mediation, or police officers on campus have little or no effect size impact at this writing, though future studies or publications might change that. The fact that the Game might only cost a few hundred dollars per classroom to implement and reduce placement in special services represents an immediate cost savings; its long-term cost-effectiveness becomes even more impressive. For example, the long-term effects on reduced special education and correctional expenditures from the use of the Game are calculable and mind-boggling. Here are a just a few of the implications of the Game, if used widely in primary grades, on projections of public expenditures in a decade, for Wyoming—a state with the smallest population of all the 50 states yet with very high rates of multiproblem behavior that merit prevention. Why the example of Wyoming? Having just completed an extremely detailed blueprint for prevention of substance abuse in Wyoming (Embry & McDaniel, 2001), the author has easy access to state budget numbers.

- A 5% reduction in special education placement, not improbable based on the results from the Baltimore Prevention Project, could potentially save $2–4 million dollars per year—which has grown from $50 million to $83 million per budget period.
- A 2% reduction in involvement with corrections, not excessive based on the Baltimore Prevention Project, might yield at least $3–10 million per year in projected savings based on an analysis of growth in arrests of juveniles for serious drug arrests by the Wyoming Statistical Analysis Center at the University of Wyoming.
- A 4% reduction in lifetime prevalence of tobacco use, again not improbable from the Baltimore Prevention Project, could save the state millions of dollars per year in deferred medical costs associated with tobacco-related diseases based on cost data calculated by the U.S. Centers for Disease Control and Prevention for the state of Wyoming.

These savings from a prevention-effect sum to something like $15–20 million per year over time in Wyoming. What might be the cost of the prevention effort? There are about 5,000 1st and 2nd graders in Wyoming total. If the Game cost $200 per child per year to implement in those grades, the annual cost of implementation would run about $1,000,000 per year and thereafter. Breakeven would occur in about 3–5 years against special-education expenses, and the lifetime savings of the prevention effort would provide even stronger cost-savings.

Packaging of a Behavioral Vaccine

Public health models versus disease or disorder models envision universal coverage. To achieve the large-scale prevention or vaccine effect from something like the Game, it will be necessary to solve a number of problems for widespread social marketing:

1. *Make the research-based prevention strategies easy to use in the real world*. The public-domain protocols for the Game are not easy to use or understand. During the past 2 years, the author and colleagues have been conducting open trials on exactly this concern. For example, we have found it necessary to build in simple behavioral cueing strategies to improve effectiveness (e.g., Posavac, Sheridan, & Posavac, 1999), because many new teachers do not know these strategies.
2. *Increase social acceptability of the science-based intervention.* Unless large numbers of people adopt or participate in the strategy, the prevention effect will be small. It has been over 30 years since the Game was first invented, and very few classrooms use it nationally. Although the underlying principles are rock solid scientifically, they do require some social marketing elements. The emphatic behavioral language of the original research manual used by Kellam is potentially off-putting to many who typically have little exposure to such language, witnessed by the fact that strong behavioral concepts can impair adoption (e.g., Tingstrom, 1994).
3. *Integrate interventions for more difficult children in the front end*. Although the Game has powerful effects for aggressive children, staff typically voice worry about the children who are seen as the "worst kid ever." Having some front-end strategies for staff to use with such children when introducing the Game or "customizing it" could provide a greater confidence for the adoption of the Game as a sound practice. Explicit links need to be built in for more intensive clinical interventions for children who require higher doses of intervention such as in classroom behavioral coaching (e.g., Kotkin, 1998). Providing explicit components for higher risk young people would also minimize the chance that local innovations might combine to produce adverse effects.
4. *Strengthen linkages to other science-based strategies*. The Game has excellent results in reducing aggression and disruptive behaviors. This is good but not good enough. For example, the reduction in problem behavior only modestly translates into improvements in academic performance, unless there are other strategies introduced. The decline in problem behavior sets the stage for potent academic interventions such as class-wide peer tutoring (e.g., Greenwood, Terry, Utley, & Montagna, 1993), peer-assisted learning (e.g., Mathes, Howard, Allen, & Fuchs, 1998), or cooperative learning (e.g., Slavin, 1992). Presently, the author and colleagues have been conducting pilot efforts on such integration, combining several strategies. Although Kellam and colleagues originally tested both the Game and Mastery Learning singly and in combination, we found Mastery Learning simply not possible to implement in the current conditions of U.S. schools. Adoption and use of the Game as a daily practice would seem to be hypothetically better (and testable) if linked explicitly with some compatible empirically driven strategies that also improved academics.
5. *Address common barriers for adoption*. A well-proven science-based strategy can elicit many practical, emotional, or logistical barriers. The private sector typically responds to such issues by figuring out how to remove barriers to purchase or adoption, which is not always the case in the public sector. Current field trials have identified some significant barriers to adoption of the Game as a behavioral vaccine. Each barrier has testable potential solutions. Barriers and potential solutions follow:
   - *Restricted staff development time*. Some states or local districts now only have a few days available for any staff development. Mass media, Internet, and other approaches might help resolve this barrier. Mini demonstrations might be another mechanism.
   - *Competing demands for staff development time*. Major federal, state, or local initiatives with funding contingencies or political consequences attached tend to compete for staff

development time. Part of the promotional elements need to test whether putative linkages to these other demands improves adoption and diffusion.

- *Existing activities or procedures that might be threatened by the Game*. The doctors of Vienna did not welcome the innovation by Semmelweiss, despite its scientific logic. Classically, innovations like the Game appeal to the "innovators" or "early adopters" in diffusion models (e.g., Rodgers, 1995)—but not if framed as bureaucratic mandate and especially if the innovators or early adopters have developed something from their own time investment, while the developers of science-based protocols diminish the potential for the practices developed by the innovators. Again, marketing appeals to different types of people on the wave of adoption postulated by Rodgers needs to be tested.
- Perceived as overwhelming by staff who may be experiencing depression or burnout. Different models of delivery need to be tested to determine how the Game or any school-based behavioral vaccine might be diffused in school settings where depression or burnout are common. (This problem cannot be underestimated. I have been shown tightly held data from various districts, suggesting that antidepressant medication use is one of the highest cost centers in their health plans—which needs to be verified in a national study.)
- Beliefs about causation that reinforce inaction (e.g., "we can't do anything until the families change"). In general, marketing research suggests that testimonial-based promotions would be effective in overcoming this barrier, yet this is not the way that science-based practices are typically promoted.
- A belief that children should not be reinforced for behavior, because of such popular books as *Punished by Rewards* (Kohn, 1993). The belief is surprisingly widespread based on the number of objections and comments I get in seminars, and already identified as a significant barrier to adoption in prior research (Tingstrom, 1994). Changing this belief and related behaviors needs to be experimental tested.

Infrastructure for a Behavioral Vaccine

All materials have to be manualized and standardized in ways compatible with current issues and concerns of potential stakeholders. The original research manuals and publications are typically not standardized. After that, a considerable amount of infrastructure must be created to support the rapid dissemination of a behavioral vaccine.

1. Training strategies must be developed that can be sustained in diverse settings and organizations. If trainers with advanced degrees, certain professional qualifications, or job titles can only successfully diffuse a strategy, then the diffusion will be inherently limited. The training capacity depends on extensive documentation, support materials, "error proof" instructions, and extensive flexibility to deal with diverse objections and problems likely to happen in the field. The materials from research projects do not typically meet these criteria.
2. Implementation strategies must include ways to reduce backfires, increase fidelity of implementation, and facilitate generalization across time, people, and places. These issues are not typically addressed in research studies. At the same time, the implementation strategies must encourage principle-driven innovation and adaptation, as this author has found an inverted U-shaped curve in past large-scale studies of the diffusion of behavioral explicit prevention strategies in school settings (e.g., Embry & Malfetti, 1982). That is, poor fidelity produced the worst results, modest levels of fidelity produced the best results, and high levels of fidelity also produced poor results. What seems to happen with medium level of fidelity, based on my observation, is that people are more focused on behavior change, adjusting their actions to produce result. Very high fidelity seems to be driven by adherence to process ("by the book"), which may not respond to poor behavioral outcomes. How to structure this kind of principle-driven implementation and adaptation in the context of fidelity of implementation is also a question that needs experimental testing in the field of behavioral science to further the diffusion of any behavioral vaccine.

3. Strategies and incentives will be required to help organizations and individuals adopt the Game as an innovation. A good idea is not enough. Local service providers for example may be wedded to their particular program or approach, which may or may not have scientific or empirical validity. Interestingly, research on other behavior approaches shows the power of incentives or other organizational strategies for increasing adoption, which is evident from long-term research initiated by Denise Gottfredson (e.g., 1988) on delinquency prevention or from the experimental analysis studies on the use of seat belts or car seats.
4. The entire package or approach must be able to ramp up to very large scale, which requires distribution, marketing, and technical support. The package or program must be sustainable in different cultural contexts. These issues have not been previously addressed as fully as they need to be but can be in the context of large-scale diffusion.
5. An entire marketing campaign must be created to encourage adoption and use, and such a campaign must have sufficient reach and exposure for effectiveness. Such campaigns are rare except by commercial products with high profitability, like prescription drugs. Marketing campaigns might test such variables as inquiries to obtain a kit, recruitment success for workshop participation, early use after training, and word-of-mouth marketing effects as a result of such campaigns.

Policies to Support a Behavioral Vaccine

Many policy issues need attention to make something like the Good Behavior Game a universal behavioral vaccine. I list a few, which have emerged in the past 2 years of field trials and state policy development work:

1. State Departments of Health need to be directed by the Governor, the Legislature, or both to implement behavioral vaccines. This might be achieved through the vehicle of the various federally mandated Governor's Advisory Boards for Title IV Safe and Drug Free Communities, the Juvenile Justice Board, etc.
2. Federal block grant funds such as Title IV to schools, juvenile justice, maternal health, and other such funds need to be consolidated by executive order to support statewide behavioral vaccines instead of Balkanized efforts so that a universal approach is justified and leveraged.
3. State Departments of Health, and Departments of Education or Public Instruction, Family or Child Services need to issue combined standards of prevention and early intervention that support a public-health approach to behavioral vaccines.
4. Legislatures may need to pass special legislation that allows governmental departments or "quangos" (quasi-governmental agencies) to mix public money and marketing funds from the private sector (sponsors) to support behavioral vaccines, so that incentives and other considerations may be undertaken.
5. State Medicaid provisions often need to be clarified so that qualifying practitioners might write a prescription for the behavioral vaccine and be appropriately reimbursed. Such provisions would allow, for example, a general practitioner to write a behavior prescription for something like the Good Behavior Game for a child's classroom, have the "prescription" paid for by Medicaid, and be reimbursed for the consult or follow-up. Presently, incentives only work for a general practitioner to write prescriptions for such things as medication for behavioral disorders, never to prescribe something like a behavioral intervention that must be purchased.
6. State Departments of Education or Public Instruction need to issue policies or procedures naming research procedures like the Game as desirable procedures for inclusion or mainstreaming of children with individual education plans (IEP's) or Section 504 Rehabilitation Plans.
7. The State Departments of Education or Public Instruction in conjunction with the State Attorney Generals may need to clarify that the public posting of team points for the Game does not violate the Family Educational Rights and Privacy Act (FERPA) regulations.

Monitoring and Evaluating the Impact of a Behavioral Vaccine

Most behavioral scientists conduct controlled experiments, typically seeking an effort with high internal validity. A behavioral vaccine, by nature, seeks to have broad community level impact—to decrease the population level indices. Public accountability as well as marketing of the vaccine also gains from high-quality monitoring.

The monitoring and evaluation might proceed with some of the following:

1. Extensive monitoring of the uptake and rates of the behavioral vaccine will be required, such as the number of Game kits requested, reusable supplies ordered (a proxy measure for fidelity), school entries in community competitions using the game, or other such markers.
2. Monitoring of archival records such as per capita rates of Schedule II medications used for treatment of disorders typically targeted by the Game collected from the state pharmacy board, Medicaid, or the local health care providers; nurses' office visits for medicine checks; etc.
3. The State Department of Health might use a standardized tool such as Strengths and Difficulties Questionnaire (Goodman, 1997), which is a brief, clinically normed instrument that compares well to the Child Behavior Checklist (e.g., Goodman & Scott, 1999) to monitor prevalence rates of key *DSM-IV* diagnoses at school enrollment, public health visits, etc.
4. A consortium of federal, state, or private groups should undertake a longitudinal sample to follow for exposure to the Game, examining the impact of the Game interacting with known polygenic cofactors predicting multiproblem behavior such as various alleles of the dopamine receptors and transporters, using such tools as buccal smears and SNP analyses (e.g., Comings et al., 2000). Such a longitudinal study might be augmented by other physiological measures that are known to be correlated with outcomes, such as heart rate and brain activity (e.g., Raine et al., 1997). Such a study would help answer some of the hypothesized interactions of behavioral outcomes between environment and polygenic mechanisms from an experimental way instead of just a correlational perspective.
5. The synergy of different types of behavior vaccines needs to be tested, because multiproblem behavior has multiple vectors (e.g., parenting) that might be ameliorated by research-based protocols for parenting that can be delivered in multiple contexts or levels with prospects of success (e.g., Sanders, 1999). It is quite conceivable that certain combinations of behavioral vaccines might confer considerable "resistance to" adverse developmental outcomes such as substance abuse, delinquency, and school failure.

The practicalities of a public-health level implementation make it difficult to have classic randomized-control group study. Several possibilities do exist to provide some element of control such as a multiple baseline across communities or age groups. Over time, epidemiological monitoring such as the commonly used Youth Risk Behavior Survey (Centers for Disease Control and Prevention, 2000) might be used to measure the longer-term impact of exposure to the Game in elementary school, by matching grade school and classroom exposure to the Game or other interventions, and the avoidance of multiproblem outcomes in a dose–response-type quasi-experimental paradigm.

The idea of behavioral vaccines—simple actions that can be repeated by nearly everyone on a daily basis with positive health effects—has face validity from the public health model. Antiseptic hand washing is a powerful example, and there have been other examples in very recent history such as seat-belt and car-seat use. The concept of a universal behavioral vaccine has intuitive appeal based on epidemiological and intervention studies of multiproblem behavior such as substance abuse, delinquency, violence, and other ills. Epidemiological studies of mutliproblem behavior suggest that there are apparent behaviors (e.g., early disruptiveness) that could be modified and reduce the future occurrence of the adverse outcomes. A behavioral vaccine for multiproblem behavior would have to be low cost, easy to use, have powerful effects, and be capable of wide distribution across the target population. A potential candidate for a behavioral vaccine against multiproblem behavior would have to have a strong history of efficacy and effectiveness studies, and be adaptable to many different circumstances. The Good Behavior Game, first reported by Barrish et al. (1969) represents a strong candidate for a behavioral vaccine, because of the simplicity and multiple replications of positive results in efficacy studies with strong long-term results in effectiveness trials. Early field replications suggest that the Game can be used in very diverse circumstances. Large-scale testing of the Game as a behavioral vaccine could provide a rich source of theory building for the diffusion of science-based prevention practices, because the Game is rare in having measurable effects based on a single classroom instead of schoolwide adoption. Against the common practice of encouraging communities to engage in an elaborate processes of prevention logic models or the abnegation of powerful behavioral vaccines used across the country or states could substantially improve developmental outcomes, benefit many diverse stakeholders, and save substantial sums of government expenditures at scale.

REFERENCES

Andersson, T., Mahoney, J. L., Wennberg, P., Kuehlhorn, E., & Magnusson, D. (1999). The co-occurrence of alcohol problems and criminality in the transition from adolescence to young adulthood: A prospective longitudinal study on young men. *Studies on Crime and Crime Prevention*, *8,* 169–188.

Ary, D. V., Duncan, T. E., Biglan, A., Metzler, C. W., Noell, J. W., & Smolkowski, K. (1999). Development of adolescent problem behavior. *Journal of Abnormal Child Psychology*, *27,* 141–150.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1,* 91–97.

Baldwin, S. (1999). Applied behavior analysis in the treatment of ADHD: A review and rapprochement. *Ethical Human Sciences and Services*, *1,* 35–59.

Barlow, D. H., & Hersen, M. (1973). Single-case experimental designs: Uses in applied clinical research. *Archives of General Psychiatry*, *29,* 319–325.

Barrish, H., Saunders, M., & Wolf, M. M. (1969). Good Behavior Game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, *2,* 119–124.

Biglan, A. (2001, August 9–10). *Palo Alto Summit on Improving the Prevalence of Successful Adolescence*. Keynote speech at 2nd National Conference on Drug Abuse Prevention Research: A Progress Update, Omni Shoreham Hotel, Washington, DC.

Bostow, D., & Geiger, O. G. (1976). Good Behavior Game: A replication and systematic analysis with a second grade class. *SALT: School Applications of Learning Theory*, *8,* 18–27.

Buss, D. M. (1984). Evolutionary biology and personality psychology: Toward a conception of human nature and individual differences. *American Psychologist*, *39,* 1135–1147.

Carlson, C. L., Pelham, W. E., Milich, R., & Dixon, J. (1992). Single and combined effects of methylphenidate and behavior therapy on the classroom performance of children with attention-deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*, *20,* 213–221.

Catalano, R. (2001, August 10). *Understanding community risk and protective factors*. Paper presented at the Second National Conference on Drug Abuse Prevention Research, National Institute on Drug Abuse and Robert Wood Johnson Foundation, Washington, DC.

Catania, A. C. (1992). *Learning* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Centers for Disease Control and Prevention (CDC). (2000). Youth risk behavior surveillance—United States, 1999. *Mortality and Morbidity Weekly Report*, *49,* No. SS-5.

Champion, L.A., Goodall, G.M., & Rutter, M. (1995). Behavioral problems in childhood and stressors in early adult life: A 10-year follow-up of London school children. *Psychological Medicine*, *25,* 231–246.

Cloninger, R. C. (1994). A systematic method for clinical description and classification of personality variants. *Current Opinion in Neurobiology*, *4,* 573–588.

Cloninger, R. C., Adolfsson, R., & Svrakic, N. (1996). Mapping genes for human personality. *Nature Genetics*, *12,* 3–4.

Cloninger, R. C., Sigvardson, S., & Bohman, M. (1988). Childhood personality predicts alcohol abuse in young adults. *Alcohol Clinical Experimental Research*, *122,* 494–505.

Comings, D. E. (1995). The role of genetic factors in conduct disorder based on studies of Tourette syndrome and ADHD probands and their relatives. *Journal of Developmental and Behavioral Pediatrics*, *16,* 142–157.

Comings, D. E., Gade, R., Muhleman, D., & MacMurray, J. (1996). Role of the HTR1A serotonin receptor gene in Tourette syndrome and conduct disorder. *Psychiatric Genetics*, *6,* 166.

Comings, D. E., Gade, R., Wu, S., Chiu, C., Dietz, G., Muhleman, D., et al. (1996). Dopamine D1 receptor gene and addictive behaviors. *Psychiatric Genetics*, *6,* 158.

Comings, D. E., Gade-Andavolu, R., Gonzalez, N., Wu, S., Muhleman, D., Blake, H., et al. (2000). Comparison of the role of dopamine, serotonin, and noradrenaline genes in ADHD, ODD and conduct disorder: Multivariate regression analysis of 20 genes. *Clinical Genetics*, *57,* 178–196.

Darveaux, D. X. (1984). The Good Behavior Game plus merit: Controlling disruptive behavior and improving student motivation. *School Psychology Review*, *13,* 510–514.

Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm peer groups and problem behavior. *American Psychologist*, *54,* 755–764.

Dishion, T. J., Spracklen, K. M., Andrews, D. W., & Patterson, G. R. (1996). Deviancy training in male adolescents friendships. *Behavior Therapy*, *27,* 373–390.

Disney, E. R., Elkins, I. J., McGue, M., & Iacono, W. G. (1999). Effects of ADHD, conduct disorder, and gender on substance use and abuse in adolescence. *American Journal of Psychiatry*, *156,* 1515–1521.

Dolan, L. J., Kellam, S. G., Brown, C. H., Werthamer-Larson, L., Rebok, G. W., Mayer, L. S., et al. (1993). The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement. *Journal of Applied Developmental Psychology*, *14,* 317–345.

Eddy, J. M., Reid, J. B., & Fetrow, R. A. (2000). An elementary-school based prevention program targeting modifiable antecedents of youth delinquency and violence: Linking the Interests of Families and Teachers (LIFT). *Journal of Emotional and Behavioral Disorders*, *8,* 165–176.

Embry, D. D. (2002). Nurturing the genius of genes: The new frontier of education, therapy, and understanding of the brain. *Brain and Mind*, *9,* 101–132.

Embry, D. D., & Flannery, D. J. (1999). Two sides of the coin: Multilevel prevention and intervention to reduce youth violent behavior. In D. J. Flannery & C. R. Huff (Eds), *Youth violence: Prevention, intervention and social policy.* Washington, DC: American Psychiatric Press.

Embry, D. D., Flannery, D. J., Vazsonyi, A. T., Powell, K. E., & Atha, H. (1996). Peacebuilders: A theoretically driven, school-based model for early violence prevention. *American Journal of Preventive Medicine*, *12,* 91–100.

Embry, D. D., & Malfetti, J. M. (1982). *A field trial of the safe playing program*. Falls Church, VA: American Automobile Association Foundation for Traffic Safety.

Embry, D. D., & McDaniel, R. (2001). *Reclaiming Wyoming: A comprehensive blueprint for substance abuse treatment, intervention and prevention*. Cheyenne, WY: Substance Abuse Division, Department of Health. (downloadable at www.wywins.net or www.paxtalk.com)

Embry, D. D., & Straatemeier, G. (2001). *The PAX Acts Game Manual: How to apply the Good Behavior Game*. Tucson, AZ: PAXIS Institute.

Fishbein, J. E., & Wasik, B. H. (1981). Effect of the Good Behavior Game on disruptive library behavior. *Journal of Applied Behavior Analysis*, *14,* 89–93.

Frick, P. J., Kuper, K., Silverthorn, P., & Cotter, M. (1995). Antisocial behavior, somatization, and sensation-seeking behavior in mothers of clinic-referred children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *34,* 805–812.

Gadow, K. D., Nolan, E. E., Sverd, J., Sprafkin, J., & Paolicelli, L. (1990). Methylphenidate in aggressive-hyperactive boys. I: Effects on peer aggression in public school settings. *Journal of the American Academy of Child and Adolescent Psychiatry*, *29,* 710–718.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38,* 581–586.

Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology*, *27,* 17–24.

Gottfredson, G. D. (1988). An evaluation of an organizational development approach to reducing school disorder. *Evaluation Review*, *11*, 739–763.

Grandy, G. S., Madsen, C. H., & De Mersseman, L. M. (1973). The effects of individual and interdependent contingencies on inappropriate classroom behavior. *Psychology in the Schools*, *10,* 488–493.

Greenberg, M., Domitrovich, C., & Bumbarger, B. (1999). *Preventing mental disorders in school-age children: A review of the effectiveness of prevention programs.* Center for Mental Health Services (CMHS), Substance Abuse Mental Health Services Administration, U.S. Department of Health and Human Services. Rockville, MD. http://www.prevention.psu.edu/CMHS.html

Greenwood, C. R., Terry, B., Utley, C. A., & Montagna, D. (1993). Achievement, placement, and services: Middle school benefits of Classwide Peer Tutoring used at the elementary school. *School Psychology Review*, *22,* 497–516.

Grossman, D. C., Neckerman, H. J., Koepsell, T. D., Liu, P., Asher, K. N., Beland, K., et al. (1997). Effectiveness of a violence prevention curriculum among children in elementary school. *JAMA*, *277,* 1605–1611.

Harris, V. W., & Sherman, J. A. (1973). Use and analysis of the "Good Behavior Game" to reduce disruptive classroom behavior. *Journal of Applied Behavior Analysis*, *6,* 405–417.

Hegerle, D. R., Kesecker, M. P., & Couch, J. V. (1979). A behavior game for the reduction of inappropriate classroom behaviors. *School Psychology Review*, *8,* 339–343.

Herrnstein, R. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior*, *13,* 243–266.

Huber, H. (1979). The value of a behavior modification programme, administered in a fourth grade class of a remedial school. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, *28,* 73–79.

Ialongo, N., Poduska, J., Werthamer, L., & Kellam, S. (2001). The distal impact of two first-grade preventive interventions on conduct problems and disorder in early adolescence. *Journal of Emotional and Behavioral Disorders*, *9,* 146–160.

Ialongo, N. S., Werthamer, L., Kellam, S. G., Brown, C. H., Wang, S., & Lin, Y. (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior. *American Journal of Community Psychology*, *27,* 599–641.

Johnson, M. R., Turner, P. F., & Konarski, E. A. (1978). The Good Behavior Game: A systematic replication in two unruly transitional classrooms. *Education and Treatment of Children*, *1,* 25–33.

Kellam, S. G., & Anthony, J. C. (1998). Targeting early antecedents to prevent tobacco smoking: Findings from an epidemiologically based randomized field trial. *American Journal of Public Health*, *88,* 1490–1495.

Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the First grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology*, *10,* 165–185.

Kellam, S. G., Ling, X., Merisca, R., Hendricks, Brown, C., Ialongo, N. (2000). Erratum: The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development & Psychopathology*, *12,* 107.

Kellam, S. G., Mayer, L. S., Rebok, G. W., & Hawkins, W. E. (1998). Effects of improving achievement on aggressive behavior and of improving aggressive behavior on achievement through two preventive interventions: An investigation of causal paths. In B. P. Dohrenwend (Ed.), *Adversity, stress, and psychopathology* (pp. 486–505). New York: Oxford University Press.

Kellam, S. G., & Rebok, G. W. (1992). Building developmental and etiological theory through epidemiologically based preventive intervention trials. In J. McCord & R. E. Tremblay (Eds.), *Preventing antisocial behavior: Interventions from birth through adolescence* (pp. 162–194). New York: Guilford Press.

Kellam, S. G., Rebok, G. W., Ialongo, N., & Mayer, L. S. (1994). The course and maleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trial. *Journal of Child Psychology and Psychiatry*, *35,* 259–281.

Kohn, A. (1993). *Punished by rewards: The trouble with gold stars, incentive plans, A's, praise, and other bribes.* Boston, MA: Houghton Mifflin Co.

Kolko, D. J., Bukstein, O. G., & Barron., J. (1999). Methylphenidate and behavior modification in children with ADHD and comorbid ODD or CD: Main and incremental effects across settings. *Journal of the American Academy of Child and Adolescent Psychiatry*, *38,* 578–586.

Kosiec, L. E., Czernicki, M. R., & McLaughlin, T. F. (1986). The Good Behavior Game: A replication with consumer satisfaction in two regular elementary school classrooms. *Techniques*, *2,* 15–23.

Kosterman, R., Hawkins, J. D., Spoth, R., Haggerty, K. P., & Zhu, K. (1997). Effects of a preventive parent-training intervention on observed family interactions: Proximal outcomes from preparing for the drug free years. *Journal of Community Psychology*, *25,* 337–352.

Kotkin, R. (1998). The Irvine Paraprofessional Program: Promising practice for serving students with ADHD. *Journal of Learning Disabilities*, *3,* 556–564.

Krueger, R. F., Moffitt, T. E., Caspi, A., Bleske, A., & Silva, P. A. (1998). Assortative mating for antisocial behavior: Developmental and methodological implications. *Behavior Genetics*, *28,* 173–186.

Lazzaro, I., Gordon, E., Whitmont, S., Meares, R., & Clarke, S. (2001). The modulation of late component event related potentials by pre-stimulus EEG theta activity in ADHD. *International Journal of Neuroscience*, *107,* 247–264.

Lipsey, M. W. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In T. D. Cook, H. Cooper, D. S. Corday, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis, & F. Musteller (Eds.), *Meta-analysis for explanation: A case book for explanation* (pp. 83–125). New York: Russell Sage.

Loeber, R., Stouthammer-Loeber, M., & White, H. R. (1999). Developmental aspects of delinquency and internalizing problems and their association with persistent juvenile substance use between ages 7 and 18. *Journal of Clinical Child Psychology*, *28,* 322–332.

Lynch, A. (2001). *Thought contagion: How belief spreads through society*. New York: Basic Books.

Madrid, A., Anderson, B., Lee, J., MacMurray, J., & Comings, D. E. (2001). A possible role for stress in mediating variability in outcome of studies of the DRD2 Taq I polymorphism and alcoholism. *Alcoholism*, *23,* 117–122.

Malott, R. W., Whaley, D. L., & Malott, M. E. (1997). *Elementary principles of behavior* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall

Mathes, P. G., Howard, J. K., Allen, S. H., & Fuchs, D. (1998). Peer-assisted learning strategies for first-grade readers: Responding to the needs of diverse learners. *Reading Research Quarterly*, *33,* 62–94.

Medland, M. B., & Stachnik, T. J. (1972). Good Behavior Game: A replication and systematic analysis. *Journal of Applied Behavior Analysis*, *5,* 45–51.

Moffitt, T. E. (1990). Juvenile delinquency and attention deficit disorder: Boys' developmental trajectories from age 3 to age 15. *Child Development*, *61,* 893–910.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, *10,* 674–701.

Northup, J., Fusilier, I., Swanson, V., Huete, J., Bruce, T., Freeland, J., et al. (1999). Further analysis of the separate and interactive effects of methylphenidate and common classroom contingencies. *Journal of Applied Behavior Analysis*, *32,* 35–50.

Oosterlaan, J., Logan, G. D., & Sergeant, J. A. (1998). Response inhibition in AD/HD, CD, comorbid AD/HD + CD, anxious, and control children: A meta-analysis of studies with the stop task. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39,* 411–425.

Patrick, C. A., Ward, P., & Crouch, D. W. (1998). Effects of holding students accountable for social behaviors during volleyball games in elementary physical education. *Journal of Teaching in Physical Education*, *17,* 143–156.

Patterson, G. R., De Baryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. *American Psychologist*, *44,* 329–335.

Patterson, G. R., Dishion, T. J., & Yoerger, K. (2000). Adolescent growth in new forms of problem behavior: Macro- and micro-peer dynamics. *Prevention Science*, *1,* 3–1.

Patterson, G. R., Forgatch, M. S., Yoerger, K. L., & Stoolmiller, M. (1998). Variables that initiate and maintain an early-onset trajectory for juvenile offending. *Development and Psychopathology*, *10,* 531–547.

Patterson, G. R., & Stouthamer-Loeber, M. (1984). The correlation of family management practices and delinquency. *Child Development*, *55,* 1299–1307.

Pelham, W. E., Jr., & Fabiano, G. A. (2000). Behavior modification. *Child and Adolescent Psychiatric Clinics of North America*, *9,* 671–688.

Phillips, D., & Christie, F. (1986). Behaviour management in a secondary school classroom: Playing the game. *Maladjustment and Therapeutic Education*, *4*, 47–53.

Posavac, H. D., Sheridan, S. M., & Posavac, S. S. (1999, April). A cueing procedure to control impulsivity in children with attention deficit hyperactivity disorder. *Behavior Modification*, *23,* 234–253.

Quist, J. F., & Kennedy, J. L. (2001). Genetics of childhood disorders: XXIII. ADHD, Part 7: The serotonin system. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40,* 253–256.

Raine, A., Venables, P. H., & Mednick, S. A. (1997). Low resting heart rate at age 3 years predisposes to aggression at age 11 years: Findings from the Mauritius Joint Child Health Project. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36,* 1457–1464.

Reid, J. B., Eddy, J. M., Fetrow, R. A., & Stoolmiller, M. (1999). Description and immediate impacts of a preventative intervention for conduct problems. *American Journal of Community Psychology*, *24,* 483–517.

Rogers, E. M. (1995). *Diffusion of Innovations* (4th Ed.). New York: The Free Press.

Rutter, M. (1985). Family and school influences on cognitive development. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *26,* 683–704.

Rutter, M., Maughan, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours: Secondary schools and their effects on children.* Somerset, UK: Open Books.

Saigh, P. A., & Umar, A. M. (1983). The effects of a Good Behavior Game on the disruptive behavior of Sudanese elementary school students. *Journal of Applied Behavior Analysis*, *16,* 339–344.

Salend, S. J., Reynolds, C. J., & Coyle, E. M. (1989). Individualizing the Good Behavior Game across type and frequency of behavior with emotionally disturbed adolescents. *Behavior Modification*, *13,* 108–126.

Sanders, M. R. (1999). Triple P—Positive Parenting Program: Towards an empirically validated multilevel parenting and family support strategy for the prevention of behavior and emotional problems in children. *Clinical Child and Family Psychology Review*, *2,* 71–90.

Schaie, K. W., & Baltes, P. B. (1975). On sequential strategies in developmental research: Description or explanation. *Human Development*, *18,* 384–390.

Shedler, J., & Block, J. (1990). Adolescent drug use and psychological health: A longitudinal inquiry. *American Psychologist*, *45,* 612–630.

Slavin, R. E. (1992). When and why does cooperative learning increase achievement? Theoretical and empirical perspectives. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 145–173). New York: Cambridge University Press.

Slutske, W. S., Heath, A. C., Dinwiddie, S. H., Madden, P. A., Bucholz, K. K., Dunne, M. P., et al. (1997). Modeling genetic and environmental influences in the etiology of conduct disorder: A study of 2,682 adult twin pairs. *Journal of Abnormal Psychology*, *106,* 266–279.

Slutske, W. S., Heath, A. C., Madden, P. A. F., Bucholz, K. K., Dinwiddle, S. H., Dunne, M. P., et al. (1995). Sex differences in the etiology of DSM-III-R conduct disorder. *Psychiatric Genetics*, *5,* S80.

Solanto, M. V., Wender, E. H., & Bartell, S. S. (1997). Effects of methylphenidate and behavioral contingencies on sustained attention in attention-deficit hyperactivity disorder: A test of the reward dysfunction hypothesis. *Journal of Child and Adolescent Psychopharmacology*, *7,* 123–136.

Stoolmiller, M., Eddy, J. M., & Reid, J. B. (2000). Detecting and describing preventative intervention effects in a universal school-based randomized trial targeting delinquent and violent behavior. *Journal of Consulting and Clinical Psychology*, *68,* 296–306.

Sulzer-Azaroff, B., & Mayer, G. R. (1991). *Behavior analysis for lasting change.* Chicago, IL: Holt, Rinehart and Winston.

Swiezy, N. B., Matson, J.L., & Box, P. (1992). The Good Behavior Game: A token reinforcement system for preschoolers. *Child and Family Behavior Therapy*, *14,* 21–32.

Tannock, R., Schachar, R. J., Car, R. P., Chajczyk, D., & Logan, G. D. (1989). Effects of methylphenidate on inhibitory control in hyperactive children. *Journal of Abnormal Child Psychology*, *17,* 473–491.

Tannock, R., Schachar, R. J., & Logan, G. (1995). Methylphenidate and cognitive flexibility: Dissociated dose effects in hyperactive children. *Journal of Abnormal Child Psychology*, *23,* 235–266.

Tingstrom, D. H. (1994). The Good Behavior Game: An investigation of teachers' acceptance. *Psychology in the Schools*, *31,* 57–65.

Tremblay, R. E., Masse, B., Perron, D., & Leblanc, M. (1992). Early disruptive behavior, poor school achievement, delinquent behavior, and delinquent personality: Longitudinal analyses. *Journal of Consulting and Clinical Psychology*, *60,* 64–72.

Tremblay, R. E., Pagani-Kurtz, L., Masse, L. C., Vitaro, F., & Pihl, R. (1995). A bimodal preventive intervention for disruptive kindergarten boys: Its impact through mid-adolescence. *Journal of Consulting and Clinical Psychology*, *63*, 560–568.

Tremblay, R. E., Pihl, R. O., Vitaro, F., & Dobkin, P. L. (1994). Predicting early onset of male antisocial behavior from preschool behavior. *Archives of General Psychiatry*, *51,* 732–739.

U.S. Surgeon General. (2001). *Surgeon General's Report on Youth Violence*. Department of Health and Human Services: Rockville, MD.

U.S. Surgeon General. (2002, June 2). An artificial eye on bioterror. *Discover*, p. 16.

Walker, H. M., Colvin, G., & Ramsey, E. (1995). *Anti-social behavior in schools: Strategies and best practices.* Pacific Grove, CA: Brooks/Cole.

Walker, H. M., Stieber, S., Ramsey, E., & O'Neill, R. (1993). Fifth grade school adjustment and later arrest rate: A longitudinal study of middle school antisocial boys. *Journal of Child and Family Studies*, *2,* 295–315.

Warner, S. P., Miller, F. D., & Cohen, M. W. (1977). Relative effectiveness of teacher attention and the Good Behavior Game in modifying disruptive classroom behavior. *Journal of Applied Behavior Analysis*, *10,* 737.

Webster-Stratton, C., & Hammond, M. (1997). Treating children with early-onset conduct problems: A comparison of child and parent training interventions. *Journal of Consulting and Clinical Psychology*, *65,* 93–109.

White, H. R., Loeber, R., Stouthamer-Loeber, M., & Farrington, D. (1999). Developmental associations between substance use and violence. *Development and Psychopathology*, *11,* 785–803.

Hammill Institute on Disabilities

Journal of Positive Behavior Interventions
2016, Vol. 18(3) 180–190

Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1098300715592355
jpbi.sagepub.com
SAGE

Promoting Positive Behavior Using the Good Behavior Game: A Meta-Analysis of Single-Case Research

Lisa Bowman-Perrott, PhD[1], Mack D. Burke, PhD[1], Samar Zaini, MS[1], Nan Zhang, MA[1], and Kimberly Vannest, PhD[1]

Abstract

The Good Behavior Game (GBG) is a classroom management strategy that uses an interdependent group-oriented contingency to promote prosocial behavior and decrease problem behavior. This meta-analysis synthesized single-case research (SCR) on the GBG across 21 studies, representing 1,580 students in pre-kindergarten through Grade 12. The TauU effect size across 137 phase contrasts was .82 with a confidence interval 95% CI = [0.78, 0.87], indicating a substantial reduction in problem behavior and an increase in prosocial behavior for participating students. Five potential moderators were examined: emotional and behavioral disorder (EBD) risk status, reinforcement frequency, target behaviors, GBG format, and grade level. Findings suggest that the GBG is most effective in reducing disruptive and off-task behaviors, and that students with or at risk for EBD benefit most from the intervention. Implications for research and practice are discussed.

Keywords

behavior(s), challenging, intervention(s), single-case designs, meta-analysis, studies

There is a need to identify effective prevention-oriented approaches to behavior management, especially at the classroom level (Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). The Good Behavior Game (GBG) is a universal behavior management strategy that uses an interdependent group-oriented contingency to promote positive classroom behaviors. First introduced by Barrish, Saunders, and Wolf (1969), the GBG was originally developed to reduce disruptive behaviors in an elementary school classroom. More recently, it has been used within prevention science (Kellam, Rebok, Ialongo, & Mayer, 1994). Embry (2002) referred to the GBG as a "behavioral vaccine" based, in part, on seminal research conducted by Kellam et al. (1994). The authors reported positive long-term impacts of the intervention on aggressive and disruptive behaviors from a large-scale epidemiological trial.

The major features of the GBG as described by Barrish et al. (1969) included the following: (a) assigning students to teams, (b) giving points to teams that exhibit inappropriate behaviors, and (c) rewarding the team that accumulated the lowest number of points (i.e., the team that exhibits the least amount of problem behavior). Depending on how the GBG is set up, more than one team can win if the criterion for winning (e.g., five or fewer points) is reached. In some instances, the GBG has been modified as follows: (a) rewarding appropriate behaviors (Crouch, Gresham, & Wright, 1985), (b) adding a merit system for simultaneously promoting academic engagement (Darveaux, 1984), (c) adding a behavioral intervention (Wright & McCurdy, 2011), (d) including a self-monitoring component (Babyak, Luze, & Kamps, 2000), (e) examining the impact of not using teams (Harris & Sherman, 1973), (f) investigating the effect of using independent and dependent (vs. interdependent) group contingencies (Gresham & Gresham, 1982), and (g) allowing individual students to earn points (Babyak et al., 2000).

The GBG is effective across a variety of problem behaviors including verbal and physical aggression (Saigh & Umar, 1983), noncompliance (Swiezy, Matson, & Box, 1992), oppositional behaviors (Leflot, van Lier, Onghena, & Colpin, 2010), hyperactive behaviors (Huizink, van Lier, & Crijnen, 2008), and out-of-seat behaviors (Medland & Stachnik, 1972). Increases in prosocial behaviors associated with the Game include on-task behaviors (Rodriguez,

[1]Texas A&M University, College Station, USA

Corresponding Author:
Lisa Bowman-Perrott, Texas A&M University, 4225 TAMU, College Station, TX 77843-4225, USA.
Email: lbperrott@tamu.edu

Action Editor: Kathleen Lane

2010), assignment completion (Darveaux, 1984), acceptance of authority (Dolan et al., 1993), and improved concentration (Dolan et al., 1993). Furthermore, positive outcomes from participation in the GBG have been observed in both general education (McGoey, Schneider, Rezzetano, Prodan, & Tankersley, 2010) and special education settings (Salend, Reynolds, & Coyle, 1989). Finally, although most of the research on the GBG has been conducted in elementary schools (Ruiz-Olivares, Pino, & Herruzo, 2010), there is promising evidence of its efficacy with middle and high school students (Kleinman & Saigh, 2011).

Previous GBG Reviews

Two GBG literature reviews and one meta-analysis have been published to date (Flower, McKenna, Bunuan, Muething, & Vega, 2014; Tankersley, 1995; Tingstrom, Sterling-Turner, & Wilczynski, 2006). Tankersley (1995) reviewed nine studies published between 1969 and 1994, including six single-case research (SCR) studies. Studies focused solely on elementary school students, with one study (Kellam et al., 1994) following up with participants in middle school. Of the nine studies, seven used the original GBG format described by Barrish et al. (1969). The remaining two studies (Fishbein & Wasik, 1981; Harris & Sherman, 1973) implemented modified versions of the Game by reinforcing positive behaviors. Tankersley (1995) reported that (a) the GBG was effective in reducing problem behaviors, (b) improvements in academic engagement could be attributed to the GBG (Darveaux, 1984), (c) social validity was high among teachers and students (Salend et al., 1989), and (d) direct observations were primarily used, with the exception of two studies which included teacher ratings of student behaviors (Kellam et al., 1994) and/or peer nominations (Dolan et al., 1993).

Tingstrom et al. (2006) synthesized 29 studies published between 1969 and 2000, 21 of which used SCR designs. Twenty of the 29 studies focused on elementary school students, 2 examined outcomes for middle and high school students, and 4 included both elementary and secondary students. One study did not report students' grade level; the remaining two studies did not investigate student outcomes. Most of the participants were "students of typical development in general education classes or students with a history of behavior problems" (Tingstrom et al., 2006, p. 241). Only a few studies examined the efficacy of the GBG with students with disabilities. In addition, studies were conducted in the United States, Germany (Huber, 1979), and Sudan (Saigh & Umar, 1983). Tingstrom et al. reported findings consistent with those of Tankersley's (1995) review. Moreover, Tingstrom et al. noted that the GBG was effective regardless of whether (a) the criterion for reinforcement was changed (Harris & Sherman, 1973) or (b) the original format described by Barrish et al. (1969) or variations of it were used (e.g., Swiezy et al., 1992). Although the literature reviews provide valuable information about the GBG, there are several limitations. First, neither reported effect sizes with confidence intervals (CIs); CIs are needed for accurately interpreting effect size data (Cooper, 2011; Thompson, 2007). Second, although they summarized several variables (e.g., reinforcement frequency, GBG format), it is not known which is more effective in promoting positive behaviors. Third, although both reviews identified the need to examine outcomes for students with disabilities, neither summarized data for this group of students. Yet, Salend et al. (1989), for example, implemented the GBG with students with emotional and behavioral disorders (EBD).

Flower et al. (2014) examined the impact of the GBG on challenging behaviors across 22 studies, 16 of which used SCR designs. They examined the impact of fidelity of implementation, training for interventionists, intervention duration, setting, and the use of rewards on problem behaviors. The authors found that although few studies reported fidelity, only one reported low fidelity. They indicated that a lower fidelity rating in that study did not minimize the benefit students gained from participating in the GBG. They concluded that (a) results did not seem to be affected by the type of interventionist training, (b) shorter interventions produced change in student behavior, (c) elementary and secondary students demonstrated a reduction in problem behaviors, and (d) the use of rewards had a positive impact on decreasing inappropriate behaviors while increasing appropriate behaviors, particularly when students found rewards reinforcing.

The Flower et al. (2014) meta-analysis extends the contribution of the literature reviews. However, several important considerations remain unaddressed. First, an investigation of the following is needed: (a) whether the GBG has differential effects for students with or at risk for EBD (as they characteristically demonstrate higher rates of challenging behavior; Reid, Gonzalez, Nordess, Trout, & Epstein, 2004), (b) whether the frequency of reinforcement moderates student outcomes, and (c) whether GBG format affects student outcomes. Second, Flower et al. only included studies published in peer-reviewed journals. As they noted, "sound implementations of the GBG conducted for dissertation . . . research may have been missed" (p. 20). Third, a measure of design quality (e.g., What Works Clearinghouse [WWC]; Kratochwill et al., 2010) was not included. Fourth, the reporting of effect sizes with CIs is missing. Given these limitations, a study that addresses these gaps and further extends the literature is essential to better understand the impact of this widely used intervention.

Purpose of the Study and Research Questions

Often, studies using SCR designs are excluded from meta-analyses (Allison & Gorman, 1993). Perhaps this is because recommended standards for quality SCR designs and evidence of treatment effects have only recently been disseminated

(Kratochwill et al., 2010). Quantitative syntheses are critical for establishing the evidence-base for effective behavioral interventions and practices (Parker & Hagan-Burke, 2007), especially with regard to SCR (Shadish, Rindskopf, & Hedges, 2008). Meta-analysis "allows researchers to arrive at conclusions that are more accurate and more credible than can be presented in any one primary study or in a non-quantitative, narrative review" (Rosenthal & DiMatteo, 2001, p. 61). The purpose of the current meta-analysis was to quantitatively analyze the SCR literature on the GBG to examine its impact: (a) for students with or at risk for EBD, (b) with regard to reinforcement frequency, (c) across target behaviors, (d) based on the GBG format used, and (e) across grade levels. Two main research questions were addressed:

Research Question 1: What is the overall effect of the GBG across studies?
Research Question 2: What are the effects of potential moderators on students' behavioral outcomes?

Method

Literature Search, Inclusion Criteria, and Design Quality

To identify relevant studies, a search of the literature was conducted using the Education Full Text, Educational Resources Information Center (ERIC), PsycINFO, and Dissertations and Theses Full Text databases. Dissertations and unpublished studies were sought for inclusion to help reduce the possibility of publication bias, the tendency for only studies yielding favorable results to be published (Rosenthal & DiMatteo, 2001). To identify the maximum number of potentially eligible studies, we used the term *Good Behavior Game*; 272 search results were obtained. The first author and two doctoral students reviewed titles and abstracts for relevance; articles were reviewed when more information was needed. Articles were excluded if they (a) included some combination of the search terms but were unrelated, (b) used a group design (as most GBG studies used SCR designs), (c) were literature reviews, (d) were duplicate studies, or (e) were studies for which a complete article copy was unavailable (e.g., older studies). Thirty-four studies remained. Studies that investigated non-behavioral outcomes or focused on teacher outcomes but did not investigate student outcomes were excluded as well, resulting in 24 studies. We also conducted an ancestral search for studies in the references of articles identified in the electronic database search; no additional studies were found. To be included, studies had to (a) implement the GBG to reduce problem behavior or increase appropriate behavior, (b) involve participants in pre-kindergarten through Grade 12, (c) be published in a peer-reviewed journal or conducted as dissertation research or an unpublished article between 1969 and 2013, (d) use an SCR design, (e) provide graphed data of student outcomes, and (f) be reported in English. To ensure that basic quality standards were adhered to from the initial pool of studies, we applied two WWC (Kratochwill et al., 2010) standards to identify studies that used a design that (a) could demonstrate experimental control (viz., reversal, multiple baseline) and (b) had at least three data points per phase. The application of these criteria yielded 21 studies for inclusion in this meta-analysis.

We then used a rubric adapted from Maggin, Chafouleas, Goddard, and Johnson (2011) to evaluate the included studies across four WWC SCR design standards (see Table 1). First, we determined whether the GBG was systematically manipulated. Second, we determined whether the design could demonstrate an experimental effect across three points in time or with three phase changes. Third, we evaluated studies using reversal designs (n = 10) to ensure that they had a minimum of four phases with at least five data points per phase to meet standards, or at least three data points per phase to meet standards with reservations. Studies using multiple baseline designs (n = 11) were evaluated to ensure that they included at least six phases and at least five data points per phase to meet standards, and three data points per phase to meet standards with reservations. Fourth, we examined each study for interobserver agreement (IOA). Based on these four standards, each study was categorized as *meets standards, meets standards with reservations*, or *does not meet standards* (Kratochwill et al., 2010). With regard to design quality, 5 of the 21 studies met standards, 9 met standards with reservations, and 7 did not meet standards (primarily because of missing IOA data on the percentage of observations; see Table 1). To establish whether there was evidence of an effect, we conducted a visual analysis of each study to determine the following: (a) the consistency of level, trend, and variability within each phase; (b) how immediate the effect was between baseline and intervention phases, the proportion of overlap, and the consistency of data across phases; and (c) whether "anomalies" existed within the data (Kratochwill et al., 2010). Together, evidence of a functional relation was used to determine whether each study provided *strong evidence, moderate evidence*, or *no evidence* of an effect (Kratochwill et al., 2010; Maggin et al., 2011; see Table 1).

Most of the studies were published in peer-reviewed journals; 4 were dissertations. Nine of the studies from the GBG literature reviews were included; 12 studies included in the Flower et al. (2014) meta-analysis were analyzed in the current meta-analysis. Two of the authors independently coded each study using the aforementioned rubric; discrepancies were discussed and resolved. Initial agreement percentages for the application of the four design standards and visual analysis were 88% and 91%, respectively. Final agreement was 100% for both. Agreement for article inclusion was 100%. The formula sum of agreement / total

Table 1. Overall Effect of the GBG.

	TauU (95% CI)						WWC design standards					
Study	LL	ES	UL	*n*	Number of phase contrasts	Number of cases	DS1	DS2	DS3	DS4	Vis	Overall[a]
Parrish (2012)	0.25	0.42	0.60	373	10	1	MS	MSR	MSR	MS	M	MSR
Kosiec, Czernicki, and McLaughlin (1986)	0.43	0.53	0.63	54	16	2	MS	MSR	MS	MS	M	MSR
Medland and Stachnik (1972)	0.55	0.75	0.94	28	2	1	MS	MSR	MS	NM	M	NM
Wright and McCurdy (2011)	0.61	0.89	1.00	37	4	2	MS	MSR	MSR	MS	M	MSR
Rodriguez (2010)	0.70	0.90	1.00	22	5	5	MS	MSR	MS	MS	M	MSR
Hunt (2012)	0.67	0.93	1.00	57	8	3	MS	MS	MSR	MS	S	MSR
Salend, Reynolds, and Coyle (1989)	0.84	0.97	1.00	19	14	3	MS	MS	MS	MS	S	MS
Kleinman and Saigh (2011)	0.68	0.97	1.00	26	6	1	MS	MS	MS	MS	M	MS
Saigh and Umar (1983)	0.70	0.98	1.00	20	6	1	MS	MSR	MS	NM	S	NM
Ruiz-Olivares, Pino, and Herruzo (2010)	0.59	0.98	1.00	15	4	1	MS	MS	MSR	MS	S	MSR
McCurdy, Lannie, and Barnabas (2009)	0.65	0.98	1.00	600	3	3	MS	MS	MS	MS	M	MS
Donaldson, Vollmer, Krous, Downs, and Berard (2011)[b]	0.78	0.98	1.00	98	5	5	MS	MS	MS	NM	S	NM
Patrick, Ward, and Crouch (1998)	0.76	0.99	1.00	67	6	3	MS	MSR	MSR	MS	M	MSR
Gresham and Gresham (1982)	0.62	1.00	1.00	12	4	1	MS	MS	MS	MS	S	MS
Tanol, Johnson, McComas, and Cote (2010)	0.73	1.00	1.00	6	8	2	MS	MSR	MSR	MS	M	MSR
Patterson (2003)	0.84	1.00	1.00	6	6	1	MS	MS	MS	MS	S	MS
Nolan, Filter, and Houlihan (2013)	0.70	1.00	1.00	6	6	3	MS	MSR	MS	MS	M	MSR
Crouch, Gresham, and Wright (1985)	0.68	1.00	1.00	22	6	1	MS	MSR	MSR	NM	M	NM
Bostow and Geiger (1976)	0.52	1.00	1.00	31	2	1	MS	MS	MS	NM	S	NM
Barrish, Saunders, and Wolf (1969)	0.64	1.00	1.00	24	4	1	MS	MSR	MS	NM	M	NM
Johnson, Turner, and Konarski (1978)	0.77	1.00	1.00	31	4	2	MS	MS	MS	NM	S	NM
Overall	0.78	0.82	0.87	1,580	137	43						

Note. GBG = Good Behavior Game; CI = confidence interval; LL = lower limit; ES = effect size; UL = upper limit; WWC = What Works Clearinghouse; DS = design standard; DS1 = GBG was systematically manipulated; DS2 = experimental effect across three points in time or three phase changes; DS3 = appropriate number of data points per phase (as described in the text); DS4 = IOA was least 80% and was reported for at least 20% of baseline and/or intervention phases; Vis = visual analysis; MS = meets standards; MSR = meets standards with reservations; NM = does not meet standards; M = moderate evidence; S = strong evidence; IOA = interobserver agreement.

[a]Includes DS1, DS2, DS3, and DS4. [b]In Donaldson et al. (2011), IOA data met standards for two of the five participating teachers but did not meet standards for three of the teachers.

number of agreements + disagreements × 100 (House, House, & Campbell, 1981) was used for these and all other instances of IOA.

Coding of Studies and Intercoder Reliability

The first author operationally defined and coded all 21 studies in an Excel spreadsheet. Two of the co-authors, doctoral students with training in SCR methodology and experience conducting SCR meta-analyses, were trained on the codes. Each student independently coded a set of studies using a separate Excel spreadsheet. Thus, each study was double- or triple-coded. Reliability was calculated for 75% of the studies across 15 study variables including the following: the five potential moderator variables, number of participants, IOA, and fidelity. Initial agreement was 96%. Disagreements were resolved after the first author and doctoral students reread and discussed the articles, resulting in 100% final agreement across all codes. IOA procedures were similar to those reported by Methe, Kilgus, Neiman, and Riley-Tillman (2012).

Publication Bias and Fixed-Effects Model

Publication bias was statistically tested in WinPepi (Abramson, 2011) using the Egger's test (Egger, Smith, Schneider, & Minder, 1997). The intercept for the Egger's test (2.88, 90% CI = [1.50, 4.16], p = .01) suggested publication bias. However, sensitivity analyses conducted in WinPepi indicated that no single study had an undue impact on the findings. Heterogeneity was measured using Higgins' and Thompson's H and I^2 statistics (Higgins & Thompson, 2002), where H = 2.0 (95% CI = [1.6, 2.5]) and I^2 = 75.5% (95% CI = [62.7, 84.0]). While these results indicate evidence of considerable heterogeneity, caution is warranted for two reasons. First, "The [H] test has poor power with few studies . . . it can therefore be difficult to decide whether heterogeneity is present or whether it is clinically important" (Higgins & Thompson, 2002, p. 1552). Second, regarding statistical heterogeneity, "there may be situations when the fixed-effects analysis is appropriate even when there is substantial heterogeneity of results (e.g., when the question is specifically about the particular set of studies that have already been conducted)" (Hedges & Vevea, 1998, p. 487).

Neither a fixed-effects nor a random-effects model is an "exact fit" for SCR data, but a fixed-effects model was preferable because the number of cases was too small for reasonable estimation of variances under the random-effects model (Greenhouse & Iyengar, 2009). Thus, the TauU effect size was calculated within a fixed-effects model (see Parker, Vannest, Davis, & Sauber, 2011) using WinPepi. Rather than being regarded as random samples, the studies in this meta-analysis were all regarded as estimates of an unknown "true" effect size. Variations in the "true" effect size were sought through moderator analysis (Hedges & Olkin, 1985).

Effect Size Estimation

TauU. TauU is an effect size measure based on non-overlap between A and B phases. One of its strengths is that it can control for confounding baseline trends (Parker et al., 2011). It performs reasonably well with autocorrelation (r_{auto}); its SE and significance level are not affected. When tested, 75% of TauU values remained unchanged after r_{auto} was cleansed (Parker et al., 2011). TauU is derived from the Kendall's Rank Correlation and the Mann-Whitney U test between groups. The Kendall's Rank Correlation is an analysis algorithm of time and score, comparing ordered scores and all possible pairs of data. Each pairwise comparison represents an improved score, a score that has not improved, or a tie. The Mann-Whitney U index represents differences in group level. With regard to SCR, the concept is applied to phases rather than groups, and scores from two phases are combined for a cross-group ranking. The rankings are statistically compared for mean differences. The Mann-Whitney U algorithm uses two continuous variables: scores and time. Replacing the time variable with a dummy code (0/1) to represent A and B phases yields an identical result. This, in turn, produces the proportion of pairwise comparisons that improve from Phase A to Phase B. TauU is better suited to short phases than most other methods (e.g., techniques relying on linear trends) because it can find reliable monotonic trend in only three or four data points.

Phase contrasts and effect size calculation. We used the GetData Digitizer program (version 2.25; http://www.getdata-graph-digitizer.com/) to scan and code graphed data. Graphed data from A and B phases were extracted from each study and transformed into raw numerical data by setting a scale based on the X and Y values for each phase. Effect size calculation involved several steps. First, an effect size was calculated for each AB contrast (e.g., an effect size for the A1/B1 contrast and a separate effect size for the A2/B2 contrast). Second, digitized data values were entered into the TauU calculator (Vannest, Parker, & Gonan, 2011) to obtain TauU and its standard error (SE_{Tau}). Third, TauU and SE_{Tau} values were entered into WinPepi using the meta-analysis function to aggregate the data and arrive at an effect size, standard error, and CI for each study. Fourth, TauU and SE_{Tau} values for each study were entered into WinPepi to obtain an omnibus effect size with standard error and CI. Finally, separate TauU, SE_{Tau}, and CI values were calculated for each level of each potential moderator (see Table 2).

TauU phase contrast intercoder agreement. Each of the 21 studies included multiple phase contrasts (e.g., A1/B1, A2/B2), resulting in 137 phase contrasts. The first author trained the doctoral students in calculating TauU and SE_{Tau} for each phase contrast using the TauU calculator from the obtained GetData values. Two of the authors independently

Table 2. Effect of the GBG Across Potential Moderators.

Potential moderators	Number of participants	Number of cases	Confidence interval (95%)			*z*	*p*
			LL	ES	UL		
Grade level							
Elementary	1,481	37	0.82	0.88	0.94		
Secondary	66	5	0.85	0.97	1.00	1.34	.18
GBG format[a]							
Not modified	321	13	0.74	0.81	0.88		
Modified	1,325	31	0.76	0.82	0.88	0.20	.84
Target behaviors[b]							
Off-task	1,140	39	0.76	0.81	0.86		
On-task	497	7	0.47	0.59	0.72	2.89	.01*
Reinforcer frequency							
Daily	984	30	0.77	0.82	0.88		
Daily and weekly	223	12	0.82	0.92	1.00	1.71	.08
EBD status							
EBD/EBD risk	88	11	0.89	0.98	1.00		
No EBD	1,492	34	0.70	0.76	0.81	3.77	.01*

Note. GBG = Good Behavior Game; LL = lower limit; ES = effect size; UL = upper limit; EBD = emotional and behavioral disorder.
[a]Gresham and Gresham (1982) and Kosiec, Czernicki, and McLaughlin (1986) used original and modified versions of the GBG. The number of participants for these studies are represented in both formats. [b]Hunt (2012) and Patrick, Ward, and Crouch (1998) reported both on- and off-task behavior data. The number of participants from these studies are represented in both types of behavior.
*p = .05

calculated these values for 20% (n = 27) of the 137 AB phase contrasts across all studies. Initial agreement for non-overlap between A and B phases ranged from 50% to 100%. The 50% agreement reflects difficulty extracting the data from the Johnson, Turner, and Konarski (1978) study. Disagreements were resolved after the authors discussed the discrepancies and recoded the data, resulting in 100% final agreement. We then calculated TauU and SE_{Tau} for the remaining phase contrasts in each study.

Statistical significance. We determined statistical significance for TauU values using 95% CI (α = .05). A 90% to 95% CI is standard for determining whether change is reliable (Nunnally & Bernstein, 1994), indicating a reasonable chance of 5% to 10% likelihood of error. Statistical significance between TauU values was determined by calculating 83.4% CI to visually test for overlap of upper and lower limits between effect sizes. Visual comparison of two effect sizes with 83.4% CI is the same as a p = .05 or a 95% confidence level test between the two scores (Payton, Greenstone, & Schenker, 2003).

Potential Moderators

We examined five potential moderators, variables hypothesized to affect students' behaviors. They were selected because they were the recommended areas of future research or had not yet been addressed in the previous reviews or meta-analysis. Potential moderators were as follows: EBD risk status, reinforcement frequency, target behaviors, GBG format, and grade level.

We calculated a reliable difference for the levels of each potential moderator. If statistically significant differences were obtained between levels, the potential moderator was confirmed as a moderator because it differentially affected students' outcomes. The reliable difference formula, (L1 − L2) / sqrt [(SE_{Tau}1sqrd) + (SE_{Tau}2sqrd)], based on the t test, was used to determine whether levels of a given moderator differed statistically from one another. A reliable difference is one that is so large that it cannot be accounted for solely by chance, given the number of participants and data points. Alpha was set at .05 and the confidence level was set at 95% to determine whether the findings were credible (viz., whether they would change substantially over several retestings). Reliable difference z test scores and p values are reported.

EBD risk status. The codes used for EBD risk status were *EBD/EBD risk* and *no EBD/no EBD risk. EBD/EBD risk* referred to students identified in a given study as having an emotional and/or behavioral disorder, or those at risk for being identified as having an emotional and/or behavioral disorder. Data for students not identified as individuals with or at risk for EBD were coded *no EBD/no EBD risk*.

Frequency of reinforcement. In some studies, daily reinforcement was awarded to the winning team(s) at the end of the class period in which the Game was played. In other studies, it was awarded at the end of the school day if it was implemented in multiple classes. Both daily and weekly reinforcers were awarded to the winning team(s) that met criteria as an extra incentive in some studies. Levels were *daily* and *daily and weekly*.

Target behaviors. Target behaviors consisted of two categories: *disruptive/off-task* and *attention to task/on-task*. Disruptive/off-task behaviors included a range of behaviors including the following: being out-of-seat, talking without permission, interrupting, fighting, name-calling, cursing, pushing, hitting, and destroying property. Attention to task/on-task behaviors referred to complying with teacher directions, working quietly, raising one's hand before asking a question, and getting instructional materials without talking.

GBG format. GBG format referred to the use of the GBG as originally described by Barrish et al. (1969) or a modification thereof. Levels were *modified* and *not modified*.

Grade level. Grade level was represented by two levels: *elementary* (pre-kindergarten through Grade 5) and *secondary* (Grades 6 through 12).

Results

Study Characteristics

The 21 SCR studies examined in this meta-analysis consisted of 43 cases and 137 phase contrasts representing 1,580 participants in pre-kindergarten through Grade 12. The majority of the studies focused on elementary school students ($n = 17$), 2 studies targeted secondary students, and 2 studies included elementary and secondary students. Participant gender was reported in 8 studies (341 males, 300 females). Participant ethnicity was reported in 4 studies, representing more than 13 ethnic groups, including African American, Caucasian, Hispanic, Asian, Native American, Mestizo, Creole, and Mayan. Although most of the studies took place in the United States, they were also carried out in Spain (Ruiz-Olivares et al., 2010), Sudan (Saigh & Umar, 1983), British Columbia (Kosiec, Czernicki, & McLaughlin, 1986), and Belize (Nolan, Filter, & Houlihan, 2013). One study (Tanol, Johnson, McComas, & Cote, 2010) conducted in a U.S. school emphasized Native American culture. Participants included students with intellectual disabilities (Gresham & Gresham, 1982), developmental disabilities (Patterson, 2003), EBD (Salend et al., 1989), students at risk for EBD (Tanol et al., 2010), and students without disabilities (Nolan et al., 2013).

Studies were conducted in general education classrooms ($n = 14$), special education classrooms ($n = 2$), school cafeterias ($n = 2$), a Head Start classroom ($n = 1$), a physical education class ($n = 1$), and a classroom for students experiencing behavioral challenges ($n = 1$). Behavioral rules and expectations were clearly defined in all studies. Content areas and activities during which the GBG was implemented included math, reading, science, social studies, language arts, art, and circle time. Eleven studies implemented a modified GBG format; 8 used the original format. Two studies (Gresham & Gresham, 1982; Kosiec et al., 1986) compared the use of the original format with a modified version. Data for original and modified formats from each of the 2 studies were analyzed separately with separate TauU and SE_{Tau} values. All studies reported direct observations of student behaviors. Twenty studies reported IOA (range = 80%–100%). The remaining study (Bostow & Geiger, 1976) reported collecting IOA data, but did not provide them. Only 14 of the 21 studies reported the percentage of observations included in calculating IOA (range = 25% –52% across baseline and/or intervention phases). Nine studies reported fidelity of implementation; average fidelity was 88%. Social validity was reported for teachers and/or students in 13 studies; ratings and interviews indicated high social validity.

Overall Effect

In response to the first research question, the overall effect of the GBG was examined across the 21 studies, yielding a mean TauU effect size of .82 ($SE = .02$, 95% CI = [.78, .87], $p = .05$). To help with effect size interpretation, we transformed the obtained TauU value to a Cohen's d of 1.99 using the formula $d = 3.464 \times [1 - \text{sqrt}(1 - \text{TauU})]$ (Rosenthal, 1994). This would be considered a large effect size based on the commonly accepted values proposed by Cohen (1992). Table 1 presents the range of effect sizes with CIs across studies at a 95% confidence level. Thus, there is a 95% certainty that the true value for the obtained effect size fell between the upper and lower limits of the calculated CI.

Potential Moderators

We calculated levels of potential moderators using the reliable difference formula. A 95% CI ($\alpha = .05$) was set for each effect size; p values are reported for z test values. The results addressed the second research question (see Table 2).

EBD risk status. Students with or at risk for EBD yielded a larger effect size (ES = .98, $SE = .05$, 95% CI = [0.89, 1.00], $p = .05$) than students not identified with or not at risk for EBD (ES = .76, $SE = .03$, 95% CI = [0.70, 0.81], $p = .05$). Reliable difference values were $z = 3.77$, $p = .01$.

Reinforcement frequency. The use of both daily and weekly reinforcement resulted in a larger effect size (ES = .92, *SE*

= .05, 95% CI = [0.82, 1.00], p = .05) than did the use of daily reinforcement alone for winning teams (ES = .82, SE = .03, 95% CI = [0.77, 0.88], p = .05). Reliable difference results for this variable were z = 1.71, p = .08.

Target behaviors. A larger effect size was obtained for disruptive/off-task behaviors (ES = .81, SE = .03, 95% CI = [0.76, 0.86], p = .05) than for attention to task/on-task behaviors (ES = .59, SE = .07, 95% CI = [0.47, 0.72], p = .05). The reliable difference values were z = 2.89, p = .01.

GBG format. Interventions using a modified format had a slightly larger effect size (ES = .82, SE = .03, 95% CI = [0.76, 0.88], p = .05) than GBG interventions using the original format (ES = .81, SE = .04, 95% CI = [0.74, 0.88], p = .05). Reliable difference values were z = .20, p = .84.

Grade level. A larger effect size was observed for secondary students (ES = .97, SE = .06, 95% CI = [0.85, 1.00], p = .05) than for elementary students (ES = .88, SE = .03, 95% CI = [0.82, 0.94], p = .05). Kosiec et al. (1986) included both elementary and secondary participants. However, their study was not included in the grade-level analysis because the data were not disaggregated for elementary and secondary students. The reliable difference z test value for grade level was 1.34, p = .18.

Discussion

The purpose of this meta-analysis was to examine the effect of the GBG and five potential moderators on elementary and secondary students' behaviors across 21 SCR studies. Specifically, this is the first meta-analysis of the GBG to examine (a) disability (viz., EBD), (b) GBG format, and (c) reinforcement frequency as potential moderator variables. There were several findings worth noting. First, the large overall effect (ES = .82) indicated that a reduction in problem behaviors and an increase in desirable behaviors may be attributed to the GBG. Second, moderator analyses revealed a statistically significant difference for two variables: EBD risk status and target behaviors. That is, students with or at risk for EBD benefited more from the GBG than their peers without EBD. Also, students who exhibited disruptive and off-task behaviors benefited most from the Game. Although there were more participants in the disruptive/off-task behaviors group (see Table 2), TauU weighs the number of observations in each study by the inverse of the variance. As such, findings were not influenced by moderator group ns. Third, findings also revealed that the GBG was more effective in reducing disruptive/off-task behaviors than increasing attention to task/on-task behaviors.

Although the remaining three potential moderators were not statistically significant, our results were similar to the findings reported by Flower et al. (2014) for grade level. We found moderate to large effects of the GBG in reducing problem behaviors for elementary and secondary students. We also discovered that, consistent with conclusions reported by Tankersley (1995) and Tingstrom et al. (2006), the GBG implemented in both its original and modified formats was effective. This offers teachers some flexibility in tailoring the Game to their students' behavioral needs. For example, the Game can be modified to award points for appropriate behaviors rather than deduct points for inappropriate behaviors (Crouch et al., 1985), use three teams versus two (Hunt, 2012), implement the intervention in a non-classroom setting (e.g., the cafeteria; McCurdy, Lannie, & Barnabas, 2009), or include an additional behavioral intervention (Ruiz-Olivares et al., 2010). However, we caution that although there seems to be some flexibility in its implementation, the core features of the Game should be adhered to as an interdependent group contingency to achieve outcomes such as those noted in the literature.

With regard to reinforcement frequency, there was a greater reduction in problem behaviors with more frequent reinforcement. This finding is particularly noteworthy for students with or at risk for EBD (Cheyney & Jewell, 2012). Last, like Flower et al., we noted that very few studies reported fidelity of implementation. It is important to report these data to help understand the degree to which and the consistency with which the GBG is implemented. Such data could inform revisions or improve implementation of the intervention.

Limitations

The following limitations should be considered in interpreting the findings of this meta-analysis. First, there are no generally agreed upon standards within SCR on the use of meta-analysis within the field to determine evidence-based practices (Horner & Kratochwill, 2012). Although the use of meta-analysis in other fields to synthesize research literature is becoming a standard practice (Parker & Hagan-Burke, 2007), its application in applied behavior analysis and SCR is more limited. Second, we did not use IOA as an inclusion criterion because so few studies reported the information needed to apply this proposed design standard. As new criteria are developed and existing protocols are revised, studies may be judged differently by others in the field. Third, effects on individual students could not be determined in some studies. For example, Barrish et al. (1969) reported that the two most behaviorally challenging students received marks for inappropriate behaviors individually rather than deducting points from their team for their disruptive behaviors and refusal to participate in the Game. Thus, the impact of the Game on all students is not reflected in the data reported in this study. Fourth, most studies did not disaggregate behaviors by type (e.g., aggressive). Rather, a range of problem behaviors were operationally defined and combined in a "disruptive behavior" category. As such, results could not be provided for some categories of problem behaviors. Finally, as a newer effect size measure, TauU has been used in few meta-analyses (e.g., Bowman-Perrott et al., 2013) and intervention studies (e.g., Rispoli et al., 2013). Also, caution should be used in comparing TauU with Cohen's *d*, as the transformation is an approximation.

Implications for Research and Practice

As additional SCR studies are conducted, complete IOA data (viz., the percentage of observations included in IOA calculations) should be reported. In addition, future research should investigate the potential mediating effect of several variables. One is gender, as boys have been found to be more likely to display disruptive behaviors than girls (McIntosh, Reinke, Kelm, & Sadler, 2013). A second variable is ethnicity, as it is related to other types of behavioral outcomes (e.g., suspension and expulsion from school; Bowman-Perrott et al., 2011). Intervention length and duration may also make a difference in students' outcomes (Kosiec et al., 1986). In future GBG studies, behavioral outcome data need to be disaggregated by type, as several studies combined physically and verbally aggressive behaviors with out-of-seat and talking out behaviors. In addition, the impact of the GBG on students' academic achievement (Flower et al., 2014; Kellam et al., 1994) should be examined in consideration of the relation between academic difficulties and behavior problems (Lane, Barton-Arwood, Nelson, & Wehby, 2008). Finally, the use of response cost versus reinforcement in promoting positive behaviors during the GBG (Tanol et al., 2010; Wright & McCurdy, 2011) should be further examined.

Identifying empirically supported practices is important in an era of increased accountability. It is critical that school personnel identify and implement classroom management and behavioral interventions that promote prosocial behaviors. In light of the evidence pointing to the GBG as an effective universal behavior management strategy, it can be a benefit to general and special education teachers. Overall, the GBG is an effective, positive behavioral support that can easily be incorporated into elementary and secondary school-based settings.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

*References with an asterisk indicate articles included in the meta-analysis.

Abramson, J. H. (2011). WINPEPI updated: Computer programs for epidemiologists and their teaching potential. *Epidemiologic Perspectives & Innovations*, *8*(1), 1–9.

Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behaviour Research and Therapy*, *31*, 621–631.

Babyak, A. E., Luze, G. J., & Kamps, D. M. (2000). The Good Behavior Game: Behavior management for diverse classrooms. *Intervention in School and Clinic*, *35*, 216–233.

*Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good Behavior Game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, *2*, 119–124.

*Bostow, D., & Geiger, O. G. (1976). Good Behavior Game: A replication and systematic analysis with a second grade class. *School Applications of Learning Theory*, *8*(2), 18–27.

Bowman-Perrott, L., Benz, M. R., Hsu, H., Kwok, O., Eisterhold, L., & Zhang, D. (2011). Patterns and predictors of disciplinary exclusion over time: An analysis of the SEELS national dataset. *Journal of Emotional and Behavioral Disorders*, *21*(2), 83–96.

Bowman-Perrott, L., Davis, H., Vannest, K. J., Williams, L., Greenwood, C. R., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review*, *42*(1), 39–55.

Cheyney, D., & Jewell, K. (2012). Positive behavior supports and students with emotional and behavioral disorders. In J. P. Bakken, F. E. Obiakor, & A. F. Rotatori (Eds.), *Behavioral disorders: Practice concerns and students with EBD* (Vol. 23; pp. 83–106). Bingley, UK: Emerald.

Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology*, *112*, 155–159.

Cooper, H. (2011). *Reporting research in psychology: How to meet journal article reporting standards* (6th ed.). Baltimore, MD: American Psychological Association.

*Crouch, P. L., Gresham, F. M., & Wright, W. R. (1985). Interdependent and independent group contingencies with immediate and delayed reinforcement for controlling classroom behavior. *Journal of School Psychology*, *23*, 177–187.

Darveaux, D. X. (1984). The Good Behavior Game plus merit: Controlling disruptive behavior and improving student motivation. *School Psychology Review*, *13*, 510–514.

Dolan, L. J., Kellam, S. G., Brown, C. H., Werthamer-Larsson, L., Rebok, G. W., . . . Wheeler, L. (1993). The short-term impact of two classroom-based preventative interventions on aggressive and shy behaviors and poor achievement. *Journal of Applied Developmental Psychology*, *14*, 317–345.

*Donaldson, J. M., Vollmer, T. R., Krous, T., Downs, S., & Berard, K. P. (2011). An evaluation of the Good Behavior Game in kindergarten classrooms. *Journal of Applied Behavior Analysis*, *44*, 605–609.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629–634.

Embry, D. D. (2002). The Good Behavior Game: A best practice candidate as a universal behavioral vaccine. *Clinical Child and Family Psychology Review*, *5*, 273–297.

Fishbein, J. E., & Wasik, B. H. (1981). Effect of the Good Behavior Game on disruptive library behavior. *Journal of Applied Behavior Analysis*, *1*(14), 89–93.

Flower, A., McKenna, J. W., Bunuan, R. L., Muething, C. S., & Vega, R. (2014). Effects of the Good Behavior Game on challenging behaviors in school settings. *Review of Educational Research. Review of Educational Research*, *84*, 546–571. doi:10.3102/0034654314536781

Greenhouse, J. B., & Iyengar, S. (2009). Sensitivity analysis and diagnostics. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 417–433). New York, NY: Russell Sage.

*Gresham, F. M., & Gresham, G. N. (1982). Interdependent, dependent, and independent group contingencies for controlling disruptive behavior. *The Journal of Special Education*, *16*, 101–110.

Harris, V. W., & Sherman, J. A. (1973). Use and analysis of the "Good Behavior Game" to reduce disruptive classroom behavior. *Journal of Applied Behavior Analysis*, *6*, 405–417.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*, 486–504.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539–1558.

Horner, R. H., & Kratochwill, T. R. (2012). Synthesizing single-case research to identify evidence-based practices: Some brief reflections. *Journal of Behavioral Education*, *21*, 266–272.

House, A. E., House, B. J., & Campbell, M. B. (1981). Measures of interobserver agreement: Calculation formulas and distribution effects. *Journal of Behavioral Assessment*, *3*(1), 37–57.

Huber, H. (1979). The value of a behavior modification program administered in the fourth grade of a school for special education. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, *28*(2), 73–79.

Huizink, A. C., van Lier, P. A. C., & Crijnen, A. A. M. (2008). Attention deficit hyperactivity disorder symptoms mediate early-onset smoking. *European Addiction Research*, *15*, 1–9.

*Hunt, B. M. (2012). Using the Good Behavior Game to decrease disruptive behavior while increasing academic engagement with a Head Start population (Unpublished doctoral dissertation). University of Southern Mississippi, Hattiesburg.

*Johnson, M. R., Turner, P. F., & Konarski, E. A. (1978). The "Good Behavior Game": A systematic replication in two unruly transitional classrooms. *Education and Treatment of Children*, *1*(3), 25–33.

Kellam, S. G., Rebok, G. W., Ialongo, N., & Mayer, L. S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trail. *Journal of Child Psychology and Psychiatry*, *35*, 259–281.

*Kleinman, K. E., & Saigh, P. A. (2011). The effects of the Good Behavior Game on the conduct of regular education New York city high school students. *Behavior Modification*, *35*, 95–105.

*Kosiec, L. E., Czernicki, M. R., & McLaughlin, T. F. (1986). The Good Behavior Game: A replication with consumer satisfaction in two regular elementary school classrooms. *Techniques*, *2*(1), 15–23.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Lane, K. L., Barton-Arwood, S., Nelson, R., & Wehby, J. (2008). Academic performance of students with emotional and behavioral disorders in a self-contained setting. *Journal of Behavioral Education*, *17*, 43–62.

Leflot, G., van Lier, P. A. C., Onghena, P., & Colpin, H. (2010). The role of teacher behavior management in the development of disruptive behaviors: An intervention study with the Good Behavior Game. *Journal of Abnormal Child Psychology*, *38*, 869–882.

Maggin, D. M., Chafouleas, S. M., Goddard, K. M., & Johnson, A. J. (2011). A systematic evaluation of token economies as a classroom management tool for students with challenging behavior. *Journal of School Psychology*, *49*, 529–544.

*McCurdy, B. L., Lannie, A. L., & Barnabas, E. (2009). Reducing disruptive behavior in an urban school cafeteria: An extension of the Good Behavior Game. *Journal of School Psychology*, *47*(1), 39–54.

McGoey, K. E., Schneider, D. L., Rezzetano, K. M., Prodan, T., & Tankersley, M. (2010). Classwide intervention to manage disruptive behavior in the kindergarten classroom. *Journal of Applied School Psychology*, *26*, 247–261.

McIntosh, K., Reinke, W. M., Kelm, J. L., & Sadler, C. A. (2013). Gender differences in reading skill and problem behavior in elementary school. *Journal of Positive Behavior Interventions*, *15*, 51–60.

*Medland, M. B., & Stachnik, T. J. (1972). Good-Behavior Game: A replication and systematic analysis. *Journal of Applied Behavior Analysis*, *5*(1), 45–51.

Methe, S. A., Kilgus, S. P., Neiman, C., & Riley-Tillman, T. C. (2012). Meta-analysis of interventions for basic mathematics computation in single-case research. *Journal of Behavioral Education*, *21*, 230–253.

*Nolan, J. D., Filter, K. J., & Houlihan, D. (2013). Preliminary report: An application of the Good Behavior Game in the developing nation of Belize. *School Psychology International*, *35*, 421–428. doi:10.1177/014303498598

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw Hill.

Parker, R., & Hagan-Burke, S. (2007). Useful effect sizes interpretations for single case research. *Behavioral Therapy*, *38*, 95–105.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining non-overlap and trend for single-case research: Tau-U. *Behavior Therapy*, *35*, 202–322.

*Parrish, R. (2012). Examining changes in appropriate social behaviors during school lunch using the lunchtime behavior game (Unpublished doctoral dissertation). Northeastern University, Boston, MA.

*Patrick, C. A., Ward, P., & Crouch, D. W. (1998). Effects of holding students accountable for social behaviors during volleyball games in elementary physical education. *Journal of Teaching in Physical Education*, *17*, 143–156.

*Patterson, K. B. (2003). The effects of a group-oriented contingency—the Good Behavior Game—on the disruptive behavior of children with developmental disabilities (Unpublished doctoral dissertation). Kent State University, Kent, OH.

Payton, M. E., Greenstone, M. H., & Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *Journal of Insect Science*, *3*(4), 1–6.

Reid, R., Gonzalez, J. E., Nordess, P. D., Trout, A., & Epstein, M. (2004). A meta-analysis of the academic status of students with emotional/behavioral disturbance. *The Journal of Special Education*, *38*, 130–143.

Rispoli, M., Lang, R., Neely, L., Camargo, S., Hutchins, N., Davenport, K., & Goodwyn, F. (2013). A comparison of within- and across-activity choices for reducing challenging behavior in children with autism spectrum disorders. *Journal of Behavioral Education*, *22*, 66–83.

*Rodriguez, B. J. (2010). *An evaluation of the Good Behavior Game in early reading intervention groups* (Unpublished doctoral dissertation). University of Oregon, Eugene.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 232–243). New York, NY: Russell Sage.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, *52*, 59–82.

*Ruiz-Olivares, R., Pino, M. J., & Herruzo, J. (2010). Reduction of disruptive behaviors using an intervention based on the good behavior game and the say-do-report correspondence. *Psychology in the Schools*, *47*, 1046–1058.

*Saigh, P. A., & Umar, A. M. (1983). The effects of a Good Behavior Game on the disruptive behavior of Sudanese elementary school students. *Journal of Applied Behavior Analysis*, *16*, 339–344.

*Salend, S. J., Reynolds, C. J., & Coyle, E. M. (1989). Individualizing the Good Behavior Game across type and frequency of behavior with emotionally disturbed adolescents. *Behavior Modification*, *13*, 108–126.

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention*, *2*(3), 188–196.

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, *31*, 351–380.

Swiezy, N. B., Matson, J. L., & Box, P. (1992). The Good Behavior Game: A token reinforcement system for preschoolers. *Child & Family Behavior Therapy*, *14*(3), 21–32.

Tankersley, M. (1995). A group-oriented contingency management program: A review of research on the Good Behavior Game and implications for teachers. *Preventing School Failure*, *40*, 19–24.

*Tanol, G., Johnson, L., McComas, J., & Cote, E. (2010). Responding to rule violations or rule following: A comparison of two versions of the Good Behavior Game with kindergarten students. *Journal of School Psychology*, *48*, 337–355.

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools*, *44*, 423–432.

Tingstrom, D. H., Sterling-Turner, H. E., & Wilczynski, S. M. (2006). The Good Behavior Game: 1969-2002. *Behavior Modification*, *30*, 225–253.

Vannest, K. J., Parker, R. I., & Gonan, O. (2011). Single-case research: Web-based calculators for SCR analysis, Version 1.0 [Web-based application]. College Station: Texas A&M University. Available from www.singlecaseresearch.org

*Wright, R. A., & McCurdy, B. L. (2011). Class-wide positive behavior support and group contingencies: Examining a positive variation of the good behavior game. *Journal of Positive Behavior Interventions*, *14*, 173–180.

Review of Educational Research
December 2014, Vol. 84, No. 4, pp. 546–571
DOI: 10.3102/0034654314536781

Effects of the Good Behavior Game on Challenging Behaviors in School Settings

Andrea Flower
The University of Texas at Austin

John W. McKenna
St. John's University

Rommel L. Bunuan
Colin S. Muething
Ramon Vega Jr.
The University of Texas at Austin

Challenging behavior at school remains a concern for teachers and administrators. Thus classroom management practices to prevent challenging behavior are sorely needed. The Good Behavior Game (GBG) has been found to be useful to positively change student behavior. However, previous reviews of the GBG have not quantified effects, have not focused solely on school and classroom behaviors, and have not examined study features that facilitate greater outcomes. Twenty-two peer-reviewed journal articles were reviewed. Study data were analyzed using effect sizes, percent of nonoverlapping data, percent of all nonoverlapping data, and hierarchical linear modeling to determine intervention effectiveness as well as study features that facilitated greater outcomes. Findings suggested that (a) moderate to large effects were found on challenging behaviors and these effects were immediate; (b) the GBG was most commonly used for disruptive behavior, off-task behavior, aggression, talking out, and out-of-seat behaviors; (c) the GBG has been implemented primarily in general education elementary school settings; and (d) correct application of reward procedures are important for intervention effectiveness. Study limitations, implications for practice, and areas for future research are presented.

Keywords: review, Good Behavior Game, challenging behavior, classroom management, hierarchical linear modeling, effect size

Today, in the context of high-stakes assessments, school reform, and improved academic achievement, teachers are under additional pressure to maximize instructional time to promote academic development (Vannest, Temple-Harvey, & Mason, 2009). Yet, in today's schools, students are not always academically or socially prepared for school (Blair & Diamond, 2008). In fact, challenging behavior often interrupts teachers' abilities to teach and students' abilities to learn (Cameron, Connor, Morrison, & Jewkes, 2008). With the move toward inclusion and legislation requiring student placement in the least restrictive environment (Individuals with Disabilities Education Improvement Act, 2004), students with challenging behavior might be served in a variety of settings including general education or special education classrooms. This change means that all teachers should be prepared to provide academic instruction as well as behavioral support (Bohannon & Wu, 2011; Witt, VanDerHeyden, & Gilbertson, 2004).

Challenging behavior at school may manifest under many conditions and in various locations throughout a school. A variety of behaviors have been identified by researchers as challenging at school including physical and verbal aggression, harassment, fighting, disrespect, and defiance (Kaufman et al., 2010; Spaulding et al., 2010), getting out of one's seat, talking without permission, and classroom rule violations (Walter, Gouze, & Lim, 2006). Harrison, Vannest, Davis, and Reynolds (2012) found that teachers identified general distractibility and difficulty following directions as the most challenging behaviors. Considering the range of problem behaviors and the time spent handling them, strategies are needed to prevent the occurrence of challenging behavior and promote the display of appropriate behaviors.

The Good Behavior Game

The Good Behavior Game (GBG) is a classroom management strategy that has been used and studied for more than 40 years. The GBG is an easy to implement group contingency procedure that includes identifying target behaviors, posting rules, identifying rewards, dividing a class into at least two equal teams, identifying rule violators and stating their infractions, debiting the offending team for infractions or awarding points for meeting expectations, and awarding daily and weekly prizes to the team with the fewest infractions (Barrish, Saunders, & Wolf, 1969; Elswick & Casey, 2012) or most points earned for prosocial behavior. The GBG allows teachers to engage in several behavior management strategies including acknowledging appropriate behavior, teaching classroom rules, providing feedback about inappropriate behavior, engaging in response cost practices, verbal praise, and providing rewards as reinforcement. Thus, the GBG is a potentially effective classroom management tool for teacher use (Elswick & Casey, 2012).

In its initial empirical evaluation (Barrish et al., 1969), researchers used the GBG to decrease out-of-seat and talking-out behaviors of fourth-grade students during mathematics and reading instruction. Since that initial investigation, the GBG has been applied numerous times to test its effects on a variety of behaviors. The GBG has been implemented by various intervention agents with varying levels of GBG training for different lengths of time, in a variety of settings, and with and without using rewards (cf., Darch & Thorpe, 1977; Elswick & Casey, 2012; McCurdy, Lannie, & Barnabas, 2009; Tanol, Johnson, & McComas, 2010). These factors might make a difference in outcomes as well.

Two earlier reviews (Embry, 2002; Tingstrom, Sterling-Turner, & Wilczynski, 2006) highlighted studies of the GBG, with both reviews concluding that the GBG promotes behavior change. Embry (2002) provided descriptive information on previously published studies as well as key findings such as intervention effectiveness in terms of immediate behavior change and long-term impact, social validity, and potential use as a prevention strategy. Tingstrom et al. (2006) also wrote a descriptive review in which they discussed the GBG variations used in experimental studies as well as the student populations included in such studies. They also suggested areas for future research such as investigating methods for limiting the amount of peer pressure placed on students who violate game rules and its potential effect on intervention effectiveness.

Neither Embry (2002) nor Tingstrom et al. (2006) specifically focused on observable and measurable challenging behaviors in school or classroom settings. A focus on such behaviors in schools and classrooms is important given the amount of time teachers and administrators spend handling challenging behavior (U.S. Department of Education, 2000), the amount of instructional time lost to these behaviors, and the amount of stress that teachers feel due to challenging behavior (Klassen & Chiu, 2010; Nelson, Maculan, Roberts, & Ohlund, 2001). Additionally, neither Embry nor Tingstrom et al. quantified the effect of the GBG. In our review, we attempt to quantify the effect of the GBG, particularly on observable and measurable challenging behaviors in school or classroom settings. Some of the studies included in our review use group design and others utilize single subject experimental designs (SSEDs).

For group designs, Cohen's *d* is a widely accepted metric for quantifying magnitude of effect. Unfortunately, a method for quantifying effects across SSEDs in meta-analyses is less clear. One promising approach for meta-analysis of SSEDs includes the use of hierarchical linear modeling (HLM; Nagler, Rindskopf, & Shadish, 2008). The HLM approach allows for the aggregation of multiple studies at the case level, which increases overall sample size. Additionally, HLM corrects for autocorrelation and allows for analysis of data from multiple cases, even when the numbers of observations vary across cases (Raudenbush & Bryk, 2002). Using HLM we can study the treatment effect using all of the data from each case.

In a resource-limited system such as a school, knowledge of the expected effects is very desirable before investing the time and resources in any intervention. School professionals might also benefit from information concerning differences in effect of the GBG based on variations in fidelity, intervention agent, duration, setting, and use of rewards. Because HLM analysis does not provide an individual effect size for each study, nonparametric statistics such as percentage of nonoverlapping data (PND) and percentage of all nonoverlapping data (PAND) were useful statistics as we examined how various study characteristics may have affected GBG outcomes.

Purpose and Research Questions

With the importance of student achievement in school, it is essential that teachers manage and change challenging classroom and school behavior so that more time can be allocated to academic instruction. Previous reviews have not quantified the effects of the GBG, nor have they examined specific characteristics of GBG interventions (i.e., fidelity, intervention agent, setting, duration, and/or use of rewards) that may affect outcomes (Embry, 2002; Tingstrom et al., 2006). Additionally, it has been 7 years since researchers have reviewed this literature. In our review, we attempted to identify additional and more recent experimental studies that examined the effect of the GBG on classroom and school challenging behaviors. The purposes of this review were to (a) describe and quantify the effect of the GBG on various challenging behaviors in school and classroom settings and (b) understand characteristics of the intervention that may affect the magnitude of the outcomes. The following questions guided this review:

Research Question 1: What is the effect of the GBG on the level and trend of challenging behaviors in school and classroom settings?

Research Question 2: Do variations in fidelity, intervention agent, duration, setting, and use of rewards affect GBG outcomes?

Method

We searched the literature to identify studies of the GBG implemented with students in kindergarten through 12th grade between 1970 and the present. Barrish et al. (1969) developed the GBG in 1969; thus, years prior to 1970 were not searched. First, we conducted an electronic search using EBSCO Research Databases including Academic Search Complete, ERIC, and PsycInfo. Next, in an effort to capture all articles concerning the GBG, the reference lists of all GBG articles including the two previous reviews (Embry, 2002; Tingstrom et al., 2006) were searched. In addition, we manually searched several journals that appeared to frequently publish GBG articles in the event that the electronic search failed to identify all relevant articles. The journals included in the manual search were *Behavior Modification*, *Education & Treatment of Children*, *Journal of Applied Behavior Analysis*, *Journal of School Psychology*, and *Psychology in Schools*.

Our initial search yielded 51 articles, including the original empirical investigation of the GBG, which potentially met our inclusion criteria. We then applied a set of inclusion criteria to identify our final pool of articles. Articles selected for inclusion met the following criteria:

1. Published in a peer-reviewed journal in education, special education, behavioral analysis, psychology, or school psychology between 1970 and spring 2013;
2. Article written in English;
3. Referred to the independent variable as the GBG and used GBG procedures;
4. The research design was either an experimental/quasi-experimental design or SSED with replication (multiple baseline or reversal);
5. Dependent variables were challenging behaviors that were a threat to learning, safety, and relationships that were observable and measurable;
6. Results of the GBG implementation could be disaggregated; and
7. Data were available for extraction or calculation of effect size.

Article Coding

Two researchers read and independently double-coded each article with regard to design, dependent variables of the study, outcomes, fidelity, interventionist, interventionist training, duration, setting, and reward use. Design referred to the research design used in the study. Dependent variables (DVs) were the specific outcome variables that concerned challenging behaviors that were a threat to learning, safety, and relationships, which were observable and measurable in the classroom or school setting. We reviewed the DVs and operational definitions given in each article. Operational definitions were closely matched across studies. For example, talking out might have been referred to as either talking out or inappropriate verbalizations (cf., Barrish et al., 1969; Salend, Reynolds, & Coyle, 1989). Operationally, both behaviors were defined as verbalizations without teacher permission. DVs defined so similarly were categorized as the same DV. Outcomes referred to the results for each DV.

Fidelity referred to the treatment integrity of the intervention, essentially whether GBG took place as intended. GBG interventionist referred to the personnel responsible for implementing the GBG with children. Training for GBG interventionists concerned the amount of training and type of training received prior to and during GBG implementation. Training for the interventionist was coded as lecture, lecture/feedback, modeling, or no response. Lecture was defined as receipt of information about the GBG without additional exposure. Lecture/feedback was defined as receipt of information with feedback from the experimenter after attempting implementation. Modeling was defined as the experimenter demonstrating how to use the GBG in the interventionist's setting.

Duration referred to the number of days for which a study was conducted. When an article reported another duration metric such as number of weeks or months, conversions were made to days in order to standardize the duration code (1 week = 5 school days). Conversions were considered to be estimates, as the authors did not specifically provide them. Four studies (Dion et al., 2011; Leflot, van Lier, Onghena, & Colpin, 2010; Leflot, van Lier, Onghena, & Colpin, 2013; Ruiz-Olivares, Pino, & Herruzo, 2010) required such estimates. Setting pertained to the level of school (elementary or secondary), grade level, type of school or class (e.g., general education, special education, traditional school campus, alternative learning center, school within a residential facility), and location of GBG implementation in the school (e.g., classroom, cafeteria). Finally, reward use concerned use of points (and/or fouls) with teams of students, use of rewards that could be exchanged for points, the type of backup rewards used, frequency of rewards, and use of preference assessments for rewards.

Codes for each component were reviewed and compared for similarity across coders. Calculation of overall and point-by-point reliability relied on this formula (number of agreements divided by agreements plus disagreements, multiplied by 100) to arrive at a percent reliability (Kazdin, 2011). After initial, independent coding, the mean agreement level was 97.2% and point-by-point agreement ranged from 94.3% to 100%. When disagreements occurred, both coders returned to the original article and recoded the source of the disagreement. The coders discussed disagreements and arrived at 100% agreement on each code. Final overall reliability and point-by-point reliability were 100%.

Data Extraction

Two members of the research team extracted data from each graph associated with the 16 SSED studies. The team used GraphClick, a data extraction program, to extract data from the graphs. Recent research (Boyle, Samaha, Rodewald, & Hoffman, 2012) has validated GraphClick as a method that yields reliable and valid data. From the 16 articles, using GraphClick, we extracted 1,439 data points that corresponded to baseline (582) and GBG (857) intervention conditions.

Use of HLM requires a single dimension across the DVs. For these 16 SSEDs, five studies used frequency as a DV, nine relied on percentage, and two used rate as dimensions for the associated analyses. Analysis with HLM required scaling of the DVs to percentages. Frequency and rate data were converted to percentages by dividing the score represented in each datapoint by the total number possible. Data were also recoded to standardize the direction of intervention effect as some studies aimed to decrease inappropriate behaviors and others aimed to increase appropriate behaviors. Reverse coding of data allowed us to reflect decreases in inappropriate behaviors across all studies. We determined the recoded value by subtracting the old value from the sum of the scale minimum and scale maximum.

Data Analysis

Data Analysis for Research Question 1

Effect size calculations allowed for a determination of the effect of the GBG on challenging school and classroom behavior. We either (a) used Cohen's *d* as provided in the article or (b) calculated the effect size for each group study. Cohen's *d* was calculated as the difference between the mean posttest score of the treatment group minus the mean posttest score of the control group divided by the pooled standard deviation (Cooper, Hedges, & Valentine, 2009). Effect sizes explain the degree to which outcomes for GBG participants differed compared with control group participants. Effect sizes were interpreted using the following criteria: $d = .80$ or greater (large), $d = .50$ (moderate), and $d = .20$ (modest; Cohen, 1988). Averaging across the individual effect sizes provided an overall measure of magnitude.

For SSEDs, we used HLM analysis to determine the effect of the GBG on the level of challenging behavior as well as trend over time. HLM analyses were conducted using SAS PROC MIXED (Little, Milliken, Stroup, & Wolfinger, 1996) to estimate the parameters of interest—β_{2jk} for the treatment effect on the time trend. The following three-level regression model has been suggested (Van den Noortgate & Onghena, 2003) as a method of summarizing such results:

$$\Upsilon_{ijk} = \beta_{0jk} + \beta_{1jk} D_{ijk} + \beta_{2jk} T_{ijk} + \beta_{3jk} D_{ijk} T_{ijk} + e_{ijk}.$$

In this equation, the variables represent the following: Υ_{ijk} is the outcome score at measurement occasion *i*, for subject *j*, from study *k;* T_{ijk} is a time-related variable that equals 0 on the first day of the treatment phase; D_{ijk} is a dummy variable that equals 0 in the baseline phase and 1 in the treatment phase; $T_{ijk}D_{ijk}$ is the interaction between T_{ijk} and D_{ijk}; β_{0jk} is the baseline intercept (i.e., the overall mean of the outcome); β_{1jk} is the linear trend during the baseline; β_{2jk} is the treatment effect on the intercept for the trend during the intervention phase (i.e., immediate treatment effect); and β_{3jk} is the treatment effect on the time trend.

Data Analysis for Research Question 2

To answer the second research question concerning whether variations in intervention characteristics (i.e., fidelity, intervention agent, duration, setting, and/or use of rewards) affected GBG outcomes, we examined effect sizes for group designs and calculated and examined PND (Scruggs, Mastropieri, & Castro, 1987) and PAND (Parker, Hagan-Burke, & Vannest, 2007) for all SSEDs. These results allowed us to review whether variations in these variables appeared to affect the outcome data. Data on these study features were compared across studies with modest effects compared with those with moderate or large effects. We were not able to enter the intervention characteristics into the HLM models for analysis because not all data were available for each study.

PND and PAND are nonparametric estimators of effect size commonly used to analyze of SSED studies (Alresheed, Hott, & Bano, 2013). According to the metrics of PND and PAND, little or no overlap between baseline and intervention is considered evidence of a treatment effect (Kratochwill et al., 2002). Calculation of PNDs required counting the number of treatment data points that were greater than the highest data point in baseline, dividing this value by the total number of treatment points, and multiplying this number by 100. For studies in which the expectation was to decrease behavior using the GBG, PND calculations required use of the lowest baseline data point and use of treatment data points below the lowest baseline data point. Criteria for PND effectiveness are as follows: PND less than 50% is considered ineffective, PND between 50% and 70% is considered mildly effective, PND between 70% and 90% is considered moderately effective, and PND greater than 90% is considered highly effective (Mastropieri, Scruggs, Bakken, & Whedon, 1996).

However, PND has some limitations, including the omission of the majority of baseline data points and the overreliance on a single data point that may be an outlier (Parker et al., 2007). PAND serves as a complementary measure of intervention effectiveness because it directly addresses the criticism leveled against PND by using all data points in the analysis. PAND refers to the percent of all data remaining after removing the overlap between the baseline and intervention phases (Parker et al., 2007). The PAND calculation requires identifying the overlapping data points, dividing the number of overlapping data points by the total number of data points and subtracting this number from 100. However, the PAND calculation requires 20 or more data points to have a minimum of five data points for each cell of a 2 × 2 table, which is the same as used for a chi-square analysis (Parker et al., 2007).

Most of the study features were inspected for whether their qualitative components affected effect size, PND, or PAND. For duration, we calculated and compared the median duration of GBG implementation across studies that indicated a moderate or high effect compared with studies with a null, questionable, or modest effect. Determination of median required ordering the number of days of duration and finding the midpoint for each group.

Results

The purpose of the present review was to (a) describe the strength of effects of the GBG on challenging behaviors in school and classroom settings and (b) to critically examine the differences in outcomes with respect to the intervention agent, setting, duration of the GBG, reward procedures. In the following sections, we present our findings concerning the effect of the GBG on school and classroom behaviors as well as how characteristics of the intervention have affected the results of each study. First, we review the corpus of studies and the findings of each study according to the challenging behavior on which the researchers focused. Then, we present our findings for each of the two research questions.

Corpus of Studies and Summary of Individual Study Findings

Twenty-two articles published in 14 journals met all inclusion criteria. Articles about the GBG and challenging behaviors in school and classroom settings were most frequently published in the *Journal of Applied Behavioral Analysis* (*n* = 5) and the *Journal of School Psychology* (*n* = 3). Two articles were published in *Psychology in the Schools* and two were published in *Behavior Modification*. The 22 articles included 16 studies with SSEDs and 6 with experimental designs. Four articles reported on two longitudinal studies (Kellam, Ling, Merisca, Brown & Ialongo, 1998; Kellam, Rebok, Ialongo, & Mayer, 1994; Leflot et al., 2010; Leflot et al., 2013).

School and classroom challenging behavior outcomes that were addressed in these articles were disruptive behavior (*n* = 8), off-task/on-task behavior (*n* = 6), aggression (*n* = 5), talking out (*n* = 4), out-of-seat behavior (*n* = 4), peer acceptance and rejection (*n* = 2), rule violations (*n* = 2), antisocial negative behaviors (*n* = 1), appropriate and inappropriate social interactions (*n* = 1), externalizing behavior (*n* = 1), and swearing or negative comments (*n* = 1). Some articles addressed more than one DV with the GBG.

Disruptive Behavior

Disruptive behavior was a combined variable consisting of multiple challenging behaviors such as talking out, out of seat, and touching others or behavior that disrupts activities of another student such as motor activities, noisemaking, verbalizations, or aggression. All the studies that addressed disruptive behavior as a DV used SSEDs to study GBG effects. Across the eight studies (see Table 1) that used the GBG in an effort to reduce disruptive behavior, only Fishbein and Wasik (1981) and Lannie and McCurdy (2007) found the GBG to be ineffective for at least one case in each of their studies. Although McCurdy et al. (2009) indicated that one case experienced an ineffective intervention according to PND, use of PAND for analysis reflected a highly effective intervention.

On-Task and Off-Task

On-task behavior was defined as engaged in tasks as requested, paying attention to academic activities, and visibly engaged in tasks. Off-task referred to behaviors that were incompatible with being engaged in assignments or instruction, failing to pay attention to academic activities, and being visibly unengaged in instructional tasks. On- and off-task behavior were typically mutually exclusive DVs in these studies—that is, a student could not be categorized as on and off-task at the same time. Six studies addressed on- and off-task behavior (see Table 1). Of these six, three used SSEDs (Darch & Thorpe, 1977; Fishbein & Wasik, 1981; Lannie & McCurdy, 2007) and three (Dion et al., 2011; Leflot et al., 2010; Leflot et al., 2013)

TABLE 1

Study Outcomes by Dependent Variable

Dependent variable	Study author(s)	Design	Effect size (*d*)	PND%	PAND%
Disruptive Behavior	Darveaux (1984)	SSED		Child 1 = 100 Child 2 = 100	
	Donaldson, Vollmer, Krous, Downs, and Berad (2011)	SSED		Experimenter: Class 1 = 100 Class 2 = 100 Class 3 = 100 Class 4 = 91.66 Class 5 = 100 Teacher: Class 1 = 100 Class 2 = 91.6 Class 3 = 88.88 Class 4 = 78.97 Class 5 = 100	Experimenter: Class 1 = 100 Class 2 = 100 Class 3 = 100 Class 4 = 96.15 Class 5 = 100 Teacher: Class 1 = N/A Class 2 = 96 Class 3 = 93.75 Class 4 = 90.9 Class 5 =100
	Fishbein and Wasik (1981)	SSED		Library = 100 Classroom = 0	
	Lannie and McCurdy (2007)	SSED		27.27	
	McCurdy, Lannie, and Barnabas (2009)	SSED		Lunch periods: K, 3rd = 100 1st, 2nd = 100 4th, 6th = 33.33	Lunch Periods: K, 3rd = 100 1st, 2nd = 100 4th, 6th = 95
	Medland and Stachnik (1972)	SSED		90	89.72
	Ruiz-Olivares, Pino, and Herruzo (2010)	SSED		Teacher 1: 100 Teacher 2: 100	
	Salend, Reynolds, and Coyle (1989)	SSED		100	100

(continued)

TABLE 1 (continued)

Dependent variable	Study author(s)	Design	Effect size (d)	PND%	PAND%
Off-task/On-task	Darch and Thorpe (1977)	SSED		100	
	Dion et al. (2011)	Exp	Attentive = .81 Inattentive = 1.22		
	Fishbein and Wasik (1981)	SSED		Library = 85.71 Classroom 28.57	
	Lannie and McCurdy (2007)	SSED		90.91	
	Leflot, van Lier, Onghena, and Colpin (2010)	Exp	Wave 2 = 0.61 Wave 4 = 0.22		
	Leflot, van Lier, Onghena, and Colpin (2013)	Exp	Wave 4 = −0.47 (for low on-task)		
Aggression	Kellam, Ling, Merisca, Brown, and Ialongo (1998)	Exp	Overall = 0.11; (up to d = 0.32 at higher levels of aggression)		
	Kellam, Rebok, Ialongo, and Mayer (1994)	Exp	−0.39		
	Kleinman and Saigh (2011)	SSED		77.78	
	Leflot et al. (2013)	Exp	0.48 (low on-task)		
	Saigh and Umar (1983)	SSED		91.67	95.83
Talking Out	Barrish, Saunders, and Wolf (1969)	SSED		Math = 100 Reading = 100	Math = 100 Reading = 100
	Leflot et al. (2010)	Exp	Wave 2 = −.62 Wave 4 = −.29		
	Saigh and Umar (1983)	SSED		75	100
	Salend et al.(1989)	SSED		Class A = 100 Class B = 87.25	Class A = 100 Class B = 89.29
Out-of-seat	Barrish et al. (1969)	SSED		Math = 100 Reading = 100	Math = 100 Reading = 100
	Kleinman and Saigh (2011)	SSED		100	
	Leflot et al. (2010)	Exp	Wave 2 = .13 Wave 4 = .08		
	Saigh and Umar (1983)	SSED		66.67	83.33

(continued)

TABLE 1 (continued)

Dependent variable	Study author(s)	Design	Effect size (*d*)	PND%	PAND%
Peer Acceptance/ Rejection (acc/rej)	Leflot et al. (2013)	Exp	0.41 (rej) Low on-task		
	Witvliet, van Lier, Cuijpers (2009)	Exp	0.34 (acc)		
Rule Violations	Tanol, Johnson, and McComas (2010)	SSED		Class 1 = 100 Class 2 = 100	
	Bostow and Geiger (1976)			100	
Antisocial Negative Behaviors	McGoey, Schneider, Rezzetano, Prodan, and Tankersley (2010)	SSED		Class 1 = 56.25 Class 2 = 31.58 Class 3 = 68.42	
Appropriate and Inappropriate Social Interactions	Patrick, Ward, and Crouch (1998)	SSED		Inappropriate = 100 Appropriate = 100	Inappropriate = 100 Appropriate = 100
Externalizing Behavior	Witvliet et al. (2009)	Exp	0.45		
Swearing or Negative Comments	Salend et al. (1989)	SSED		100	100

Note. PND% = percent nonoverlapping data points; PAND% = percent all nonoverlapping data points; Exp = experimental design; SSED = single subject experimental design.

used experimental designs to study the effects of the GBG. The researchers for four studies specifically measured on-task behavior (Darch & Thorpe, 1977; Dion et al., 2011; Leflot et al., 2010; Leflot et al., 2013) and two measured off-task behavior (Fishbein & Wasik, 1981; Lannie & McCurdy, 2007). Only one study (Fishbein & Wasik, 1981) found a limited effect of the GBG intervention on off-task behavior. For the other studies, the GBG realized moderate to large effects.

Aggression

Five studies addressed aggressive behavior. The definition for aggression was physical contact such as hitting, kicking, tapping, tripping, pinching, throwing objects in the classroom, and destroying the property of others. Two studies (Kleinman & Saigh, 2011; Saigh & Umar, 1983) used SSEDs and three studies (Kellam et al., 1994; Kellam et al., 1998; Leflot et al., 2013) had experimental designs. The researchers who conducted the two SSED studies suggested that the GBG was a moderately to highly effective intervention for reducing aggression. All three experimental studies demonstrated that use of the GBG had modest effects on aggression. The study by Kellam et al. (1994) also indicated that the GBG had a greater reductive effect in more aggressive children, whereas Leflot et al.'s (2013) findings suggested that the modest effect was only for students who had low baseline on-task behavior and no effect on students who had high on-task behavior at baseline.

Talking Out

Four studies addressed talking out which defined primarily as talking without permission. Three of these four studies used SSEDs to study the effects of GBG implementation on talking out behavior (Barrish et al., 1969; Saigh & Umar, 1983; Salend et al., 1989). Overall, GBG implementation had moderately positive effects on the reduction of talking out. Barrish et al. (1969) suggested that the GBG was highly effective for reducing talking out behavior. Salend et al. (1989) indicated that the GBG was highly effective for one class and moderately effective for another class. Saigh and Umar (1983) and Leflot et al. (2010) found that the GBG was moderately effective for reducing talking out.

Out of Seat

Four studies focused on out-of-seat behavior. Out-of-seat behavior was defined as leaving one's seat or seated position without permission. Three studies employed SSEDs (Barrish et al., 1969; Kleinman & Saigh, 2011; Saigh & Umar, 1983) and one study (Leflot et al., 2010) used an experimental design. Barrish et al. (1969) and Kleinman and Saigh (2011) indicated that the GBG was a very effective intervention to reduce out-of-seat behavior. Saigh and Umar (1983) found the GBG to be a mildly effective intervention based on PND and a moderately effective intervention for out-of-seat behavior according to PAND. Leflot et al. (2010) indicated a near null effect of the GBG intervention on out-of-seat behavior.

Peer Acceptance and Rejection

Two studies (Leflot et al., 2013; Witvliet, van Lier, & Cuijpers, 2009) addressed peer acceptance and/or rejection. The definition for peer acceptance was that other students liked the student. Peer rejection referred to being liked least by classmates. Leflot et al. (2013) asked students to nominate all other students that they liked least (rejection). Witvliet et al. (2009) requested the opposite and asked students to nominate all other students that they liked most (acceptance). Both studies used experimental designs. Findings suggested that the GBG had a modest effect on increasing acceptance and decreasing rejection. In the case of decreasing peer rejection, Leflot et al.'s (2013) findings suggested that this was particularly true for students with low levels of on task behavior at baseline.

Rule Violations

Two studies (Bostow & Geiger, 1976; Tanol et al., 2010), both SSEDs, concerned rule violations in the classroom. Rule violations were generally defined as not following rules or engaging in behaviors against classroom expectations. Although the specific behaviors of interest for other studies could also be considered rule violations, those other behaviors were more specific than the definitions used in these two studies. These researchers found the GBG to be a highly effective intervention for reducing rule violations.

Antisocial/Negative Behavior

One study (McGoey, Schneider, Rezzetano, Prodan, & Tankersley, 2010) focused on antisocial/negative behavior using a SSED. Antisocial/negative behavior was defined as a composite of several behaviors including negative social interactions, off-task behavior, and tantrumming. Their findings suggested that the GBG is an ineffective intervention against these antisocial/negative behaviors.

Appropriate and Inappropriate Social Interactions

One study (Patrick, Ward, & Crouch, 1998) used an SSED to address appropriate and inappropriate social interactions. These behaviors were incompatible categories of behavior where appropriate behavior was defined as supportive verbal, physical, or gestural acts. Inappropriate social interactions were defined as aggressive verbal, physical, or gestural acts. In this study, the GBG appeared to be a highly effective intervention for increasing appropriate interactions and decreasing inappropriate interactions.

Externalizing Behavior

Using an experimental design, one study (Witvliet et al., 2009) addressed the effect of the GBG on externalizing behavior. Externalizing behavior referred to oppositional and conduct problems. Using teacher rating scales researchers measured students' oppositional and conduct problems as observed by teachers, findings revealed that use of the GBG had a modest effect on externalizing behavior.

Swearing/Negative Comments to Others

Using an SSED, Salend et al. (1989) used the GBG in an attempt to reduce swearing and negative comments among 19 high school students with emotional disturbance. Swearing referred to verbal statements or gestures pertaining to body parts designed for sexual activity or waste elimination, uncomplimentary references to others' parentage. Negative comments referred to negative verbal

TABLE 2

Parameter Estimates (With Standard Errors) From HLM Meta-analysis of GBG SSEDs

	Intercept	Baseline slope	Immediate treatment effect	Treatment effect on trend
Fixed effects	51.88** (8.05)	−0.51 (0.40)	−20.38* (7.30)	−0.03 (0.48)

Note. SSED = single subject experimental design; HLM = hierarchical linear modeling; GBG = Good Behavior Game.
*$p < .01$. **$p < .001$.

comments, complaints about assignments, or complaints about the instruction. Salend et al. (1989) suggested that the GBG was a highly effective intervention for reduction of swearing and negative comments.

Research Question 1: Effects of the GBG

The average Cohen's *d*, calculated from all group design studies, revealed a moderate effect ($d = .50$) of the GBG intervention on challenging behaviors in classroom and school settings. Using the data extracted from the graphs from the SSEDs, the HLM analysis revealed that the overall baseline mean (intercept) for challenging behavior was 51.88%. Across studies the immediate treatment effect (β_{2jk}) was −20.38%. These results indicated that a high rate of challenging behavior during the baseline phase was evident and an immediate decrease in the behavior occurred with introduction of the treatment. Both the baseline mean ($p < .001$) and immediate treatment effect ($p < .01$) were statistically significant. The treatment effect on trend (β_{3jk}) also decreased slightly, by .03% during the treatment phase; however, this effect was not statistically significant, $p > .05$ (see Table 2).

Research Question 2: Outcome Differences by Study Features

The GBG was found to be a highly effective intervention across a range of challenging classroom and school behaviors (see Table 1). According to Cohen's *d*, PND, and PAND, 10 studies indicated a null, modest, or mild effect for at least one case (Fishbein & Wasik, 1981; Kellam et al., 1994; Kellam et al., 1998; Lannie & McCurdy, 2007; Leflot et al., 2010; Leflot et al., 2013; McCurdy et al., 2009; McGoey et al., 2010; Saigh & Umar, 1983; Witvliet et al., 2009). Through this research question we examined some potential reasons for these differing results, including a review of characteristics including fidelity, who served as the intervention agent, the type of training provided to the GBG interventionist, setting, intervention duration, and rewards provided to students. Table 3 summarizes these findings.

Fidelity

Fidelity was reported in eight studies (Dion et al., 2011; Donaldson, Vollmer, Krous, Downs, & Berad, 2011; Lannie & McCurdy, 2007; Leflot et al., 2010; Leflot et al., 2013; McCurdy et al., 2009; Salend et al., 1989; Tanol et al., 2010). Fourteen studies did not report fidelity results. Fidelity scores were near or above 80% in all

TABLE 3

Features of GBG Studies

Study	Teacher implement	Interventionist training: Lecture	Lecture and feedback	Model	NR	Setting (elem/sec)	Duration (in days)	Rew (Y/N)	Fid	Eff
Barrish et al. (1969)	•				•	Elem	56	Y	N	High, High
Bostow and Geiger (1976)	•	•				Elem	25	Y	N	High
Darch and Thorpe (1977)	Student teacher				•	Elem	27	Y	N	High
Darveaux (1984)	•		•			Elem	20	Y	N	High
Dion et al. (2011)	•		•			Elem	120	Y	Y	High
Donaldson et al. (2011)	•			•		Elem	Unk	Y	Y	High
Fishbein and Wasik (1981)	Librarian and teacher				•	Elem	65	Y (librarian only)	N	High, Null
Kellam et al. (1998)	•	•				Elem	Unk	N	N	Null, Modest
Kellam et al. (1994)	•	•				Elem	Unk	N	N	Modest
Kleinman and Saigh (2011)	•	•				Sec	21	Y	N	High-Mod
Lannie and McCurdy (2007)	•		•			Elem	17	Y	Y	Ineff-High
Leflot et al. (2010)	•		•			Elem	320	N	Y	Null-Mod
Leflot et al. (2013)	•		•			Elem	320	N	Y	Modest-Modest
McCurdy et al. (2009)	Lunch assistants	•				Elem	20	Y	Y	Ineff, High
McGoey et al. (2010)	•				•	Elem	37	Y	N	Ineff-Mild
Medland and Stachnik (1972)	•				•	Elem	55	Y	N	Moderate
Patrick et al. (1998)	•				•	Elem	20	N	N	High
Ruiz-Olivares et al. (2010)	•	•			•	Elem	30	Y	N	High
Saigh and Umar (1983)					•	Elem	25	N	N	Mild-High
Salend et al. (1989)	•				•	Sec	14	Y	Y	Moderate-High
Tanol et al. (2010)	•		•			Elem	40	Y	Y	High
Witvliet et al. (2009)	•		•			Elem	Unk	Y	N	Modest

Note. Teacher Implement refers to whether a teacher was the primary implementer of the GBG intervention. NR under interventionist training indicates: not recorded or not given. In the setting column, elem refers to elementary school and sec refers to secondary school (middle school/high school). Rew refers to whether rewards were indicated and/or described. Fid refers to whether fidelity was measured. Eff refers to the effectiveness. A dash (-) between two effectiveness levels (e.g., mod-high) describes effectiveness across two DVs. A comma (,) between two effectiveness levels (mod, high) indicates effectiveness across cases on one DV.

except for one (Donaldson et al., 2011) of the eight studies that reported fidelity outcomes. Donaldson et al. (2011) measured fidelity and determined that fidelity averaged 60%; however, the lower fidelity in this study did not appear to affect the intervention outcomes as PND scores were found to be more than 90%. Two groups of researchers (McGoey et al., 2010; Ruiz-Olivares et al., 2010), both conducting SSEDs, indicated that they did not formally assess fidelity, but both acknowledged the lack of fidelity assessment as a limitation. One additional study (Patrick et al., 1998) suggested training to ensure fidelity but did not measure fidelity.

Intervention Agent

School staff served as intervention agents in the majority of studies (n = 21) and teachers were the most common (n = 19). In one study (Darch & Thorpe, 1977), a student teacher was the intervention agent and the GBG was found to be highly effective for off-task and out-of-seat behavior. In another study (Fishbein & Wasik, 1981), the school librarian was the primary implementer of the intervention rather than the classroom teacher even though effects were measured in both the classroom and the library. The GBG appeared to reduce disruptive and off-task behavior in the library, but not in the classroom. Finally, McCurdy et al. (2009) focused on behavior in the school cafeteria where lunchtime supervision staff implemented the GBG intervention and, overall, found large effects.

Thirteen articles mentioned interventionist GBG training or training materials. Seven studies (Darveaux, 1984; Dion et al., 2011; Lannie & McCurdy, 2007; Leflot et al., 2010; Leflot et al., 2013; Tanol et al., 2010; Witvliet et al., 2009) referred to training as including a combination of a lecture and or follow-up consultation and feedback with teachers upon implementation. Donaldson et al. (2011) also indicated that training procedures included the experimenter implementing the GBG in the presence of the teacher prior to the teacher assuming responsibility for the GBG. See Table 3 for a summary of the training methods used in these studies.

Of the 10 articles that reported limited to null effects on various outcome variables, two did not indicate how intervention agents were trained (Fishbein & Wasik, 1981; McGoey et al., 2010). McCurdy et al. (2009) did indicate training for the interventionist. However, the training appeared to be quite brief as interventionists participated in one 90-minute training procedure that included role-play and feedback. In one study (Leflot et al., 2010), GBG effects were minimal on one of the DVs (out-of-seat behavior); however, interventionist training appeared to be somewhat intensive as intervention agents were provided manuals, three half-day trainings, and 10 one-hour observations of their implementation.

Duration

Duration data were available, calculated, or estimated for 18 of 22 of the studies. Seven articles with duration information (Fishbein & Wasik, 1981; Lannie & McCurdy et al., 2007; Leflot et al., 2010; Leflot et al., 2013; McCurdy et al., 2009; McGoey et al., 2010; Saigh & Umar, 1983) had modest, questionable, or null effects on at least one DV. The duration data for the highly or moderately effective studies were compared with the duration data from studies with modest or questionable effects. For the studies with high or moderate effects, the median number of intervention days was 26 (range = 14-120). The median for duration for studies with modest, questionable, or null effects was 37 days (range = 17-320). This suggests that longer duration of GBG implementation does not necessarily mean better outcomes.

Setting

All 22 articles described the instructional context in which the GBG study took place. Twenty of the studies were conducted in elementary schools and classrooms. Two studies were conducted in secondary school settings—that is, a ninth-grade history class (Kleinman & Saigh, 2011) and a residential setting with high school students with emotional disturbance (Salend et al., 1989). Overall, school setting did not appear to affect GBG findings as moderate to large effects were found across elementary and secondary settings.

Rewards

The GBG naturally provides teachers with a vehicle through which to administer rewards to their students. Rewards in the GBG serve to increase students' use of appropriate behaviors versus more challenging behaviors. Rewards or winning were referred to in 16 articles (see Table 3). Most articles indicated if verbal, tangible, or social/activity rewards were used. However, Saigh and Umar (1983) did not mention the type of rewards given beyond saying they used rewards. Three articles indicated verbal praise (Darch & Thorpe, 1977; Dion et al., 2011; McCurdy et al., 2007). Tangibles were the most commonly used type of reward (n = 14; Barrish et al., 1969; Bostow & Geiger, 1976; Darveaux, 1984; Donaldson et al., 2011; Fishbein et al., 1981; Kellam et al., 1994; Kleinman & Saigh, 2011; Lannie & McCurdy, 2007; McGoey et al., 2010; Medland & Stachnik, 1972; Ruiz-Olivares et al., 2010; Salend et al., 1989; Tanol et al., 2010; Witvliet et al., 2009), and were often combined with verbal praise.

Studies that utilized tangible rewards appear to have had the highest effects as 9 of the 14 studies that used tangibles had high or moderate effects. Five articles (Fishbein & Wasik, 1981; Kellam et al., 1994; Lannie & McCurdy, 2007; McGoey et al., 2010; Witvliet et al., 2010) indicated use of tangible rewards but demonstrated no effect or modest or questionable effects. In these five studies, rewards were delivered daily (Fishbein & Wasik, 1981; McGoey et al., 2010; Witvliet et al., 2009), daily and weekly (Kellam et al., 1994), or the schedule was unspecified (Lannie & McCurdy, 2007). Two of these studies (Lannie & McCurdy, 2007; McGoey et al. 2010) made use of edibles for rewards, particularly candy. McGoey et al. (2010) appeared to offer a variety of options but the teacher initially selected these rewards.

In total, only four studies made use of preference assessments (Kleinman & Saigh, 2011; Lannie & McCurdy, 2007; Saigh & Umar, 1983; Salend et al., 1989). Three of the studies that indicated use of preference assessments had moderate or large effects (Kleinman & Saigh, 2011; Saigh & Umar, 1983; Salend et al., 1989) and one suggested modest effects (Lannie & McCurdy, 2007). Perhaps one of the most poignant findings with regard to the GBG and reward use is that one study (Fishbein & Wasik, 1981) found that the GBG had a large effect when a reward was used and a null effect when not used. Rewards used with the GBG appear be a critical component with regard to increasing students' appropriate behaviors and simultaneously decreasing challenging behavior.

Discussion

By evaluating Cohen's *d*, PND, PAND, and results of an HLM analysis, we concluded that, overall, the GBG appears to have a moderate to large effect on challenging behaviors in school and classroom setting. Results of the HLM analysis for SSEDs suggested that at baseline challenging behavior was high and GBG implementation resulted in an immediate and significant change in level. The GBG also appeared to have a continued effect throughout the intervention phase as the trend continued to evidence an extremely slight decrease in challenging behavior. Use of HLM to evaluate the overall effect of SSEDs proved useful as all data points from SSEDs could be used and simultaneously analyzed, neither of which is possible using overlapping data metrics such as PND or PAND.

Effectiveness of the GBG

An examination of effects at the individual study level revealed PNDs and PANDs with moderate to large effects overall. Of 45 PNDs across 16 studies, 37 were in the moderate to large effect range and 31 PNDs were over 90%. Eight PNDs from five studies (Fishbein & Wasik, 1981; Lannie & McCurdy, 2007; McCurdy et al., 2009; McGoey et al., 2010; Saigh & Umar, 1983) were indicative of mild or null effects; however, two of these (McCurdy et al., 2009; Saigh & Umar, 1983) were artifacts of the PND metric, as PAND suggested moderately to highly effective interventions. In these cases, PND's overreliance on outlying baseline data points appeared to be a problem. These results indicate that only 6 of 45 PNDs reflected mild or null effects. Interestingly, five of these six PNDs were for DVs that consisted of a combination of multiple behaviors, for example, disruptive behavior and antisocial/negative behaviors. This observation may speak to the need for researchers to improve their operationalization of the DVs and measure effectiveness with greater specificity regarding behaviors and classes of behavior.

Across the group design studies, Cohen's *d* tended to be in the moderate range. Shadish, Rindskopf, and Hedges (2008) suggested that these effect size differences (i.e., between Cohen's *d* and PND/PAND) are somewhat common and unsurprising because SSEDs contain within-subject variability as the primary source of variation. On the other hand, experimental designs produce between subject variability. Data from an individual are expected to be much less variable than data based on a group. Another possible explanation might be simply that the single subject studies for which PND and PAND were calculated were typically based on observational data, whereas large *N*, experimental studies used rating scales more frequently. Although the use of rating scales and surveys allows researchers to efficiently obtain a large quantity of information, these data may be vulnerable to various sources of error, such as the error of recency, where a rater may remember only the most recent events or behaviors exhibited by a student.

Although large values for Cohen's *d* were not commonly found, Dion et al. (2011) did find large effects of the GBG on students' on-task behavior. In this study a larger effect was found for the GBG's effect on inattentive students' attention compared with the attention of attentive students, although *d* for both groups was large. This finding suggests that the GBG helped both groups of students to be more attentive but that the group that had more room to improve made more progress. Leflot et al. (2013) also found that the GBG had greater effects (on on-task behavior, aggression, and peer rejection) for students who had low levels of on-task behavior at baseline. Kellam et al. (1998) found an overall null effect in terms of effect size, but when they analyzed data by groups of students based on baseline aggression levels, they found that the GBG had modest effects on the aggression levels of students with higher initial levels of aggression. These findings have implications for use of the GBG among students with aggression and attention problems and various disabilities—that is, the GBG may be of particular use among students with attention- and aggression-related school problems.

Another interesting finding from analysis of the group design studies comes from the work by Leflot et al. (2010). Leflot et al. found a moderate effect of the GBG on talking out and on-task behavior; however, stronger effects were observed at the first post intervention period (Wave 2) than at the end of the following year with the same participants (Wave 4). These differences between waves are probably best explained by considering that there may be floor effects for DVs. For example, if students talk out at 0% of the time by Wave 2, they cannot improve any further which will eliminate subsequent effects. Findings from Leflot et al. (2010) are also reflective of the HLM findings from analysis of SSEDs in our study, as we found that the GBG effects were immediate and largely stable with only a very slight decrease over time. Overall, the GBG appears to be effective at reducing a variety of challenging behaviors in school settings. Disruptive behaviors including off-task, aggression, talking out, and out of seat, behaviors could potentially impede teaching and learning. A simple management procedure, such as the GBG, may facilitate increased time devoted to teaching and learning.

Impact of Study Features on Outcomes

We examined various features of every study to assess the GBG's impact on the outcome. One of these variables was fidelity. When considering intervention studies, fidelity is an important component as low fidelity scores threaten internal validity. After comparing findings for articles that contained fidelity information with those that did not contain that information, we found that the lack of fidelity data did not appear to have an effect on outcomes, as modest and large findings were found across articles with the information and across articles without the information. Nonetheless, the lack of fidelity information was a surprising finding, particularly given that measuring fidelity is a common research standard.

We compared studies with high/moderate effects as given by Cohen's *d* or PND with studies with modest, questionable or null outcomes to understand how various study features (interventionist experience and training, setting, duration, and reward use) contributed to magnitude of effect. Given that most studies provided some level of training for interventionists, it is difficult to determine how training affected outcomes, as there did not appear to be any particular differences in findings that could be associated with type of training provided. Researchers implemented the GBG primarily in elementary general education settings with moderate to high effects. Those GBG implementations conducted in secondary settings were also highly effective. Findings for duration suggested that GBG implementation did not require lengthy intervention, as studies with modest or null effects had longer durations than studies with moderate or large effects. This finding is also consistent with results of our HLM analysis, which indicated that GBG results in an immediate drop in challenging behavior, but only slight additional change over time.

Of the study features studied here, use of rewards appears to have some effect on magnitude of outcome. It appears that where modest or null effects were found, rewards were not used or were used in a limited way. For example, Fishbein and Wasik (1981) found that the GBG was highly effective in the library where the librarian administered rewards to students. The GBG was not effective in the classroom as the librarian was not present and rewards were not administered. Another issue concerns making the rewards of interest to the students. For example, when the teacher chose the rewards (McGoey et al., 2010), the students may not have a clear understanding of what rewards they would earn. Also, as highlighted, many of the researchers did not use preference assessments to identify rewards preferred by the students. Of the four articles that reported use of preference assessments, three indicated large or moderate effects. These strong effects support the use of preference assessments with GBG implementation.

Limitations

Like all studies this review has some limitations to address. First, this study relied on the already published peer reviewed literature. This decision means that sound implementations of the GBG conducted for dissertation or thesis research may have been missed. Furthermore, we were only able to use the information provided in the included articles. For example, information on study duration was not included in all articles. It is possible that these omissions introduced some error into our analysis of the effect of duration on GBG outcomes. Second, several studies were eliminated from this review because they did not include challenging behaviors in school and classroom settings as dependent variables. These additional studies illustrate merits of the GBG, for example, the relationship between the GBG and later substance abuse or depression. Also, studies were only included when the effect of the GBG could be isolated. Finally, a number of studies were excluded as they were not written in English (e.g., Pérez, Rodríguez, De la Barra, & Fernández, 2005, which was written in Spanish) and we were not able to code them. It is likely that these studies would also be useful in evaluating the GBG effects on challenging behaviors in school and classroom settings.

Implications for Practice

Several implications for practice must be highlighted as a result of this review. First, the GBG appears to be an effective intervention to address a variety of challenging behaviors that could potentially cause an interruption to the teaching and learning process in the classroom. With GBG implementation, teachers may be able to spend more time on teaching and less time responding to behavioral incidents in the classroom. This outcome is important given the academic growth requirements that schools face (Vannest et al., 2009). Additionally, the GBG appears to have immediate and moderate to large effects over short periods of time. The fact that the GBG shows effects after a short time may be particularly important for teachers in classrooms where challenging behavior is frequent and teachers are struggling with classroom management. In these situations, GBG use may facilitate changes in student behavior so that teaching and learning can take place.

Another important implication concerns the types of behaviors addressed with the GBG. Through our review, we found that most GBG research focused on externalizing, challenging behaviors. This finding is not surprising as this type of behavior is generally characterized as noncompliant, disruptive, and often aggressive. Students with this type of behavior may be more noticeable or present greater concern for teachers compared with other behaviors or needs (Cullinan, Evans, Epstein, & Ryser, 2003). Externalizing behavior problems are commonly referred to the school office for intervention (McIntosh, Campbell, Carter, & Zumbo, 2009), as they are likely to interfere with teaching, learning, safety, and/or relationships. Harrison et al. (2012) found that teachers consistently cite student distractibility as a major concern. With the GBG demonstrated as effective for decreasing off-task behaviors and increasing the amount of time that students pay attention, teachers who are concerned with student distractibility may also view its use positively and aim to incorporate it into their classroom management systems. The effectiveness of the GBG on externalizing and off-task behaviors offers promise for teaching and learning.

As this review demonstrates, the GBG has been implemented by individuals in a variety of school roles such as classroom teachers, student teachers, librarians, and lunchtime staff. The different school roles of the interventionists highlight the ease in which the GBG can be implemented under a variety of conditions. Additionally, the relatively brief training for interventionists evidenced in these articles suggests that the GBG can be used successfully without extensive GBG training. When teachers consider using the GBG as part of their classroom management procedures they should consider how they will reward desired behavior and whether they will conduct preference assessments to understand student interests. Given the results synthesized here, the GBG might be considered as a promising practice for the classroom.

Implications for Future Research

This review has illustrated the magnitude of the effect of the GBG on various school and classroom behavioral outcomes. We examined the effects of various GBG study features on the associated outcomes. One of the quality indicators for SSEDs is the inclusion of social validity measurement. Future researchers might examine the effects of high or low social validity in SSEDs on GBG outcomes as well. Even though researchers have studied the GBG numerous times, there are still gaps in the literature in terms of implementation and effect on problem behavior across grade levels, school settings, and disability status. Since most studies reviewed here focused on elementary students, further use in secondary schools to determine effects on problem behaviors such as attendance for older adolescents might be useful. Additionally, use of the GBG to address internalizing behaviors is another avenue for exploration. Future studies should also provide detailed information on study participants and various training procedures, as this information was often lacking in studies that met inclusion criteria. Finally, with the academic achievement requirements required by NCLB, future research might consider the effects of the GBG on academic outcomes.

References

References marked with an asterisk indicate studies included in the review.

Alresheed, F., Hott, B., & Bano, C. (2013). Single subject research: A synthesis of analytic methods. *Journal of Special Education Apprenticeship*, *2*, 1–18.

*Barrish, H., Saunders, M., & Wolf, M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, *2*, 119–124.

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: Promotion of self-regulation and the prevention of early school failure. *Development and Psychopathology*, *20*, 899–911. doi:10.1017/S0954579408000436

Bohannon, H., & Wu, M.-J. (2011). Can prevention programs work together? An example of school-based mental health with prevention initiatives. *Advances in School Mental Health Promotion*, *4*(4), 35–46.

*Bostow, D., & Geiger, G. (1976). Good behavior game: A replication and systematic analysis with a second grade class. SALT: *School Applications of Learning Theory*, *8*(2), 18–27.

Boyle, M. A., Samaha, A. L., Rodewald, A. M., & Hoffman, A. N. (2012). Evaluation of the reliability and validity of GraphClick as a data extraction program. *Computers in Human Behavior*, *29*, 1023–1027. doi:10.1016/j.chb.2012.07.031

Cameron, C., Connor, C., Morrison, F., & Jewkes, A. (2008). Effects of classroom organization on letter-word reading in first grade. *Journal of School Psychology*, *6*, 173–192. doi:10.1016/j.jsp.2007.03.002

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.

Cullinan, D., Evans, C., Epstein, M., & Ryser, G. (2003). Characteristics of emotional disturbance of elementary school students. *Behavioral Disorders*, *28*, 94–110.

*Darch, C., & Thorpe, H. (1977). The principal game: A group consequence procedure to increase classroom on-task behavior. *Psychology in the Schools*, *14*, 341–347. doi:10.1002/1520-6807(197707)14:3<341::AID-PITS2310140315>3.0.CO;2-3

*Darveaux, D. (1984). The Good Behavior Game plus merit: Controlling disruptive behavior and improving student motivation. *School Psychology Review*, *13*, 510–514.

*Dion, E., Roux, C., Landry, D., Fuchs, D., Wehby, J., & Dupéré, V. (2011). Improving attention and preventing reading difficulties among low-income first-graders: A randomized study. *Prevention Science*, *12*, 70–79. doi:10.1007/s11121-010-0182-5

*Donaldson, J. M., Vollmer, T. R., Krous, T., Downs, S., & Berad, K. P. (2011). An evaluation of the Good Behavior Game in kindergarten classrooms. *Journal of Applied Behavior Analysis*, *44*, 605-609. doi:10.1901/jaba.2011.44-605

Elswick, S., & Casey, L. B. (2012). The Good Behavior Game is no longer just an effective intervention for students: An examination of the reciprocal effects on teacher behaviors. *Beyond Behavior*, *21*, 36–46.

Embry, D. D. (2002). The Good Behavior Game: A best practice candidate as a universal behavioral vaccine. *Clinical Child and Family Psychology Review*, *5*, 273–297.

*Fishbein, J., & Wasik, B. (1981). Effect of good behavior game on disruptive library behavior. *Journal of Applied Behavior Analysis*, *14*, 89–93. doi:10.1901/jaba.1981.14-89

Harrison, J., Vannest, K., Davis, J., & Reynolds, C. (2012). Common problem behaviors of children and adolescents in general education classrooms in the United States. *Journal of Emotional and Behavioral Disorders*, *20*, 55–64. doi:10.1177/1063426611421157

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446. (2004). Retrieved from http://idea.ed.gov/explore/view/p/%2Croot%2Cstatute%2C

Kaufman, J. S., Jaser, S. S., Vaughan, E. L., Reynolds, J. S., Di Donato, J., Bernard, S. N., & Hernandez-Brereton, M. (2010). Patterns in office discipline referral data by grade, race/ethnicity, and gender. *Journal of Positive Behavior Interventions*, *12*, 44–54. doi:10.1177/1098300708329710

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.

*Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology*, *10*, 165–185. doi:10.1017/S0954579498001564

*Kellam, S., Rebok, G., Ialongo, N., & Mayer, L. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trial. *Journal of Child Psychology and Psychiatry*, *35*, 259–281. doi:10.1111/j.1469-7610.1994.tb01161.x

Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology*, *102*, 741–756. doi:10.1037/a0019237

*Kleinman, K. E., & Saigh, P. A. (2011). The effects of the Good Behavior Game on the conduct of regular education New York City high school students. *Behavior Modification*, *35*, 95–105. doi:10.1177/0145445510392213

Kratochwill, T., Stoiber, K., Christenson, S., Durlak, J., Levin, J., Talley, R., & Shadish, W. R. (2002). In T. R. Kratochwill & K. C Stoiber (Eds.), *Procedural and coding manual for review of evidence-based interventions* (Task force on the Evidence Based Interventions in School Psychology Sponsored by Division 16 of the American Psychological Association and Society for the Study of Psychology). Retrieved from http://www.indiana.edu/~ebi/documents/_workingfiles/EBImanual1.pdf

*Lannie, A. L., & McCurdy, B. L. (2007). Preventing disruptive behavior in the urban classroom: Effects of the Good Behavior Game on student and teacher behavior. *Education & Treatment of Children*, *30*, 85–98. doi:10.1353/etc.2007.0002

*Leflot, G., van Lier, P. A. C., Onghena, P., & Colpin, H. (2010). The role of teacher behavior management in the development of disruptive behaviors: An intervention study with the Good Behavior Game. *Journal of Abnormal Child Psychology*, *38*, 869–882.

*Leflot, G., van Lier, P. A. C., Onghena, P., & Colpin, H. (2013). The role of children's on-task behavior in the prevention of aggressive behavior development and peer rejection: A randomized controlled study of the Good Behavior Game in Belgian elementary classrooms. *Journal of School Psychology*, *51*, 187–199. doi:10.1016/j.jsp.2012.12.006

Little, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS® System for mixed models*. Cary, NC: SAS Institute.

Mastropieri, M. A., Scruggs, T. E., Bakken, J. P., & Whedon, C. (1996). Reading comprehension: A synthesis of research in learning disabilities. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities* (*Vol. 10*, Part B, pp. 201–227). Greenwich, CT: JAI Press.

*McCurdy, B., Lannie, A. L., & Barnabas, E. (2009). Reducing disruptive behavior in an urban school cafeteria: An extension of the Good Behavior Game. *Journal of School Psychology*, *47*, 39–54. doi:10.1016/j.jsp.2008.09.003

*McGoey, K. E., Schneider, D. L., Rezzetano, K. M., Prodan, T., & Tankersley, M. (2010). Classwide intervention to manage disruptive behavior in the kindergarten classroom *Journal of Applied School Psychology*, *26*, 247–261. doi:10.1080/15377903.2010.495916

McIntosh, K., Campbell, A., Carter, D., & Zumbo, D. (2009). Concurrent validity of office discipline referrals and cut points used on schoolwide positive behavior support. *Behavioral Disorders*, *34*, 100–113.

*Medland, M., & Stachnik, T. (1972). Good behavior game: A replication and systematic analysis. Journal of Applied Behavior Analysis, *5*, 45–51.

Nagler, E. M., Rindskopf, D. M., & Shadish, W. R. (2008). *Analyzing data from small N designs using multilevel models: A procedural handbook*. Washington, DC: U.S. Department of Education.

Nelson, J. R., Maculan, A., Roberts, M. L., & Ohlund, B. J. (2001). Source of occupational stress for teachers of students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders*, *9*, 123–131. doi:10.1177/106342660100900207

Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percent of all non-overlapping data (PAND): An alternative to PND. *Journal of Special Education*, *40*, 194–204.

*Patrick, C., Ward, P., & Crouch, D. (1998). Effects of holding students accountable for social behaviors during volleyball games in elementary physical education. *Journal of Teaching in Physical Education*, *17*, 143–156.

Pérez, V., Rodríguez, J., De la Barra, F., & Fernández, A. M. (2005). Efectividad de una estrategia conductual para el manejo de la agresividad en escolares de enseñanza básica [Effectiveness of a behavioral strategy of aggression management in elementary school children]. *Psykhe*, *14*, 55–62. doi:10.4067/S0718-22282005000200005

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear modeling: Applications and data analysis procedures* (2nd ed.). Thousand Oaks, CA: Sage.

*Ruiz-Olivares, R., Pino, M. J., & Herruzo, J. (2010). Reduction of disruptive behaviors using an intervention based on the good behavior game and the say-do-report correspondence. *Psychology in the Schools*, *47*, 1046–1058. doi:10.1002/pits.20523

*Saigh, P., & Umar, A. (1983). The effects of the good behavior game on the disruptive behavior of Sudanese elementary school students. *Journal of Applied Behavior Analysis*, *16*, 339–344. doi:10.1901/jaba.1983.16-339

*Salend, S., Reynolds, C., & Coyle, E. (1989). Individualizing the good behavior game across type and frequency of behavior with emotionally disturbed adolescents. *Behavior Modification*, *13*, 108–126. doi:10.1177/01454455890131007

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education*, *8*, 24–33.

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention*, *2*, 188–196.

Spaulding, S. A., Irvin, L. K., Horner, R. H., May, S. L., Emeldi, M., Tobin, T. J., & Sugai, G. (2010). Schoolwide social-behavioral climate, student problem behavior, and related administrative decisions: Empirical patterns from 1,510 schools nationwide. *Journal of Positive Behavior Interventions*, *12*, 69–85. doi:10.1177/1098300709332345

*Tanol, G., Johnson, L., & McComas, J. (2010). Responding to rule violations or rule following: A comparison of two versions of the Good Behavior Game with kindergarten students. *Journal of School Psychology*, *48*, 337–355. doi:10.1016/j.jsp.2010.06.001

Tingstrom, D., Sterling-Turner, H., & Wilczynski, S. (2006). The Good Behavior Game: 1969-2002. *Behavior Modification*, *30*, 225–253. doi:10.1177/0145445503261165

U.S. Department of Education. (2000). *Annual report to Congress on the implementation of the Individuals with Disabilities Act*. Washington, DC: Office of Special Education and Rehabilitative Services.

Van den Noortgate, W., & Onghena, P. (2003). Combining single case experimental data using hierarchical linear modeling. *School Psychology Quarterly*, *18*, 325-346. doi:10.1521/scpq.18.3.325.22577

Vannest, K., Temple-Harvey, K., & Mason, B. (2009). Adequate yearly progress for students with emotional and behavioral disorders through research-based practices. *Preventing School Failure*, *53*, 73–84.

Walter, H. J., Gouze, K., & Lim, K. G. (2006). Teachers' beliefs about mental health needs in inner city elementary schools. *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*, 61–68. doi:10.1097/01.chi.0000187243.17824.6c

Witt, J., VanDerHeyden, A., & Gilbertson, D. (2004). Troubleshooting behavioral interventions. A systematic process for finding and eliminating problems. *School Psychology Review*, *49*, 156–167.

*Witvliet, M., van Lier, P. A. C., & Cuijpers, P. (2009). Testing links between childhood positive peer relations and externalizing outcomes through a randomized controlled intervention study. *Journal of Consulting and Clinical Psychology*, *77*, 905–915. doi:10.1037/a0014597

Authors

ANDREA FLOWER is an assistant professor of special education at The University of Texas at Austin, 1 University Station/D5300, Austin, TX 78712; e-mail: *AndreaFlower@austin.utexas.edu*. Her primary research interests focus on the use of positive behavior support strategies to facilitate academic and behavioral change. She is also interested in teacher preparation for managing challenging behavior in school settings.

JOHN W. MCKENNA completed his doctoral degree at the University of Texas at Austin. He is an assistant professor in the Department of Special Education at St. John's University, 8000 Utopia Pkwy, Queens, NY 11439; e-mail: *mckennj1@stjohns.edu*. His primary research interests are positive behavior supports, effective instructional strategies for students with EBD and at risk. He is also interested in responsible inclusion and wrap around services.

ROMMEL L. BUNUAN is a doctoral candidate in the Department of Educational Psychology, The University of Texas at Austin, 1 University Station/D5300, Austin, TX 78712; e-mail: *rommel.l.bunuan@gmail.com*. His research interests include multilevel modeling with a focus on mediation and meta-analysis of single-subject experimental designs studies.

COLIN S. MUETHING is a doctoral student in the Department of Educational Psychology, The University of Texas at Austin, 1 University Station/D5300, Austin, TX 78712; e-mail: *colinmuething@gmail.com*. His research interests involve the assessment and treatment of challenging behavior in individuals with developmental disabilities.

RAMON VEGA JR. is a doctoral student in the Department of Special Education, The University of Texas at Austin, 1 University Station/D5300, Austin, TX 78712; e-mail: *ramon.vega@utexas.edu*. His primary research interests are behavioral interventions for students with challenging behavior as well as examining the degree of contingency between events and behavior in a functional behavior assessment.