行政院國家科學委員會專題研究計畫 成果報告

代幣制與增強物之效果:以"超越中數線率"法對個案實驗研究所作後設分析(第 2 年)

研究成果報告(完整版)

計 畫 類 別 ： 個別型
計 畫 編 號 ： NSC 96-2413-H-004-003-MY2
執 行 期 間 ： 97 年 08 月 01 日至 98 年 07 月 31 日
執 行 單 位 ： 國立政治大學教育學系

計 畫 主 持 人 ： 馬信行

計畫參與人員： 碩士班研究生-兼任助理人員：張婉玟

公 開 資 訊 ： 本計畫可公開查詢

中 華 民 國 98 年 08 月 04 日

RUNNING HEAD: META-ANALYSIS OF REINFORCERS

A Comparison of The Relative Effectiveness of Different Kinds of Reinforcers: A PEM Approach

Hsen-Hsing Ma

Department of Education, National Chengchi University, Taiwan

Address: Department of Education, National Chengchi University, Wen-Shan District (116), Chi-Nan Road, Section 2, No. 64. Taipei, Taiwan

Telephone: 011-886-2-2938-7191 (Office **Number** and Fax); 011-886-2-2939-6704 (H)

E-mail: gxyzwgcd@nccu.edu.tw

Author's Note

This research was supported by grants from the National Science Council, Taiwan (96B160/NSC96-2413-H-004-003-MY2 and 96B161/NSC96-2413-H-004-003-MY2). The assistance of part-time assistants, Mr. Chia-Wei Tang and Miss Wan-Wen Chang is deeply appreciated. Correspondence should be sent to Hsen-Hsing Ma, Department of Education, National Chengchi University, Wen-Shan District (116), Chi-Nan Road, Section 2, No. 64, Taipei City, Taiwan. E-mail:gxyzwgcd@nccu.edu.tw

Abstract

The purpose of the present study was to apply the PEM approach to compare the relative effectiveness of different kinds of reinforcers used in behavior modification. Altogether 153 studies were located, which produced 1091 effect sizes. The grand mean of the PEM scores was .92. An important finding was that among the positive reinforcers, “activities” was the most effective while “edibles” and “objects” were the least effective. The feasibility of the PEM approach suggests that authors of operant research describe the effectiveness of treatment in terms of PEM scores and the criterion of Scruggs et al. (1986).

Descriptors: PEM approach; meta-analysis of single-case experiments; reinforcers

A Comparison of the Relative Effectiveness of Different Kinds of Reinforcers: A PEM Approach

In between-group research, since Glass (1976) proposed a formula using standard deviation as the unit for comparison, hundreds of meta-analyses have been conducted to draw conclusions about the efficacy of psychological, educational, and behavioral treatment. But studies using single-case experimental designs can not be analyzed with the conventional methodology of meta-analysis for between group researches because of the fact that the data points in the baseline and treatment phases are not randomly (independently) distributed but have autocorrelation. Mastropieri and Scruggs (1985-1986) thus proposed a nonparametric method, the percentage of nonoverlapping data (PND) approach, as a tool to calculate the effect size of experimental treatment for the meta-analysis of single-case experimental designs. The PND is the percentage of data points in the treatment phase over the highest point of the distribution in the baseline phase (or below the lowest point of data points in the baseline phase if the undesirable behavior is expected to decrease after the treatment is introduced). However, the PND approach has a serious problem in that if one or more outlier data points in the baseline phase reach ceiling/floor level, then the PND scores will be 0%. In order to improve this weakness of the PND approach, Ma (2006) has suggested the use of the percentage of data points exceeding the median of baseline phase (PEM) approach. To compute the PEM scores, one needs only to draw a horizontal median line in the baseline phase. This horizontal median line will hit the median when the number of data points in the baseline phase is odd and fall between the two middle points if the number of data points is even. The median line will then stretch out horizontally to the treatment phase. Then the percentage of data points of treatment phase above the median line can be calculated as the effect size scores. The null hypothesis is that if the treatment is not effective, then the data points in the treatment phase will fluctuate around the stretched median line and each data point will have a probability of .5 above the line. If instances of the undesired behavior are expected to decrease after the intervention is introduced, then the PEM score will be the percentage of data points below the median line in the treatment phase. Figure 1 demonstrates the calculation and comparison of the PEM and PND scores. In the upper panel of Figure 1, the median of the baseline is 18%, two data points are above and two other points below the median. Twelve data points in the treatment phase are above the stretched median line, hence the PEM score is 12÷15 = .8, a moderate effect. According to the criterion set by Scruggs, Mastropieri, Cook, & Escobar (1986), a score greater or equal to .9 is highly effective, a score greater than or equal to .7 but less than .9 is moderately effective, and a score less than .7 is questionable or not effective. However, because there was an outlier data point (50%) in the baseline phase and the highest data point in the treatment phase was also 50%, no data point in the treatment phase exceeded the highest point in the baseline phase, therefore the PND score is 0 ÷ 15 = .0, not effective. These results show why the PEM approach is more justifiable than the PND approach. The middle and the lower panels of Figure 1 show a PEM score of high and no effect of treatment, respectively. A comparison of the calculation of PEM and PND scores was conducted by Ma in (2006).

Figure 1 about here

Thus, using a PEM score to express the effectiveness of a behavioral treatment not only enables researchers to quantify their judgments, which is currently only based on visual inspection, but also makes the meta-analysis of within-subject studies possible.

The use of meta-analysis for studies employing single-case experimental designs not only allows for a comparison of the effectiveness of more than two different interventions, which is not easy, although not impossible in a single-case experimental design, but essentially it can also allow for the consolidation of the findings resulting from empirical within-subject research. Moreover, the validity of the PEM approach in terms of significant correlation with the judgment of original authors was confirmed by Gao and Ma (2006), Chen and Ma (2007), and Ma (2009).

Relative Effectiveness of Different Kinds of Reinforcer

The purpose of the present study is to apply the PEM approach (Ma, 2006) to compare the relative effectiveness of different kinds of reinforcers used in behavior modification. Can reinforcement really facilitate students' learning? Forness, Kavale, Blum, and Lloyd (1997) conducted a mega-analysis summarizing 18 meta-analyses to report the relative effectiveness of different interventions aiming at the improvement of behaviors of special education students and found that positive reinforcement combined with systematic ongoing assessment produced a mean effect size of 1.12 compared to a grand mean effect size of 0.54 of all 18 mean effect sizes.

Why should teachers in the educational settings use reinforcement to promote students' learning? It is desirable that all students know the value of knowledge and skills and are intrinsically motivated to learn, but it is rarely the case that such is so. If students have to wait until later to realize that knowledge and skills can be exchanged for the opportunity of advanced study in school or for higher status in the labor market, it may be too late to learn. Therefore, teachers in school settings have to employ extrinsic reinforcers to match a good performance by students in order to motivate students to learn. After the frequent use of reinforcement of good school performance, it can be expected that the act of learning will become a conditioned response, and that the learning can become intrinsically motivated with less extrinsic reinforcement. Is extrinsic reinforcement harmful to intrinsic motivation? Cameron and Pierce (1994) conducted a meta-analysis and found that rewards given for task completion or for quality of performance are not detrimental to intrinsic motivation.

Do different reinforcers have different effectiveness? It would be of practical value if the results of the present study could show that some reinforcers are more effective than others.

Reinforcement can be classified into four kinds: (a) positive reinforcement (giving a positive reinforcer), (b) punishment (giving a negative reinforcer), (c) punishment (withdrawing a positive reinforcer), and (d) negative reinforcement (withdrawing a negative reinforcer). In order to avoid the satiation resulting from the consecutive consumption of positive reinforcers, especially edibles, a token system has been frequently employed. The present study concentrates mostly on a comparison of the effectiveness of positive reinforcers including edible foods, tangible objects, activities, and tokens. Less attention is paid to the effectiveness of negative (aversive) reinforcers, but the mean effect sizes of punishment are presented for the purpose of comparison.

There have been many studies reporting success in the use of primary reinforcers to modify the behavior of participants. Williams, Koegel, and Egel (1981) successfully used M & M's and raisins as reinforcers under a functional condition (the target behavior became directly functional in procuring the reinforcer) to train appropriate imitation by severely handicapped children with autism. Kern, Ringdahl, Hilt, and Sterling-Turner (2001) used appropriate requests for a break by one child and for toy by another as reinforcers to replace incompatible inappropriate behavior. Osborne (1969) employed free time as a reinforcer in an examination of the effectiveness of reinforcement. The results of his study showed a sharply reduced frequency of out-of-seat behavior of six deaf participants. Attention may also function as a positive reinforcer to increase the instruction-following behavior of the participants (Schutte & Hopkins, 1970).

A token is a limited generalized means of payment. “Generalized” means that the token can be used to purchase different kinds of objects or privileges. “Limited” means that the token can only used in a certain setting. A token can be in the form of points, plastic money, or other kinds of symbolic currency. A token is a kind of secondary reinforcer that acquires its reinforcing properties through association with backup (primary) reinforcers.

McGinnis, Friman, and Carlyon (1999) applied contingent token rewards upon intrinsic mathematic motivation (measured as the amount of time spent on math tasks) and found that the effect was high. The token they used was stars, which could be exchanged for desired items, such as a sticker or yoyo.

Is the effectiveness of a positive reinforcer really higher than that of token economy? Both kinds of reinforcer have their strength in that a reinforcer in the form of “edibles”, “objects”, or “activities” would have a larger effect size than a token, because the participants receiving a reinforcer can consume it immediately while the participants receiving a token would experience a delay in the reduction of their deprivation. However, because of the avoidance of satiation, the participants reinforced with the token would have prolonged motivation. The null hypothesis was that there would be no significant difference in effectiveness between the immediate and delayed consumption of a reinforcer.

Assessment of the Needs (Deprivation) of Participants

According to the principle of behavior modification, in order to expect a desirable behavior to happen in the future, three conditions must be fulfilled as follows: A discriminative stimulus must be present; there must be a contingency for reinforcement of the target behavior, and the reinforcer must be able to satisfy the need of the individual. This principle can be described with the following formula:

$$\forall B : (S^D \wedge {}_BC_R \wedge R_N) \Rightarrow (S^D \Rightarrow (P(B)\uparrow)) \qquad (1)$$

where $\forall B :$ = This formula is valid for all operant behaviors which can be operationally defined; S^D = discriminative stimulus or situation under which the emission of a behavior will have a chance of being reinforced; ${}_BC_R$= contingency of reinforcement on the target behavior; R_N = the reinforcer can satisfy the need (or reduce the deprivation) of the individual; $\wedge$ = and; $\Rightarrow$ = if ...then; $P(B)\uparrow$ = the probability of the emission of the target behavior under the discriminative situation in the future will increase.

Researchers in the field of behavior analysis have recently paid more attention to the third condition of formula (1). Neef and Lutz (2001) found that the effect of more preferred reinforcers was higher than that of less preferred reinforcers. The results of Pace, Ivancic, Edwards, Iwata, and Page's (1985) study confirmed their belief that the success of reinforcement depends on the selection of suitable reinforcement schedules and contingencies. Their study demonstrated that providing a reward in the form of a preferred reinforcer had a higher effectiveness than providing one in the form of a non-preferred reinforcer. Glynn (1970) compared self-determined, experimenter-determined, and chance-determined token reinforcement interventions with a no-token intervention. He found that the effect of self-determined and experimenter-determined token intervention on the learning of history and geography material was superior to that of chance-determined and no-token interventions, and that the difference in the effectiveness of self-determined and experimenter-determined intervention was not significant.

How can the needs of a participant be assessed? Several approaches have been taken by researchers. Some researchers have depended on their professional judgment; others have conducted a survey of the significant others of the participant, such as parents and teachers; others have implemented a preference test; and others have used money as a reinforcer as it is a generalized medium with which the participant can purchase what he or she wants or needs outside the experimental setting. Another method used to assess the needs of the participant is functional analysis, whereby researchers investigate the environmental factors affecting the problem behavior. It explores the possible antecedents, consequences, and sequences of behavior that might have contributed to the formation of the problem behavior and a hypothesis of the cause of the problem behavior can be used in the formulation of an intervention strategy. Application of the Premark principle is also a way to find a reinforcer to meet the needs of a participant, such as incorporating the ritualistic behavior of child with autism into games to increase interaction in social play with siblings (Baker 2000). It is hypothesized that the effect size of an intervention will be larger when the reinforcer is determined by the participant rather by the interventionist because the participant-determined reinforcer can better meet the needs of the participant and serve as a mechanism to increase his or her motivation.

Which way of assessing the needs of participant is more effective? It is hypothesized that every author did his or her best to find a reinforcer that could reduce the deprivation of the participant; therefore the null hypothesis is that there will be no significant difference between different ways of assessing the needs of participants.

Method

Procedures for Locating Studies

The single-case experimental studies investigating the effect of reinforcers analyzed in this synthesis were obtained through a computer-assisted search of the relevant databases, including EBSCOhost, ERIC, and ProQuest. Descriptors included “token economy, token system, reinforcement, or reinforcer”. Additionally, a hand search of relevant journals of behavior analysis such as the Journal of Applied Behavior Analysis; Behavior Disorders; Behavior Modification; Behavior Assessment; Behavior Therapy; Behavior, Research and Therapy; and the Journal of Special Education was conducted. Studies that met the criterion that the data of baseline and treatment phases of a reversal or a multiple-baseline design were graphically displayed for individual participants in a time series format enabling the computation of PEM scores were included in this synthesis. Studies which employed an AB design were excluded because such a design lacks internal validity and alternative interpretations of a result can not be ruled out. Altogether 153 studies were included in the meta-analysis. They are listed in the Appendix.

Procedure for Coding a Sample Study

Variables in each of the following areas were coded:

1. Author(s)' conclusion on the overall effectiveness of an intervention: 2 = effective, including "highly effective", "successful intervention", "all data points of inappropriate behavior during the treatment phase were below the mean that occurred during the baseline phase", "noticeable reduction of inappropriate behavior"; 1 = moderately effective, including "slightly effective", "gradually improved", "above baseline level but was unstable, was not immediately effective but effective later"; and 0 = questionable or not effective. "Questionable" and "no effect" were combined together in this present study because it is hard to distinguish between them.
2. Categorization of independent variables: The reinforcers were coded regardless of whether they were delivered by the interventionist or by the participant him- or herself in the case of an intervention with self-management. The coding number and operational definition of the independent variables are listed as following:

   11. Edibles: Providing food; various edible reinforcers.

12. Objects: Providing objects, such as a toy; obsession (those items that the participants continually sought out or verbally requested).
13. Activities: Providing activities, such as interactive play; allowing a choice of activities; incorporating echolalia into a task response; presenting varied tasks instead of constant tasks; choosing books or stories to be read by the experimenter; sitting in a therapy ball instead of in a traditional classroom chair; playing electrovideo games; choice of preferred game (or toy); free time after remaining in seat; given preferred reading material; rhythmic entertainment; music; puzzles.
14. Giving a secondary reinforcer: Praise; attention (making statement or physical gesture to the participant); nonverbal approval, such as a smile and physical contact.
21. Giving negative reinforcers: Giving aversive stimulus including a reprimand; a stern “no”; icing on facial area contingent on bruxism; over-correction; positive practice overcorrection; loud noise; response blocking; shock; electric stimulation (Self-Injurious Behavior Inhibiting System); suppression (sharply saying “No” and briefly holding the part of the child’s body when the child performed self-stimulation; requiring the child to stand up and sit on the floor five to ten times contingent on an inappropriate behavior.
22. Withdrawal of positive reinforcers: Time out (isolating the participant from the reinforcing situation); extinction (ignoring the inappropriate behavior); using earlier curfew contingent on entering a residence late; withdrawal of attention; escape extinction (participant could escape only after a completing task); brief escape from dental treatment contingent upon co-operative behavior; break from the task only after completion of a part of the overall task; and sensory extinction.

23. Differential reinforcement of alternative appropriate behavior (DRA) with reinforcers other than tokens: Extinction of inappropriate target behavior and reinforcement of appropriate behavior.

30. Package of positive reinforcement and punishments other than removal of a token: Reinforcement contingent on desirable behavior and punishments with the exception of the removal of a token contingent on the undesirable behavior as a package.

40. Token: Tokens which can be redeemed for back-up preferred primary reinforcers at a later point in time, including points, lottery, and money.

41 Package of a token reward plus the withdrawal of a token: Token reinforcement contingent on appropriate behavior and removal of token contingent on inappropriate behavior.

42. Package of a token reward plus punishments other than the removal of a token: Token reinforcement contingent on appropriate behavior and other kind of punishments (other than the removal of a token) contingent on inappropriate behavior.

43. Differential reinforcement of alternative appropriate behavior (DRA) with tokens: Extinction of inappropriate target behavior and reinforcement of alternative appropriate behavior with tokens as a package.

Because the withdrawal of a token from a participant in the token economy might cause an aversive experience, and could diminish the reinforcing power of tokens, the present study also intends to find whether the use of a punishment measure other than the withdrawal of a token produces a greater effect than using the withdrawal of a token as a punishment measure in the token economy.

3. Categorization of dependent variables: Target behaviors were classified into six categories:

51. Quality of academic behavior as measured by accuracy, including: The learning of sign language, making appropriate verbal responses, showing skills in matching, and fluency in speaking.

52. Quantity of academic behavior as measured by the number of tasks completed, including: Following instructions, and percentage of times of taking medication.

53. Socially desirable behavior: Class attendance, attentive behavior in the classroom, engagement in interactive play, following of a dressing routine in the family, being on-task, showing compliance, making eye contact, making appropriate requests, being in-seat, appropriately recruiting teacher attention, payment of fines, and consumption of tokens.

61. Problem behavior, including: Anti-social behavior, returning too late to the dormitory, disruptive behavior, tantrums, perseverative speech, making a noise, not attending, making inappropriate movements of the body, talking rudely, packing food into the mouth without swallowing, thematic ritualistic activities, being off-task, expulsion and refusals during mealtime, social avoidance behavior, aggressive behavior, stealing, and stamping.

62. Self-injury including, Bruxism, and ingesting pills.

63. Self-stimulation, including: Rocking and hand-flapping, stereotyped behaviors, hand-clapping, object-mouthing, thumb-sucking, and excessive alcohol consumption.

4. Settings: Intervention settings were classified as (a) home; (b) institution, including clinic and various therapeutic centers, laboratory, residential facility, hospital classroom, room in an institution, infirmary playroom, home-style rehabilitation setting, and achievement placement; (c) school, including facilities in different levels of school, such as classroom and cafeteria, day-care program, co-operative student dormitory; and (d) other places, including unspecified room or playroom,

semi-naturalistic setting, facilities in the community, such as outdoor cafeteria, factory, recreation center, and empty meeting room.

5. Interventionists: Interventionists were classified into: (a) experimenter, including treatment provider, facilitator; research assistant, and nonprofessional staff, educational staff, observer, and recorder; (b) specialist, including author, researcher, therapist; instructor; counselor, clinician, and teaching parent; (c) teacher, including swimming coach and trainer; (d) tutor, including peer teacher and home tutor; and (e) parents, including caregivers.
6. Participants: Participants were classified as those with: (a) Attention deficit hyperactivity disorder (ADHD); (b) Autism Spectrum Disorder, Asperger's Syndrome, Down Syndrome; (c) mental illness, including psychiatric and psychotic patients; (d) emotional or behavior disorders; (e) learning disabilities; (f) mental retardation of different degrees of severity including mongoloid, educable mental retardation, developmentally disabled, global developmental delay, organic brain syndrome, left hemi paresis, a left visual field defect, brain injury, multiple handicaps, speech and language development delay, and developmental disabilities; (g) normal intelligence including participants with disruptive behaviors or deficient in sustaining attention, pre-delinquent behaviors, asthma, psychological problems, physical handicap; and (h) deafness and hearing impairment.
7. Age of participant. Age was divided into five groups: below 7, 7-12, 13-15, 16-18, and beyond 18 years old.
8. The length of the treatment phase was coded in order to examine whether a longer treatment phase has a higher effect.
9. The first pair of baseline-treatment phases and the pairs after that were coded so that the effect of the orthogonal slope change on the effect size of the second pair of baseline-treatment phases described by Scruggs, Mastropieri, and Casto (1987)

could be examined. They assumed that the data points of an appropriate target behavior in the second baseline would show a gradual downward trend and that those in the second treatment would show a gradual upward trend, and hence form an orthogonal slope change.

10. Methods of assessing the preference of reinforcers: This moderator refers to the approaches taken by the original authors to choosing a reinforcer that would satisfy the needs of a participant: 0 = no mention of assessment (it was assumed that the reinforcer was decided by the author either based on a review of the literature or on his or her professional judgment); 1 = the reinforcer was suggested by significant others of the participant such as a parent or teacher (the information was gathered through interview or questionnaire); 2 = the reinforcer was chosen based on a functional analysis (after an informal interview and observation, an experimentally manipulated multi-element design was conducted to investigate the environmental events that led to subsequent behavior changes in order to identify the factor which was critical in reinforcing the problem behavior; 3 = preference test (a list of potential reinforcers was compiled and arrayed to let the participant show his or her preference by consuming them or playing with them), including pair-wise comparison of preference; the reinforcer list was decided by the participant or formulated through discussion with the participant; the participant was able to redeem tokens in a “store” of back-up reinforcers; provision of choice-making opportunities in arranging the schedule of activities; observation of the behavior of the participant to comprehend what the participant preferred as reinforcers; use of the Premark principle (use of a high frequency activity as a reinforcer to reinforce less preferred activities); incorporating thematic ritualistic behaviors preferred by the child with autism into games to facilitate social play; 4. = using money as a reinforcer.

Reliability. One student in a doctoral program and one in a master's program in education serving as part-time research assistants independently conducted the calculation and coding of 42% of the scores of the PEM scores and judgment of original author(s). The percentage of agreement is calculated by the formula: agreements ÷ (agreements + disagreements). The reliabilities of coding for the reinforcer data were as follows: PEM scores = 344 ÷ 449 = .77, judgment = 428 ÷ 449 = .95, and, for the token data: PEM scores = 346 ÷ 365 = .95, and judgment = .83. In order to let the reliability approach 1.00 in the final results of the present study, the two assistants were asked to carry out all the calculation and coding of all of the PEM scores and judgment scores and the present author made the last check and resolved any disagreements.

Results

In the use of statistics to analyze the data and explain the findings of a meta-analysis of single-case experimental designs, there should be no violation of the three basic assumptions of parametric statistics (normality, independence, and variance homogeneity of the distribution of residuals); otherwise, a nonparametric statistic should be used, such as employing the Kruskal-Wallis analysis of variance by ranks to test the significance of difference between multiple groups and applying the Mann-Whitney U test to test that of two groups. Normally, data coming from the same body have autocorrelation and hence violate the assumption of the independent distribution of residuals. The lag 1 autocorrelation is the correlation between i and I + 1 of the same set of data and can be used as an indicator of the independence of the distribution of the residuals. If an article contains several effect sizes, then these effect sizes will likely have autocorrelation. The Levene statistic which is available in the SPSS package can be used to test the assumption of the variance homogeneity of the residuals.

When a mean effect size was used to represent the effect size of the results of each located study, the lag 1 autocorrelation of -.09 with a standard error of .08, $p > .05$, depicted that the data was independent and did not violate the assumption of the independent distribution of the residuals, which were produced by subtracting the mean PEM score of each study from the grand mean of the 153 studies. A t-test for single group resulted in $t(152) = 54.94$, $p < .01$ indicating that the grand mean effect size of .92 was significantly different from the hypothetical .5 PEM score. There were 1091 effect sizes from the 153 studies when every effect size in each study was used as a unit of analysis. The lag 1 autocorrelation of .16 with a standard error of .03, $p < .05$, was significantly different from zero, indicating that that the data violated the assumption of the independent distribution of the residuals. Therefore, non-parametric statistics had to be employed to analyze the 1091 effect sizes.

Validity

The Spearman's rank correlations between the judgments and the PEM scores were $r(1082) = .39$, $p < .01$. Table 1 exhibits that all three categories of the mean effectiveness judged by the original authors fall into the criteria suggested by Scruggs et al. (1986). Ninety percent of the treatments which utilized reinforcement showed a high effectiveness.

…………………………….

Insert Table 1 about here

…………………………

Analyses of Independent Variables, Dependent Variables, and Moderators

The results of the analyses of independent variables, dependent variables, and moderators are displayed in Table 2.

…………………………….

Insert Table 2 about here

…………………………

The Mean Effect Size of Independent Variables

The grand mean effect size of 1091 effect sizes was .90 with a standard deviation of .22. By testing the homogeneity of variances of the residuals of the 12 categories of independent variables (interventions or treatments), a Levene statistic revealed $F(11, 1079) = 10.97$, $p < .01$, indicating that the assumption of the homogeneity of variances of residuals was also violated. Applying nonparametric statistics demonstrated that the difference between the mean rank of effect sizes of 12 categories of independent variables was significant, with the Kruskal-Wallis analysis of variance by ranks showing $\chi^2 (11, N = 1091) = 57.79$, $p < .01$. The independent variables, of which the mean effect size was higher than .90, were “activities”, “token plus punishment”, “negative reinforcer”, “token”, “DRA with token”, and “positive reinforcer plus punishment”. The most effective reinforcer was “activities”, while the least effective interventions were those that involved using edibles, tangible objects, and token plus removal of token.

There were 148 effect sizes resulting from the intervention “activity”. Further analysis exhibited no significant difference between the mean ranks of the effect size of the six dependent variables, with Kruskal-Wallis analysis of variance by ranks showing $\chi^2 (5, N=148) = .77$, $p = .98$. This result depicts that the intervention “activities” can be as effective when used for the establishment of desirable behaviors as for eliminating undesirable ones. Multiple *post hoc* comparisons by means of the Mann-Whitney U test resulted in (13, 30, 21, 40) > (14, 41, 11, 12); 13 > (42, 22, 23); (21, 40) > 22; 43 > 11; and 21 > 23. The numbers represent the coding number of each subcategory of independent variable are shown in Table 2. The number in the parentheses refers to the fact that the mean ranks of these subcategories are all significantly larger than those of the subcategories behind the “>”. Within the parentheses, though their mean ranks are arranged in ranking order, their differences are not significant from each other. For instance, (21, 40) > 22 represents that the mean rank of the effect sizes of “negative reinforcer” and “ token” were significantly higher than that of “withdrawal of positive reinforcer”, and that the mean rank of the effect size of “negative reinforcer” was higher than that of “token”, but the difference was not significant. Table 2 shows that among the positive reinforcers, “activities” was the most effective while “edibles” and “objects” were the least effective. The effect of “activities” was also significantly greater than that of “differential reinforcement of alternative appropriate behavior (DRA)” , which makes use of an indirect way to reduce an inappropriate target behavior by means of ignoring the target behavior but reinforcing an alternative appropriate behavior, although the superiority of "activities” over the “DRA with token” did not reach a significant level.

Normally, the rank order of the mean effect sizes corresponds with that of the mean ranks; however, inconsistency can occasionally occur due to the different number of effect sizes as well as the variability and the outliers of the effect sizes in the treatment phase of the categories to be compared. For example, the reason why the category “token plus punishment” showed a higher mean effect size (.96) than that of “activities” (.95), but showed a lower mean rank of 525 when compared to the 624 of “activities” may be due to the different number of the outlier effect sizes in both categories. Among the 18 effect sizes of “token plus punishment” there were 10 (55.56%) effect sizes with a PEM score of 1.00 while out of 148 effect sizes of the “activities” there were 125 (84.46%) effect sizes with a PEM score of 1.00.

The reinforcers were classified into two kinds: One with immediate consumption of reinforcers and the other with delayed consumption of reinforcers including tokens, a package of token reward plus removal of token, a package of token reward plus punishment, and a differential reinforcement of appropriate behavior with tokens. The mean effect size of the token-reinforcement program (.89) with a standard deviation of .24 was slightly lower than that of the immediate consumption of reinforcers (M = .90, SD = .21), but the difference was not significant. A Mann-Whitney U test showed Z = -.38, p = .70. This demonstrates that token reinforcement showed only a scarce reduction in the power of the reinforcement, but that it can also prevent the participant from becoming satiated with primary reinforcers.

By comparing the package of "token plus punishment" with that of "token plus removal of token", it was found that the former strategy was more effective than the latter one. A Mann-Whitney U test showed that Z = -2.11, p = .04, depicting that the difference was significant. The rationale may be that if the token is removed as a method of punishment, it will be associated with an aversive experience, and hence its power of reinforcement will be diminished compared to a previous state prior to such use where its power was a positive one.

In the analysis of the assessment of preference for reinforcers, the difference between the mean ranks of effect sizes was on the edge of significance. A Kruskal-Wallis analysis of variance by ranks showed that $\chi^2(4, N=1091) = 8.42$, p = .08. But the post hoc comparison displayed a significant difference between "preference test" and "parent's suggestion" in favor of the former. The Mann-Whitney U test showed Z = -2.72, p < .01.

The Mean Effect Size of Dependent Variables

Employing nonparametric statistics, a test by means of Kruskal-Wallis analysis of variance by ranks revealed a significant difference between the mean effect sizes of the six dependent variables, $\chi^2(5, N = 1091) = 13.88$, p = .02. A Mann-Whitney U test showed that (63, 52, 53, 61) > 51, expressing that there was no significant difference among the effectiveness of interventions on the four dependent variables "self-stimulation", "works completed", "desirable social behaviors", and "problem behaviors", as well as that the effectiveness of the interventions on all of the four dependent variables **was** significantly higher than that of "works demanding accuracy". This result implies that it is more difficult to improve "accuracy of work" than to effect change in other kinds of dependent variables because accuracy of work is a matter of "can" rather than just "will".

The Effect of Other Moderators

By employing the Kruskal-Wallis analysis of variance by ranks to test whether the effectiveness of the interventions on the dependent variables was influenced by moderators, no significant difference was found for the moderators of setting, interventionist, category of participants, age and gender of participants, implying that the effectiveness of the interventions on the dependent variables can be generalized to different settings, interventionists, categories of participants, and ages and genders of participants. In a test as to whether it was the case that the longer the length of the treatment the higher the effectiveness, a Pearson correlation coefficient was calculated and no significant correlation between the length of treatment phase and the effect size was found, $r(1089) = -.01$, $p = .78$.

The only two moderators which showed significance were the order of pair of phases and the kind of design. The second pair of the reversal designs showed a significantly larger mean effect size than the first one. The Mann-Whitney U test showed that $Z = -2.36$, $p = .02$. The results in the research employing reversal design demonstrated a higher mean effect size than those using multiple-baseline designs. The Mann-Whitney U test showed that $Z = -2.98$, $p = .01$.

Discussion

The effectiveness in terms of the mean effect size of the 12 reinforcing strategies investigated in the present study ranges from .83 to .96, i.e., from a moderate to large effect size as compared with the criterion suggested by Scruggs et al. (1986)·

The finding that the effectiveness of “activities” and “token” were significantly higher than that of “praise”, “edible”, and “object” may possibly be explained in that the American participants in the relevant studies were not deprived personally in ordinary daily life in respect to food and object reinforcers, and thus it may not generalize the finding to individuals of less wealthy countries. Praise as a secondary reinforcer acquires its power for reinforcement after being paired by association with a primary reinforcer. Agents who deliver positive reinforcers frequently, though not always, conjoin them with praise. Such a combination functions as a kind of conditioned positive reinforcer. A nearly large mean effect size of .89 of praise found in the present study supports the findings in Reinke, Lewis-Palmer and Merrell’s (2008) study, which indicated that the performance feedback on the rate of teachers’ praise helped the teachers to increase behavior-specific praise and that increased praise by the teachers contingent on appropriate behaviors led to the reduction of disruptive behaviors of their students, and that, on the contrary, reprimanding inappropriate behavior led to an increase in disruptive behavior of their students. Hence, the best strategy of classroom management for a teacher should be to focus on praising an appropriate behavior rather than on reprimanding inappropriate behavior. The effect of this strategy was demonstrated experimentally as early as 1968 by Thomas, Becker, and Armstrong.

The finding that there was no significant correlation between the length of a treatment phase and its effectiveness is the same as in the results of the study conducted by Vegas, Jenson, and Kircher (2007). What is important to the magnitude of the effectiveness is not the length of the treatment phase but the power of the reinforcement provided by a reinforcer. The results of the present study, that the mean rank of the effect size of the second pair of baseline-treatment phases was higher than that of the first pair can serve to reduce concern on the orthogonal slope change mentioned by Scruggs, Mastropieri, & Casto (1987). They anticipated that the orthogonal slope change would threaten the effect size of the second baseline treatment pair. But the result of the present study demonstrates that such is not the case. In Ma's (2006) study, the mean effect size of the second baseline-treatment pair was higher than that of the first one, although the difference was small, and in the present study the same result was not only replicated but furthermore, the difference reached a significant level of .02.

The result that not all positive reinforcers have the same effectiveness also justifies the importance of the element of R_N in the formula (1), that is, the reinforcer must satisfy the need (or reduce the deprivation) of the individual. This conclusion was also partially supported by the result of the present study in that an intervention showed a higher effectiveness if the reinforcer was determined through the use of a preference test rather than simply by asking for the suggestion of significant others such as the parent. The finding that reinforcer "activities" had the highest mean effect size (.95) confirmed indirectly the Premark principle, which states that a high frequency activity can be used to reinforce a low frequency activity. Logically inferred, the Premark principle can be alternatively expressed as the use of an activity towards which the student feels a strong intrinsic motivation as an extrinsic reinforcer to motivate him or her to learn an important academic or social behavior for which his or her intrinsic motivation is weak. This finding has practical implications. As suggested by Kern, Babara, and Fogt (2002), academic activities can be associated with opportunities to make choices, such as choice of activity, choice of teaching of learning materials, and choice of task sequence, and the findings can be used to modify class-wide curricula. Their research demonstrated that curricular modification resulted in increased levels of engagement and decreased levels of destructive behavior and that it can be compatible with school policy.

The present study does not address the controversy between intrinsic and extrinsic motivation. However, all the studies located for this present research were conducted to improve behaviors in completing a task or performing an action according to an acceptable standard and were based on the assumption that the participant had weak intrinsic motivation. Thus the results of the present study imply that extrinsic reinforcement may motivate participants with a weak intrinsic motivation to improve quantitatively and/or qualitatively in their academic or social behaviors. Deci (1975) defined intrinsic motivation as a motivation to satisfy the individual's needs for feelings of competence and self-determination. Table 2 shows that "activity" was the most effective reinforcer. Among the activities listed in the independent variable coded "13", a prominent part was "allowing choice of activities", which by operational definition is in accord with the element of "self-determination" in intrinsic motivation. Although the average effect size of the reinforcer on the quality of an academic behavior was only moderate (.85), depicting that such quality is among the most difficult to be changed because it is also influenced by the prior achievements of the individual student, however, this result also shows that when the quality of an academic behavior was reinforced, a feeling of competence would be elicited in the student, which would facilitate the building of intrinsic motivation. Cameron and Pierce (1996) stated that the negative effect of a reward on intrinsic motivation was caused by non-contingent reward and could be prevented by rewarding people for completing work, solving problems successfully, or attaining a specified level of performance. The dependent variables in the current study "quantity of academic behaviors" and "quality of academic behaviors" fall within the category of "completing work" and "solving problems successfully, or attaining a specified level of performance", respectively and therefore would not be harmful to intrinsic motivation. Hence, by logical inference, intrinsic and extrinsic motivation could be mutual facilitators, that is, a student demonstrating excellence in intrinsically motivated academic behavior would be likely to accept extrinsic primary or secondary reinforcers delivered by significant others, such as parents, teachers and/or peers in the natural setting, and conversely, an extrinsically reinforced academic behavior would be likely to turn into a intrinsically motivated one because a frequently reinforced behavior would be likely to become a habit, which resembles an intrinsically motivated behavior. Future research should be done on the mutual fostering of both kinds of motivation.

The feasibility of the PEM approach suggests that PEM scores can be used to describe and judge the effectiveness of a treatment based on the figures provided in an article employing a single-case experimental design in accordance with the criterion set by Scruggs et al. (1986).

The analysis of whether a different schedule, duration, intensity, or amount of reinforcement would produce a different effect remains for further study as the number of studies included in the present investigation is too small to allow for such analysis.

References

Baker, M. J. (2000). Incorporating the thematic ritualistic behaviors of children with autism into games: Increasing social play interactions with siblings. *Journal of Positive Behavior Interventions, 2*(2), 66-84.

Cameron, J. & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research, 64*, 369-423.

Cameron, J. & Pierce, W. D. (1996). The debate about rewards and intrinsic motivation: Protests and *accusations… . Review of Educational Research, 66*, 39-51.

Chen, C. -W., & Ma, H. -H. (2007). Effect of intervention on disruptive behaviors: A quantitative synthesis of single-subject researches using the PEM approach. *The Behavior Analyst Today, 8*(4), 380-397.

Deci, E.L.(1975). Notes on the theory and meta-theory of intrinsic motivation. *Organizational Behavior and Human Performance, 15,* 130-145.

Forness, A. R., Kavale, K. A., Blum, I. M., & Lloyd, J. W. (1997). Mega-analysis of meta-analyses: What works in special education and related services. *Teaching Exceptional Children, 29(6),* 4-9.

Gao, Y. -J., & Ma, H. -H. (2006). Effectiveness of interventions influencing academic behaviors: A quantitative synthesis of single-subject researches using the PEM approach. *Behavior Analysts Today. 4,* 572-594.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Research, 5,* 3-8.

Glynn, E. L. (1970). Classroom applications of self-determined reinforcement. *Journal of Applied Behavior Analysis, 3*, 123-132.

Kern, L., & Bambara, L., & Fogt, J. (2002). Class-wide curricular modification to improve the behavior of students with emotional or behavioral disorders.

Behavioral Disorders, 27(4), 317-326.

Kern, L., Ringdahl, J. E., Hilt, A., & Sterling-Turner, H. E. (2001). Linking self-management procedures to functional analysis results. *Behavioral Disorders, 23*(3), 214-226.

Ma, H. -H. (2006). An alternative method for quantitative synthesis of single-subject researches: Percentage of data points exceeding the median. *Behavior Modification, 30* (5), 598-617.

Ma, H. -H. (2009). The effectiveness of intervention on the behavior of individuals with autism: A meta-analysis using percentage of data points exceeding the median of baseline phase (PEM). *Behavior Modification, 33*(3), 339-359.

Mastropieri, M. A. & Scruggs, T. E. (1985-86). Early intervention for socially withdrawn children. *The Journal of Special Education, 19(4),* 429-441.

McGinnis, J. C., Friman, P. C., & Carlyon, W. D. (1999). The effect of token rewards on "intrinsic" motivation for doing math. *Journal of Applied Behavior Analysis, 32*, 375-379.

Neef, N. A., & Lutz, M. N. (2001). Assessment of variables affecting choice and application to classroom interventions. *School Psychology Quarterly, 16*(3), 239-252.

Osborne, J. G. (1969). Free-time as a reinforcer in the management of classroom behavior. *Journal of Applied Behavior Analysis, 2*, 113-118.

Pace, G. M., Ivancic, M. T., Edwards, G. L., Iwata, B. A., & Page, T. J. (1985). Assessment of stimulus preference and reinforcer value with profoundly retarded individuals. *Journal of Applied Behavior Analysis, 18*, 249-255.

Schutte, R. C., & Hopkins, B. L. (1970). The effects of teacher attention on following instructions in a kindergarten class. *Journal of Applied Behavior Analysis, 3*, 117-122.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education, 8*(2), 24-33.

Scruggs, T. E., Mastropieri, M. A., Cook, S. B., & Escobar, C. (1986). Early interventions for children with conduct disorders: A quantitative synthesis of single-subject research. *Behavior Disorders, August,* 260-271.

Thomas, D. R., Becker, W. C., and Armstrong, M. (1968). Production and elimination of disruptive classroom behavior by systematically varying teacher's behavior. *Journal of Applied Behavior Analysis, 1,* 35-45.

Vegas, K. C., Jenson, W. R., & Kircher, J. C. (2007). A single-subject meta-analysis of the effectiveness of time-out in reducing disruptive classroom behavior. *Behavioral Disorders, 21*, 109-121.

Williams, J. A., Koegel, R. L., & Egel, A. L. (1981). Response-reinforcer relationships and improved learning in autistic children. *Journal of Applied Behavior Analysis, 14*, 53-60.

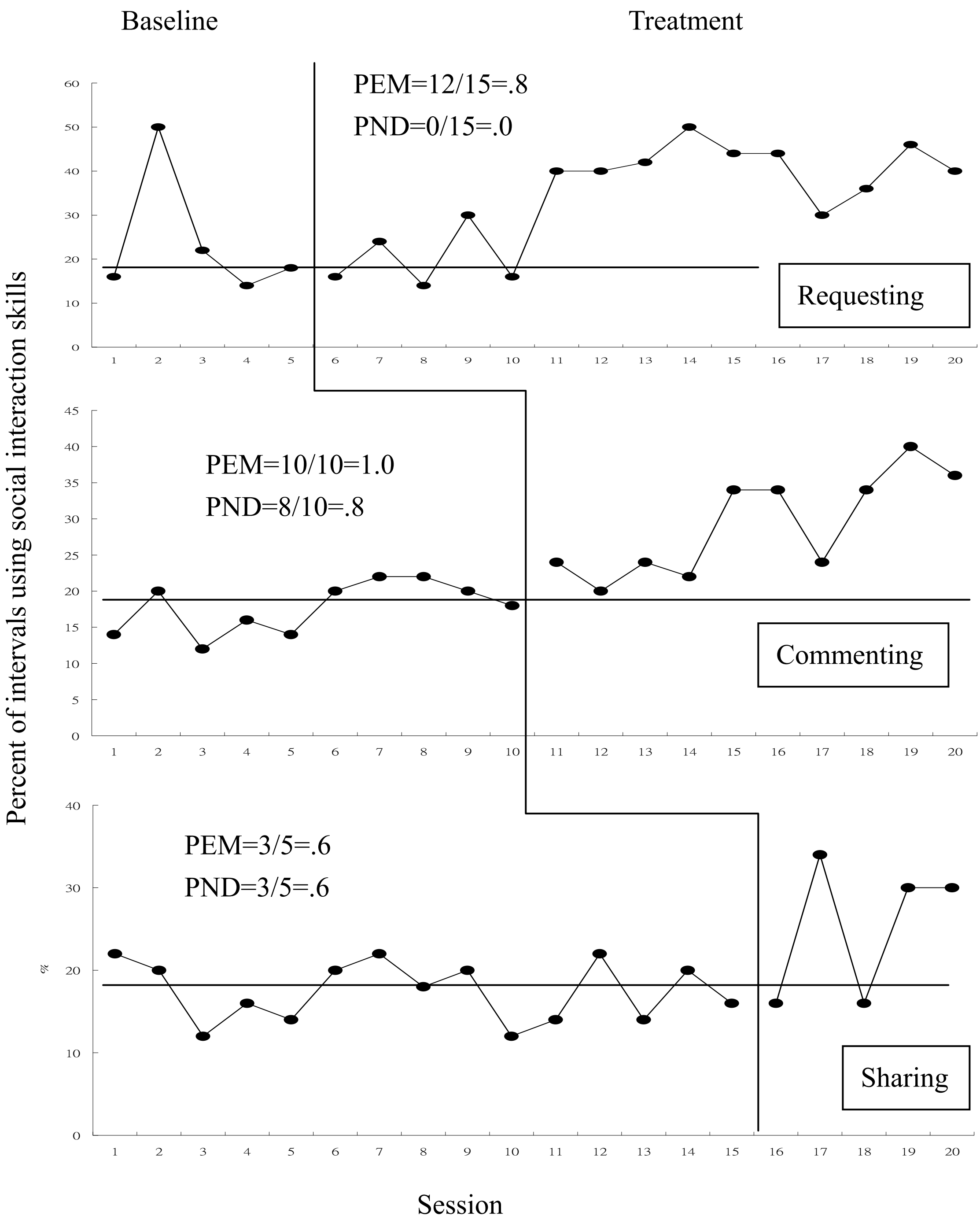
Baseline
Treatment
PEM=12/15=.8
PND=0/15=.0
Requesting
PEM=10/10=1.0
PND=8/10=.8
Commenting
PEM=3/5=.6
PND=3/5=.6
Sharing
Percent of intervals using social interaction skills
Session

Fig. 1. Fabricated data for the purpose of the demonstration of the method of calculation of PEM and PND scores.

Table 1

Comparisons of means of PEM scores with criteria suggested by Scruggs et al. (1986) at each level of effectiveness judged by original authors

Original author(s)' judgment	N (percentage)	Mean	SD	The criterion of Scruggs et al. (1986)
Highly effective	970 (90%)	.93	.18	≧ .90
Moderately effective	78 (7%)	.79	.24	≧ .70 < .90
Questionable or not effective	35 (3%)	.33	.28	< .70

Table 2

Analyses of independent variables, dependent variables, and moderators

Subcategories of variables with coding number	*N*	Mean	*SD*	Mean rank
Independent Variables				
13. Activities	148	0.95	0.12	624
30. Positive reinforcer plus punishment	25	0.93	0.22	620
21. Negative reinforcer	140	0.94	0.16	612
40. Token	190	0.92	0.19	586
43. DRA with token	50	0.91	0.23	565
42. Token plus punishment	18	0.96	0.05	525
22. Withdrawal of positive reinforcer	92	0.88	0.21	509
23. DRA	24	0.86	0.27	503
14. Praise	73	0.89	0.17	490
41. Token plus removal of token	165	0.84	0.30	488
11. Edibles	142	0.83	0.29	470
12. Objects	24	0.85	0.25	458
Total	1091	0.90	0.22	
Dependent variables				
63. Self-stimulation	94	0.92	0.20	589
52. Quantity of academic behaviors	105	0.90	0.24	584
53. Social behaviors	267	0.92	0.17	556
62. Self-injury	71	0.90	0.23	556
61. Problem behaviors	352	0.90	0.22	543
51. Quality of academic behaviors	202	0.85	0.26	494

Moderators				
Settings				
Home	83	0.94	0.14	587
Other place	77	0.91	0.18	544
Institution	328	0.88	0.26	531
School	583	0.90	0.21	531
Interventionists				
Teacher	386	0.92	0.16	551
Parent	47	0.91	0.16	547
Assistant	358	0.89	0.24	539
Specialist	260	0.87	0.26	514
Tutor	20	0.76	0.35	449
Category of participants				
Deaf	6	1.00	0.00	711
Mental patient	34	0.96	0.17	653
Normal intelligence	257	0.92	0.17	548
With autism	402	0.88	0.25	544
Mental retardation	251	0.90	0.21	537
Learning disability	14	0.84	0.28	517
Emotional disorder	42	0.89	0.24	576
ADHD	78	0.87	0.23	508
Age of participant				
7 - <13	432	0.92	0.19	548
< 7	321	0.88	0.24	521
≧18	121	0.87	0.28	518

13 - <16	152	0.88	0.22	502
16 - <18	30	0.90	0.23	500
Order of pair of phases				
Second pair	286	0.91	0.20	577
First pair	805	0.89	0.23	535
Assessment of reinforcer				
Preference test	329	0.90	0.24	571
Decided by author	411	0.90	0.21	546
Functional analysis	179	0.91	0.18	541
Using money	57	0.92	0.16	525
Parent's suggestion	115	0.84	0.28	493
Gender of participant				
Female	308	0.89	0.22	469
Male	666	0.89	0.23	496
Kind of *design*				
Reversal	677	0.91	0.20	564
Multiple	414	0.87	0.25	516

Note. The independent variable "punishment" refers to the punishments other than removal of token.

Appendix

The List of Studies Located in the Meta-Analysis

Alber, S. R., Heward, W. L., & Hippler, B. J. (1999). Teaching middle school students with learning disabilities to recruit positive teacher attention. *Exceptional Children, 65*(2), 253-270.

Alexander, R. N., Corbett, T. F., & Smigel, J. (1976). The effects of individual and group consequences on school attendance and curfew violations. *Journal of Applied Behavior Analysis, 9*, 221-226.

Allen, K. D., Loiben, T., Allen, S. J., & Stanley, R. T. (1992). Dentist-implemented contingent escape for management of disruptive child behavior. *Journal of Applied Behavior Analysis, 25*(3), 629-639.

Baer, A. M., Rowbury T., & Baer D. M. (1973). The development of instructional control over classroom activities of deviant preschool children. *Journal of Applied Behavior Analysis*, 6, 289-298

Baker, M. J. (2000). Incorporating the thematic ritualistic behaviors of children with autism into games: Increasing social play interactions with siblings. *Journal of Positive Behavior Interventions, 2*(2), 66-84.

Blair, K. C., Umbreit, J., & Bos, C. S. (1999). Using functional assessment and children's preferences to improve the behavior of young children with behavioral disorders. *Behavioral Disorders, 24*(2), 151-166.

Blount, R. L., Drabman, R. S., Wilson, N., & Stewart, D. (1982). Reducing severe diurnal bruxism in two profoundly retarded females. *Journal of Applied Behavior Analysis, 15*, 565-571.

Boyajian, A. E., DuPaul, G. J., Handler, M. W., Eckert, T. L., & McGoey, K. E. (2001). The use of classroom-based brief functional analyses with preschoolers at-risk for Attention Deficit Hyperactivity Disorder. *School Psychology Review, 30*(2), 278-293.

Breyer, N. L., & Allen, G. J. (1975). Effects of implementing a token economy on teacher attending behavior. *Journal of Applied Behavior Analysis, 8*, 373-380.

Broden, M. C., Bruce, C., Mitchell, M. A., Carter, V., & Hall, R. V. (1970). Effects of teacher attention on attending behavior of two boys at adjacent desks. *Journal of Applied Behavior Analysis, 3*, 199-203.

Buckley, S. D., & Newchok, D. K. (2005). An evaluation of simultaneous presentation and differential reinforcement with response cost to reduce packing. *Journal of Applied Behavior Analysis, 38*, 405-409.

Calloway, C. J., & Simpson, R. L. (1998). Decisions regarding functions of behavior: scientific versus informal analyses. *Focus on Autism and Other Development Disabilities, 13*(3), 167-175.

Carr, E. G., & Kologinsky, E. (1983). Acquisition of sign language by autistic children Ⅱ: Spontaneity and generalization effects. *Journal of Applied Behavior Analysis, 16*, 297-314.

Carr, E. G., Newsom, C. D., & Binkoff, J. A. (1980). Escape as a factor in the aggressive behavior of two retarded children. *Journal of Applied Behavior Analysis, 13*, 101-117.

Carter, C. M. (2001). Using choice with game play to increase language skills and interactive behaviors in children with autism. *Journal of Positive Behavior Interventions, 3*(3), 131-151.

Carton, J. S., & Schweitzer, J. B. (1996). Use of a token economy to increase compliance during hemodialysis. *Journal of Applied Behavior Analysis, 29*, 111-113.

Chapman, S., Fisher, W., Piazza, C. C., & Kurtz, P. F. (1993). Functional assessment and treatment of life-threatening drug ingestion in a dually diagnosed youth.

Journal of Applied Behavior Analysis, 26, 255-256.

Charlop, M. H. (1983). The effects of echolalia on acquisition and generalization of receptive labeling in autistic children. *Journal of Applied Behavior Analysis, 16, 111-126*

Charlop, M. H., Burgio, L. D., Iwata, B. A., & Ivancic, M. T. (1988). Stimulus variation as a means of enhancing punishment effects. *Journal of Applied Behavior Analysis,* 21, 89-95.

Charlop, M. H., Kurtz, P. F., & Milstein, J. P. (1992). Too much reinforcement, too little behavior: Assessing task interspersal procedures in conjunction with different reinforcement schedules with autistic children. *Journal of Applied Behavior Analysis, 25*, 795-808.

Charlop-Christy, M. H., & Haymes, L. K. (1996). Using obsessions as reinforcers with and without mild reductive procedures to decrease inappropriate behaviors of children with autism. *Journal of Autism and Developmental Disorders, 26*(5), 527-546.

Charlop-Christy, M. H., & Haymes, L. K. (1998). Using objects of obsession as token reinforcers for children with autism. *Journal of Autism and Developmental Disorders, 28*(3), 189-198.

Christensen, L., Young, R. K., & Marchant, M. (2004). The effects of a peer-mediated positive behavior support program on socially appropriate classroom behavior. *Education & Treatment of Children, 27*(3), 199-234.

Clark, H. B., Rowbury, T., Baer, A. M., & Baer, D. M. (1973). Timeout as a punishing stimulus in continuous and intermittent schedules. *Journal of Applied Behavior Analysis, 6*, 443-455.

Clarke, S., Dunlap, G., & Vaughn, B. (1999). Family-centered, assessment-based intervention to improve behavior during an early morning routine. *Journal of*

Positive Behavior Interventions, 1(4), 235-241

Clevenger, T. M., & Graff, R. B. (2005). Assessing object-to-picture and picture-to-object matching as prerequisite skills for pictorial preference assessments. *Journal of Applied Behavior Analysis, 38*, 543-574.

Coles, E. K., Pelham, W. E., Gnagy, E. M., Burrows-MacLean, L., Fabiano, G. A., Chacko, A., Wymbs, B. T., Tresco, K. E., Walker, K. S., & Robb, J. A. (2005). A controlled evaluation of behavioral treatment with children with ADHA attending a summer treatment program. *Journal of Emotional and Behavioral Disorders, 13*(2), 99-112

Corte, H. E., Wolf, M. M., & Locke, B, J. (1971). A comparison of procedures for eliminating self-injurious behavior of retarded adolescents. *Journal of Applied Behavior Analysis, 4*, 201-213.

DaCosta, I., G., Rapoff, M. A., & Goldstein, G. L. (1997). Improving adherence to medication regimens for children with asthma and its effect on clinical outcome. *Journal of Applied Behavior Analysis, 30*, 687-691.

Didden, R., Moor, J. D., & Bruyns, W. (1997). Effectiveness of DRO tokens in decreasing disruptive behavior in the classroom with five multiply handicapped children. *Behavioral Interventions, 12*(2), 65-75.

Doleys, D. M., Wells, K. C., Hobbs, S. A., Roberts, M. W., & Cartelli, L. M. (1976). The effects of social punishment on noncomplicance: *A* comparison with timeout and positive practice. *Journal of Applied Behavior Analysis, 9*, 471-482.

Doty, D. W., McInnis, T., & Paul, G. L. (1974). Remediation of negative side effects of an on-going response-cost system with chronic mental patients. *Journal of Applied Behavior Analysis, 7*, 191-198.

Dunlap, G., & Koegel, R. L. (1980). Motivating autistic children through stimulus variation. *Journal of Applied Behavior Analysis, 13*, 619-627.

Dunlap, G., Koegel, R. L., Johnson, J., & O'Neill, R. E. (1987). Maintaining performance of autistic clients in community settings with delayed contingencies. *Journal of Applied Behavior Analysis, 20*, 185-191.

Dyer, K., Dunlap, G., & Winterling, V. (1990). Effects of choice making on the serious problem behaviors of students with severe handicaps. *Journal of Applied Behavior Analysis, 23*, 515-524.

Edwards, W. H., Magee, S. K., & Ellis, J. (2002). Identifying the effects of idiosyncratic variables on functional analysis outcomes: A case study. *Education & Treatment of Children, 25*(3), 317-330.

Egel, A. L. (1981). Reinforcer variation: Implications for motivating developmentally disabled children. *Journal of Applied Behavior Analysis, 14*, 345-350.

Ellingson, S. A., Miltenberger, R. G., Stricker, J., Galensky, T. L., & Garlinghouse, M. (2000). Functional assessment and intervention for challenging behaviors in the classroom by general classroom teachers. *Journal of Positive Behavior Interventions, 2*(2), 85-97.

Ellis, J., & Magee, S. K. (1999). Determination of environmental correlates of disruptive classroom behavior: Integration of functional analysis into public school assessment process. *Education & Treatment of Children, 22*(3), 291-316.

English, C. L., & Anderson, C. M. (2006). Evaluation of the treatment utility of the analog functional analysis and the structured descriptive assessment. *Journal of Positive Behavior Interventions, 8*(4), 212-229.

Evans, R. D. Y. (2004). *Effects of reciprocal peer tutoring on academic achievement and social interaction of elementary students with emotional-behavioral disorders*. Unpublished doctoral dissertation, North Carolina State University, North Carolina.

Fantuzzo, J. W., & Clement, P. W. (1981). Generalization of the effects of teacher- and self-administered token reinforcers to nontreated students. *Journal of Applied Behavior Analysis, 14,* 435-447.

Favell, J. E., McGimsey, J. F., & Jones, M. L. (1978). The use of physical restraint in the treatment of self-injury and as positive reinforcement. *Journal of Applied Behavior Analysis, 11*, 225-241.

Fisher, W., Piazza, C., Cataldo, M., Harrell, R., Jefferson, G., & Conner, R. (1993). Functional communication training with and without extinction and punishment. *Journal of Applied Behavior Analysis, 26*(1), 23-36.

Foxx, R. M. (1977). Attention training: The use of overcorrection avoidance to increase the eye contact of autistic and retarded children. *Journal of Applied Behavior Analysis, 10*, 489-499.

Foxx, R. M., & Azrin, N. H. (1973). The elimination of autistic self-stimulatory behavior by overcorrection. *Journal of Applied Behavior Analysis, 6*, 1-14.

Galensky, T. L., Miltenberger, R. G., Stricker, J. M., & Garlinghouse, M. A. (2001). Functional assessment and treatment of mealtime behavior problems. *Journal of Positive Behavior Interventions, 3*(4), 211-224.

Garcia, E., Guess, D., & Byrnes, J. (1973). Development of syntax in a retarded girl using procedures of imitation, reinforcement, and modelling. *Journal of Applied Behavior Analysis, 6*, 299-310.

Glynn, E. L. (1970). Classroom applications of self-determined reinforcement. *Journal of Applied Behavior Analysis, 3*, 123-132.

Grace, N. C., Kahng, S., & Fisher, W. W. (1994). Balancing social acceptability with treatment effectiveness of an intrusive procedure: A case report. *Journal of Applied Behavior Analysis, 27*, 171-172.

Greene, R. J., & Hoats, D. L. (1969). Reinforcing capabilities of television distortion. *Journal of Applied Behavior Analysis, 2*, 139-141.

Hall, R. V., Axelrod, S., Tyler, L., Grief, E., Jones, F. C., & Robertson, R. (1978). Modification of *behavior problems in the home with a parent as observer and experimenter. Journal of Applied Behavior Analysis,* 5, 53-64.

Haring, T. G., & Kennedy, C. H. (1990). Contextual control of problem behavior in students with severe disabilities. *Journal of Applied Behavior Analysis, 23*, 235-243.

Harris, S. L., & Wolchik, S. A. (1979). Suppression of self-stimulation: Three alternative strategies. *Journal of Applied Behavior Analysis, 12*, 185-198.

Higgins, J. W., Williams, R. L., & McLaughlin, T. F. (2001). The effects of a token economy employing instructional consequences for a third-grade student with learning disabilities: *A* data-based case study. *Education and Treatment of Children, 24*(1), 99-106

Hinton, L. M., & Kern, L. (1999). Increasing homework completion by incorporating student interests. *Journal of Positive Behavior Interventions, 1*(4), 231-234.

Hobbs, T. R., & Holt, M. M. (1976). The *effects* of token reinforcement on the behavior of delinquents in cottage settings. *Journal of Applied Behavior Analysis, 9,* 189-198.

Hoff, K. E., Ervin, R. A., & Friman, P. C. (2005). Refining functional behavioral assessment: Analyzing the separate and combined effects of hypothesized controlling variables during ongoing classroom *routines*. *School Psychology Review, 34*(1), 45-57.

Hohnhorst, A., & Roberts, M. W. (1992). Evaluation of a brief work chore discipline procedure. *Behavioral Residential Treatment, 7*(1), 55-69.

Houten, R. V., & Rolider, A. (1988). Recreating the scene: An effective way to provide delayed punishment for inappropriate motor behavior. *Journal of Applied Behavior Analysis, 21*, 187-192.

Hutchinson, S. W., Murdock, J. Y., Williamson, R. D., & Cronin, M. E. (2000). Self-recording plus encouragement equals improved behavior. *Teaching Exceptional Children, 32*(5), 54-58.

Ingham, R. J., & Andrews, G. (1973). An analysis of a token economy in stuttering therapy. *Journal of Applied Behavior Analysis, 6,* 219-229.

Iwata, B. A., Pace, G. M., Kalsher, M, J., Cowdery, G. E., & Cataldo, M. F. (1990). Experimental analysis and extinction of self-injurious escape behavior. *Journal of Applied Behavior Analysis, 23*, 11-27.

Jackson, D. A., & Wallace, R. F. (1974). The modification and generalization of voice loudness in a fifteen-year-*old* retarded girl. *Journal of Applied Behavior Analysis, 7*, 461-471.

Johnson, S. P., Welsh, T. M., Miller, L. K., & Altus, D. E. (1991). Participatory management: *Maintaining* staff performance in a university housing cooperative. *Journal of Applied Behavior Analysis, 24,* 119-127.

Jones, K. M., Young, M. M., & Friman, P. C. (2000). Increasing peer praise of socially rejected delinquent youth: *Effects* on cooperation and acceptance. *School Psychology Quarterly, 15*(1), 30-39.

Kamps, D., Wendland, M., & Culpepper, M. (2006). Active teacher participation in functional behavior assessment for students with emotional and behavior *disorders risks* in general education classrooms. *Behavioral Disorders, 31*(2), 128-146.

Kaplan, H., Hemmes, N. S., Motz, P., & Rodriguez, H. (1996). Self-reinforcement and persons with developmental disabilities. *The Psychological Record, 46*(1), 161-178.

Kazdin, A. E., & Klock, J. (1973). The effect of nonverbal teacher approval on student attentive behavior. *Journal of Applied Behavior Analysis, 6*(4), 643-654.

Kelley, M. L., & McCain, A. P. (1995). Promoting academic performance in inattentive children: The relative efficacy of school-home notes with and without response cost. *Behavior Modification, 19*, 357-375.

Kern, L., & Bambara, L., & Fogt, J. (2002). Class-wide curricular modification to improve the behavior of students with emotional or behavioral disorders. *Behavioral Disorders, 27*(4), 317-326.

Kern, L., Ringdahl, J. E., Hilt, A., & Sterling-Turner, H. E. (2001). Linking self-management procedures to functional analysis results. *Behavioral Disorders, 23*(3), 214-226.

Knapczyk, D. R., & Livingston, G. (1973). Self-recording and student teacher supervision: *Variables* within a token economy structure. *Journal of Applied Behavior Analysis, 6*, 481-486.

Knight, M. F., & McKenzie, H. S. (1974). Elimination of bedtime thumbsucking in home settings through contingent reading. *Journal of Applied Behavior Analysis, 7*, 33-38.

Koegel, R. L., & Covert, A. (1972). The relationship of self-stimulation to learning in autistic children. *Journal of Applied Behavior Analysis, 5*, 381-387.

Koegel, R. L., & Rinvover, A. (1974). Treatment of psychotic children in a classroom environment:I. Learning in a large group. *Journal of Applied Behavior Analysis, 7*, 45-59.

Koegel, R. L., Dyer, K., & Bell, L. K. (1987). The influence of child-preferred activities on autistic children's social behavior. *Journal of Applied Behavior Analysis, 20*, 243-252.

Koegel, R. L., Firestone, P. B., Kramme, K. W., & Dunlap, G. (1974). Increasing spontaneous play by suppressing self-stimulation in autistic children. *Journal of Applied Behavior Analysis, 7*, 521-528.

Kotkin, R. (1998). The Irvine paraprofessional program: promising practice for serving students with ADHD. *Journal of Learning Disabilities, 31*(6), 556-564.

Kubany, E. S., & Sloggett, B. B. (1973). Coding procedure for teachers. *Journal of Applied Behavior Analysis*, 6, 339-344.

Laski, K. E., Charlop, M. H., & Schreibman, L. (1988). Training parents to use the natural language paradigm to increase their autistic children's speech. *Journal of Applied Behavior Analysis, 21*, 391-400.

Lerman, D. C., Vorndran, C., Addison, L., & Kuhn, S. A. C. (2004). A rapid assessment of skills in young children with autism. *Journal of Applied Behavior Analysis, 37*, 11-26.

Leung, J., & Wu, K. (1997). Teaching receptive naming of Chinese characters to children with autism by incorporating echolalia. *Journal of Applied Behavior Analysis, 30*, 59-68.

Linscheid, T. R., Iwata, B. A., Ricketts, R. W., Williams, D. E., & Griffin, J. C. (1990). Clinical evaluation of the Self-injurious Behavior Inhibiting System (SIBIS). *Journal of Applied Behavior Analysis, 23*, 53-78.

Lo, Y., & Loe, S. A. (2002). The effects of social skills instruction on the social behaviors of students at risk for emotional or behavioral disorders. *Behavioral Disorders, 27*(4), 371-385.

Lohrmann-O'Rourke, S., & Yurman, B. (2001). Naturalistic assessment of and intervention for mouthing behaviors influenced by establishing operations. *Journal of Positive Behavior Intervention, 3*(1), 19-27.

Losada-Paisey, G., & Paisey, T. J. H. (1988). Program evaluation of a comprehensive treatment package for mentally retarded offenders. *Behavioral Residential Treatment, 3*(4), 247-264.

Luce, S. C., Delquadri, J., & Hall, R. V. (1980). Contingent exercise: *A* mild but powerful procedure for suppressing inappropriate verbal and aggressive behavior. *Journal of Applied Behavior Analysis, 13*, 583-594.

Maag, J. W., & Larson, P. J. (2004). Training a general education teacher to apply functional assessment. *Education and Treatment of Children, 27*(1), 26-36.

Mancina, C., Tankersley, M., Kamps, D., Kravits, T., & Parrett, J. (2000). Brief report: Reduction of inappropriate vocalizations for a child with autism using a self-management treatment program. *Journal of Autism and Developmental Disorders, 30*(6), 599-606.

Matheson, A. S., & Shriver, M. D. (2005). Training teachers to give effective commands: Effects on student compliance and academic behaviors. Sc*hool Psychology Review, 14*(2), 202-219.

McComas, J. J., Goddard, C., & Hoch, H. (2002). The effects of preferred activities during academic work breaks on task engagement and negatively reinforced destructive behavior. *Education and Treatment of Children, 25*(1), 103-112.

McGinnis, J. C., Friman, P. C., & Carlyon, W. D. (1999). The effect of token rewards on "intrinsic" motivation for doing math. *Journal of Applied Behavior Analysis, 32*, 375-379.

McGoey, K. E., & Dupaul, G. J. (2000). Token reinforcement and response cost procedures: Reducing the disruptive behavior of preschool children with attention-deficit- hyperactivity disorder. *School Psychology Quarterly, 15*(3), 330-343.

Milan, M. A., & McKee, J. M. (1976). The cellblock token economy: Token reinforcement procedures in a maximum security correctional institution for adult male felons. *Journal of Applied Behavior Analysis, 9*, 253-275.

Moore, B. L., & Bailey, J. S. (1973). Social punishment in the modification of a pre-school child's "autistic-like" behavior with a mother as therapist. *Journal of*

Applied Behavior Analysis, 6, 497-507.

Musser, E. H., Bray, M. A., Kehle, T. J., & Jenson, W. R. (2001). Reducing disruptive behaviors in students with serious emotional disturbance. *School Psychology Review, 30*(2), 294-304.

Neef, N. A., & Lutz, M. N. (2001). Assessment of variables affecting choice and application to classroom interventions. *School Psychology Quarterly, 16*(3), 239-252.

Nelson, G. I., & Cone, J. D. (1979). Multiple-baseline analysis of a token economy for psychiatric inpatients. *Journal of Applied Behavior Analysis, 12*, 255-271.

Newman, B., Tuntigian, L., Ryan, C. S., & Reineche, D. R. (1997). Self-management of a DRO procedure by three students with autism. *Behavioral Interventions, 12*(3), 149-156.

Noell, G. H., VanDerHeyden, A. M., Gatti, S. L., & Whitmarsh, E. L. (2001). Functional assessment of the effects of escape and attention on students' compliance during instruction. *School Psychology Quarterly, 16*(3), 253-269.

O'Neill, R. E., & Sweetland-Baker, M. (2001). Brief report: An assessment of stimulus generation and contingency effects in functional communication training with two students with autism. *Journal of Autism and Developmental Disorders, 31*(2), 235-240.

O'Reilly, M., Sigafoos, J., Lancioni, G., Edrisinha, C., & Andrews, A. (2005). An examination of the effects of a classroom activity schedule on levels of self-injury and engagement for a child with severe autism. *Journal of Autism and Developmental Disorders, 35*(3), 305-311.

Orr, T. J., Myles, B. S., & Carlson, J. K. (1998). The impact of rhythmic entrainment on a person with autism. *Focus on Autism and Other Developmental Disabilities, 13*(3), 163-166.

Osborne, J. G. (1969). Free-time as a reinforcer in the management of classroom behavior. *Journal of Applied Behavior Analysis, 2*, 113-118.

Pace, G. M., Ivancic, M. T., Edwards, G. L., Iwata, B. A., & Page, T. J. (1985). Assessment of stimulus preference and reinforcer value with profoundly retarded individuals. *Journal of Applied Behavior Analysis, 18*, 249-255.

Packard, R. G. (1970). The control of “classroom attention”: A group contingency for complex behavior. *Journal of Applied Behavior Analysis, 3*, 13-28.

Pelios, L. V., MacDuff, G. S., & Axelrod, S. (2003). The effects of a treatment package in establishing independent academic work skills in children with autism. *Education and Treatment of Children, 26*(1), 1-21.

Pendergrass, V. E. (1972). Timeout from positive reinforcement following persistent, high-rate behavior in retardates. *Journal of Applied Behavior Analysis. 5*, 85-91.

Pfiffner, L. J., & O'Leary, S. G. (1987). The efficacy of all-positive management as a function of the prior use of negative consequences. *Journal of Applied Behavior Analysis*, 20, 265-271.

Phillips, E. L. (1968). Achievement place: Token reinforcement procedures in a home-style rehabilitation setting for "pre-delinquent" boys. *Journal of Applied Behavior Analysis, 1*, 213-223.

Phillips, E. L., Phillips, E. A., Fixsen, L. D., & Wolf, M. M. (1971). Achievement place: *M*odification of the behaviors of pre-delinquent boys within a token economy. *Journal of Applied Behavior Analysis, 4*, 45-59.

Pierce, C. H., & Risley, T. R. (1974a). Improving job performance of neighborhood youth corps aides in an urban recreation program. *Journal of Applied Behavior Analysis, 7*, 207-215.

Pierce, C. H., & Risley, T. R. (1974b). Recreation as a reinforcer: Increasing membership and decreasing disruptions in an urban recreation center. *Journal of*

Applied Behavior Analysis, 7, 403-411.

Plummer, S., Baer, D. M., & LeBlanc, J. M. (1977). Functional considerations in the use of procedural timeout and an effective alternative. *Journal of Applied Behavior Analysis, 10*, 689-705.

Reese, R. M., Sherman, J. A., & Sheldon, J. B. (1998). Reducing disruptive behavior of a group-home resident with autism and mental retardation. *Journal of Autism and Developmental Disorders, 28*(2), 159-165.

Rehfeldt, R. A., & Chambers, M. C. (2003). Functional analysis and treatment of verbal perseverations displayed by an adult with autism. *Journal of Applied Behavior Analysis, 36*, 259-261.

Reid, D. H., Parsons, M. B., Phillips, J. F., & Green, C. W. (1993). Reduction of self-injurious hand mouthing using response blocking. *Journal of Applied Behavior Analysis, 26*, 139-140.

Rincover, A., Cook, R., Peoples, A., & Packard, D. (1979). Sensory extinction and sensory reinforcement principles for programming multiple adaptive behavior change. *Journal of Applied Behavior Analysis, 12*, 221-233.

Robinson, P. W., Newby, T. J., & Ganzell, S. L. (1981). A token system for a class of underachieving hyperactive children. *Journal of Applied Behavior Analysis, 14*, 307-315.

Roca, J. V. & Gross, A. M. (1996). Report-Do-Report: Promoting setting and setting-time generalization. *Education & Treatment of Children, 19*(4), 408-424.

Rolider, A., & Houten, R. V. (1985). Movement suppression time-out for undesirable behavior in psychotic and severely developmentally delayed children. *Journal of Applied Behavior Analysis, 18*, 275-288.

Rowbury, T. G., Baer, A. M., & Baer, D. M. (1976). Interactions between teacher guidance and contingent access to play in developing preacademic skills of deviant preschool children. *Journal of Applied Behavior Analysis, 9*, 85-104.

Russo, D. C., & Koegei, R. L. (1977). A method for integrating an autistic child into a normal public-school classroom. J*ournal of Applied Behavior Analysis, 10*, 579-590.

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.

Sasso, G. M., Reimers, T. M., Cooper, L. J., Wacker, D., Berg, W., Steege, M., Kelly, L., & Allaire, A. (1992). Use of descriptive and experimental analyses to identify the functional properties of aberrant behavior in school settings. *Journal of Applied Behavior Analysis, 25*, 809-821.

Schilling, D. L., & Schwartz, I. S. (2004). Alternative seating for young children with autism spectrum disorder: Effects on classroom behavior. *Journal of Autism and Developmental Disorders, 34*(4), 423-432.

Schutte, R. C., & Hopkins, B. L. (1970). The effects of teacher attention on following instructions in a kindergarten class. *Journal of Applied Behavior Analysis, 3*, 117-122.

Smith, B. W., & Sugai, G. (2000). A self-management functional assessment-based behavior support plan for a middle school student with EBD. *Journal of Positive Behavior Interventions, 2*(4), 208-217.

Smith, L. K. C., & Fowler, S. A. (1984). Positive peer pressure: *The* effects of peer monitoring on children's disruptive behavior. *Journal of Applied Behavior Analysis, 17*, 213-227.

Solnick, J. V., Rincover, A., & Peterson, C. R. (1977). Some determinants of the reinforcing and punishing effects of timeout. *Journal of Applied Behavior Analysis, 10*, 415-424.

Swain, J. & McLaughlin, T. F. (1998). The effects of bonus contingencies in a classwide token program on math accuracy with middle-school students with behavioral disorders. *Behavioral Interventions, 13*, 11-19.

Tanner B. A. & Zeiler M. (1975). Punishment of self-injurious behavior using aromatic ammonia as the aversive stimulus. *Journal of Applied Behavior Analysis*, 8, 53-57.

Tessing, J. L., Napolitano, D. A., McAdam, D. B., DiCesare, A., & Axelrod, S. (2006). The effects of providing access to stimuli following choice making during vocal preference assessments. *Journal of Applied Behavior Analysis, 39*, 501-506.

Tiger, J. H., Hanley, G. P., & Bessette, K. K. (2006). Incorporating descriptive assessment results into the design of a functional analysis: A case example involving a preschooler's hand mouthing. *Education and Treatment of Children, 29*(1), 107-124.

Truchlicka, M., McLaughlin, T. F., & Swain, J. C. (1998). Effects of token reinforcement and response cost on the accuracy of spelling performance with middle-school special education students with behavior disorders. *Behavioral Interventions, 13*, 1-10.

VanDerHeyden, A. M., Witt, J. C., & Gatti, S. (2001). Descriptive assessment method to reduce overall disruptive behavior in a preschool classroom. *School Psychology Review, 30*(4), 548-567.

Wahler, R. G. (1969). Setting generality: Some specific and general effects of child behavior therapy. *Journal of Applied Behavior Analysis, 2*, 239-246.

Wasik, B. H., Senn, K., Welch, R. H., & Cooper, B. R. (1969). Behavior modification with culturally deprived school children: Two case studies. *Journal of Applied Behavior Analysis, 2*, 181-194.

Watanabe, M., & Sturmey, P. (2003). The effect of choice-making opportunities during activity schedules on task engagement of adults with autism. *Journal of Autism and Developmental Disorders, 33*(5), 535-538.

Weiner, R. K., Sheridan, S. M., & Jenson, W. R. (1998). The effects of conjoint behavioral consultation and a structured homework program on math completion and accuracy in junior high students. *School Psychology Quarterly, 13*(4), 281-309.

Wells, K. C., Forehand, R., Hickey, K., & Green, K. D. (1977). Effects of a procedure derived from the overcorrection principle on manipulated and nonmanipulated behaviors. *Journal of Applied Behavior Analysis, 10*, 679-687.

Wilder, D. A., Chen, L., Atwell, J., Pritchard, J., & Weinstein, P. (2006). Brief functional analysis and treatment of tantrums associated with transitions in preschool children. *Journal of Applied Behavior Analysis, 39*, 103-107.

Wilder, D. A., Normand, M., & Atwell, J. (2005). Noncontingent reinforcement as treatment for food refusal and associated self-injury. *Journal of Applied Behavior Analysis, 38*, 549-553.

Williams, J. A., Koegel, R. L., & Egel, A. L. (1981). Response-reinforcer relationships and improved learning in autistic children. *Journal of Applied Behavior Analysis, 14*, 53-60.

Wilson, G. T., Leaf, R. C., & Nathan, P. E. (1975). The aversive control of excessive alcohol consumption by chronic alcoholics in the laboratory setting. *Journal of Applied Behavior Analysis, 8*, 13-26.

Winkler, R. C. (1970). Management of chronic psychiatric patients by a token reinforcement system. *Journal of Applied Behavior Analysis, 3*, 47-55.

Wolfe, B. D., Dattilo, J., & Gast, D. (2003). Effects of a token economy system within the context of cooperative games on social behaviors of adolescents with emotional and behavioral disorders. *Therapeutic Recreation Journal, 37*(2), 124-141.

Wood, R., & Flynn, J. M. (1978). A self-evaluation token system versus an external evaluation token system alone in a residential setting with predelinquent youth. *Journal of Applied Behavior Analysis, 11*, 503-512.

Zlomke, K., & Zlomke, L. (2003). Token economy plus self-monitoring to reduce disruptive classroom behaviors. *The Behavior Analyst Today, 4*(2), 177-143.