Title: Improving Intervention Implementation and Fidelity in Evidence-Based Practice:

Integrating Teacher Preference into Intervention Selection

Author(s): Joseph H. Wehby, Daniel M. Maggin, LeAnne Johnson, & Frank J. Symons

Abstract Body
Limit 5 pages single spaced.

Background/context:
There is a recognized need to improve the efficiency with which evidence-based interventions are brought to scale in educational settings (Durlak & DuPree, 2008; Gresham, 2004; Han & Weiss, 2005; Smith, Daunic, & Taylor, 2007). Although recent efforts to address the diffusion of evidence-based practices within education has focused on their adoption at a systems level (e.g., Sugai & Horner, 2008), research has shown that implementation at the classroom level continues to be a major challenge (Fagan, Hanson, Hawkins, & Arthur, 2008; Fairbanks, Simonsen, & Sugai, 2008; Kincaid, Childs, Blase, & Wallace, 2007). Unfortunately, the fundamental problem of how to increase the use of effective programs is rarely addressed in a systematic way within school-based research. It is true that substantial research has been directed toward characterizing barriers to diffusion (and, indirectly, implementation) but there remain very few experimental or quasi-experimental studies directly targeting diffusion, particularly implementation variables in the delivery of evidence-based programs (Pentz, 2004).

In school settings, teachers function as the front-line implementers of innovative programs. Methods designed to increase the use of evidence-based practices in schools have traditionally relied on teacher training (McCormick et al., 1995; Perry, Murray, & Griffin, 1990) and / or consultant delivered performance feedback (Mortenson & Witt, 1998; Noell et al., 1997; Witt et al., 1997; Jones et al., 1997). However, neither of these methods has been associated with sustained implementation by teachers. A potential limitation of traditional approaches to increase the implementation and sustainability of research-based practices is the failure to consider teacher preferences. Teacher preference has been discussed in previous educational research exploring treatment acceptability and the association between it and teachers' (a) willingness to implement (Broughton & Hester, 1993), (b) fidelity of implementation (Sterling-Turner, et al. 2002), and treatment effects (Remiers, Waker, & Koeppl, 1987). But, educational research has lagged behind research in clinical psychology and healthcare which have provided several direct demonstrations that research participants show enhanced fidelity with a treatment protocol when allowed to express a preference between multiple treatment options (Janevic, et al., 2003; Ward, et al., 2000; Wills & Holmes-Rovner, 2006). The purpose of this study, therefore, was to conduct an experiment in which teacher expressed preference was the independent variable and examined in relation to a set of implementation outcomes. Specifically, we used a randomized design to evaluate whether providing teachers the opportunity to select among two different emerging evidence-based practices (the Good Behavior Game & teacher self-monitoring - described in detailed below) resulted in improved implementation and fidelity.

Purpose / objective / research question / focus of study:
The purpose of the present study was to the effect that teacher choice of intervention has on their level of procedural implementation and quality of implementation. The following research questions helped to guide the study:
1. Do teachers randomly assigned to the intervention "choice" group have higher ratings of procedural fidelity than teachers assigned to the "no choice" group at any of three different time points?

2. Do teachers randomly assigned to the intervention "choice" group have higher ratings of implementation quality than teachers assigned to the "no choice" group at any of three different time points?

3. Are teachers randomly assigned to the intervention "choice" group more likely to adopt the intervention at any of three different time points?

Setting:
Participants were identified from 14 schools distributed across three large metropolitan school districts from three regions of the United States. Districts ranged in size from 23,200 – 70,140 enrolled students with an average racial demographic of 24% White, 58% Black, 12% Hispanic, 4 % Asian, and 2% American Indian. Across all three districts, an average of 12% of students had limited English proficiency, 15% received special education services, and 69% qualified for free or reduced lunch. Teachers from each district were invited to participate given their expressed interest in learning new strategies to address the disruptive behavior of their students.

Population / Participants / Subjects:
A total of 69 teachers (88% female; 68% general education, 32% special education) working with kindergarten through 6th grade students participated in this study. Participating teachers averaged 12 years of teaching experience ($SD = 9.5$) with 44% having obtained Master's degrees or higher, 33% of teachers were Black, 58% White, 2% Hispanic, and 7% did not report their racial background. Teachers were trained to implement all components of the intervention in their classrooms during their typical language arts instruction. All observations (described below) of teachers' instructional behavior occurred in the classroom during language arts instruction.

Intervention / Program / Practice:
Two behaviorally-based interventions were used as the basis for teacher preference and treatment selection. The two interventions were from a larger ongoing parent study on reducing severe behavior problems in schools by changing classroom ecologies. Both interventions have sufficient evidence (peer-reviewed replicated results obtained from experimental and quasi-experimental research designs across independent research groups) to be considered consistent with evidence-based practice guidelines.

Good Behavior Game. The Good Behavior Game (GBG: Barrish, Saunders, & Wold, 1969; Kellam, Ling, Merisca, Brown, & Ialongo, 1998) is a group-contingency classroom management procedure designed to reduce problem behavior in the classroom. Research has documented the effectiveness of the GBG in decreasing levels of aggression and disruption as well as increasing on-task behavior during instructional times (Dolan et al, 1993; Harris & Sherman, 1973; Ialongo et al., 1999; Kellam, et al., 1998; Medland & Staknik, 1972). GBG is based on implementation of explicit rules as well as explicit and consistent teacher responses to students' rule following and rule violating behavior. For this study, teachers were trained to implement the GBG during 10 minutes of their typical classroom routine. This short duration was identified to, 1) create discrete opportunities for each teacher to focus on very systematic and consistent responding to students' behavior, and 2) allow the behavior consultant to have precise opportunities to deliver

feedback to the teachers about the observed schedules of reinforcement that may be influencing the efficacy of a teacher's efforts to elicit appropriate student behaviors.

Teacher Self Monitoring. Teacher self-monitoring (TSM) of their instruction focused on two critical behaviors: the frequency of praise statements and opportunities to respond (OTRs) embedded in their teaching. There is reliable evidence that the increased use of praise and OTRs increases student engagement and achievement (Sutherland & Wehby, 2001). Likewise, self-monitoring has proven to be an effective method of increasing the use of targeted teacher instructional behaviors (Hoover & Carroll, 1987; Kilbourn, 1991; Sutherland, Alder, & Gunter, 2003; Sutherland & Wehby, 2001b; Sutherland, Wehby, & Copeland, 2000). For this study, the TSM intervention was based on procedures originally described by Sutherland and Wehby (2001). Teachers were trained to use a microcassette recorder to audio tape 15 minutes of their language arts instruction three days per week. With each recording, teachers were trained to monitor their use of praise and opportunities given to students to respond to instruction by listening to a five minute sample from the tape, tallying occurrences of each behavior, and graphing their performance. TSM is an effective method of increasing the use of targeted teacher instructional behaviors and improving student outcomes (Hoover & Carroll, 1987; Kilbourn, 1991; Sutherland, Alder, & Gunter, 2003; Sutherland & Wehby, 2001b; Sutherland, Wehby, & Copeland, 2000).

Assignment to Preference vs. No Preference Groups for Intervention Implementation
Each participating teacher was assigned randomly to one of three groups: 'preference group' (P; N = 25), 'no preference -GBG' (NP-GBG; N = 21), or 'no preference-TSM' (NP-TSM; N = 23). Random assignment occurred at the classroom level using a random number generator in which teachers with odd numbers were assigned to preference and even numbers were assigned to no preference. Teachers randomly assigned to the 'preference group' were given a choice between implementing GBG or TSM. Twenty-two (88%) selected the GBG to implement. Consequently, all of the following procedural description and subsequent analyses of the effect of preference on implementation are based only on the participants that implemented GBG either because they self-selected the procedure (P-GBG) or were assigned to it (NP-GBG).

Research Design:
The design used for the present study was a randomized field trial .Teachers were randomly assigned to either the "choice" or "no choice" condition. Members of the "choice" group were asked to select one of the two evidence-based practices while members of the "no choice" group were assigned to one of the two evidence-based practices. Random assignment was conducted using a random number table. Each teacher was randomly assigned a number. Those teachers with even numbers were assigned to the "no choice" group and those with odd numbers were assigned to the "choice" group.

Data Collection and Analysis:
Direct observation of teacher's implementation was monitored at three pre-assigned observation intervals; initial week of implementation, immediately post six weeks of consultation, and at a four-week follow-up. During the six weeks of consultation, observations of implementation of the GBG occurred each week to guide feedback and support for implementation. Implementation was evaluated using three different variables at each of the three pre-specified observation

intervals, 1) percent of procedural items, 2) quality of implementation, and 3) number of actual implementers.

Percent of procedural items. Teachers were trained to implement the intervention procedure in direct adherence to the procedures listed on the fidelity checklist. The GBG fidelity checklist consisted of 14 procedural including getting the students ready for the start of the game by explicitly reviewing team membership, rules, and criteria for receiving the reward at the end of the game, procedures for responding to rule violations, and procedures for providing an explicit end to the game with delivery of feedback and rewards for students. Each procedural item was dichotomously coded as observed or not observed. A column was also included for documenting if the teacher provided other evidence for completion of the procedures there were not directly observed by the behavior consultant. For analysis, these columns were collapsed such that affirmation that a specific item was either directly observed or the teacher provided other evidence were counted as completion of that procedural item when summarized into an overall percent of procedural items implemented for the week. Four procedural items could only be coded if disruptive behavior occurred while the intervention was being implemented. If no disruptive behavior was observed, those items were not coded and the total number of items used to calculate the percent of procedural items implemented for the week was reduced to 10.

Quality of implementation. For each procedural item observed, a quality of implementation score was also assigned to quantify variability in the degree to which individual teachers implemented each procedure. For example, one procedural item on the GBG checklist states, "Refer to requirements to win." For this item, if the teacher made any reference to how teams win the game, the dichotomous coding that was used to calculate the percent of procedural items implemented would reflect that that particular item was implemented. However, there are quantitative and qualitative differences when a teacher simply states, "you all remember how to win" versus stating, "when the timer goes off, if your team has less than three checks on the board, your whole team will win for the day and your names go up on the leader board." To quantify this variability in implementation, each procedural item was assigned one of five possible scores on a scale of 0 – 100 percent. Zero represented that the item was not implemented at all, 25% represented minimal quality of implementation with significant room for improvement, 50% represented half/partial quality with some room for improvement, 75%, represented good quality with only minimal room for improvement, and a score of 100% represented that the item was implemented with the highest possible quality. For analysis, individual teacher's quality scores across all procedural items were summed and then divided by 14 to obtain a mean quality of implementation score. As with the percent of procedural items variable, if the four procedural items that were dependent on the occurrence of disruptive behavior were not scored by the observer because disruptive behavior did not occur during the observation period, the total items were reduced to 10 when calculating the mean quality of implementation score.

Number of actual implementers. During each observation interval, the behavior consultant conducted a general assessment of whether or not the intervention was being implemented. If the intervention was not being implemented, the consultant documented the reason as per teacher report. If the intervention was being implemented, the consultant documented the frequency and actual duration of each implementation as reported by the teacher. For analysis, the number of

teachers that were actually implementing the intervention was summed to provide a total number of actual implementers at each of the three pre-determined observation intervals.

Data Analysis

Repeated measures analysis of variance (RM-ANOVA) was used to examine overall differences between the P-GBG and NP-GBG groups on percent of procedural items implemented and implementation quality across the three observation intervals. Planned post hoc analyses of variance (ANOVA) were conducted at each time point to further examine differences between the groups. Chi-square tests were conducted for each observation interval to test for any differences in the number of teachers actually implementing the intervention.

Findings / Results:

Data gathered across each time point (initial exposure, post 6-week consult, and 4-week follow up) were screened for missing values, normality, and homogeneity of variance, within each time point. Screening of the data prior to the analysis confirmed that implementation data were missing at the third observation interval for one case due to a maternity leave. This case was excluded from subsequent analyses. Figure 1 provides a visual display of the mean performance of both groups on each of three implementation variables over time.

Percent of Procedural Items

Repeated measures ANOVA was used to examine within and between group differences in percent of procedural items implemented across time (Figure 1, Panel 1a). Distributions for both the P-GBG and NP-GBG groups met the assumptions of normality with nominal negative skews for the P-GBG group (range of -.81 to -1.9) and homogeneity of variance over time based on Levene's Test (F (1, 40) range = 1.33 – 2.58, p range = .12 - .26). The omnibus test resulted in statistically significant main effects for group membership (P-GBG/NP-GBG), F (1, 40) = 5.56, p = .02, partial η^2 = .12 with observed power of .63, and time, F (2, 39) = 6.3, p < .01, partial η^2 = .24 with observed power of .87. A statistically significant interaction between group and time was not observed (F (2, 80) = 1.43, p = .25, partial η^2 = .03). Planned comparisons of the differences in the estimated means between groups at each time point were conducted using three univariate ANOVAs. At the initial observation, the P-GBG group implemented a significantly higher percent of the procedural items compared to the NP-GBG group (P-GBG: X = 70%, SD = 41%; NP-GBG: X = 35%, SD = 43%; F (1, 41) = 7.31, p = .01, d = .83). There were no significant differences immediately following the six-week consultation (P-GBG: X = 77%, SD = 28%; NP-GBG: X = 66%, SD = 39%; F (1, 42) = 1.23, p = .27, d = .33). At the final follow up, the P-GBG group again demonstrated significantly higher levels of fidelity than the NP-GBG group (P-GBG: X = 64%, SD = 40%; NP-GBG: X = 36%, SD = 42%; F (1, 39) = 4.85, p = .03, d = .68).

Quality of Implementation

Repeated measures ANOVA was used to examine within and between group differences in implementation quality across time (Figure 1, Panel 1b). The distributions for both groups met the assumptions of normality and homogeneity of variance based on Levene's Test (F (1, 40) range = .007 – 1.132, p range = .29 - .94). The omnibus test results were similar to those examining level of fidelity, with statistically significant main effects for group membership (P-GBG/NP-GBG), F (1, 40) = 5.46, p = .02, partial η^2 = .12 with an observed power of .63, and

time, F (2, 39) = 8.69, $p < .01$, partial $\eta^2 = .31$ with observed power of .96. A statistically significant interaction between group and time was not observed (F (2, 38) = 1.95, p = .17, partial $\eta^2 = .09$). Planned comparisons of the differences in the estimated means between groups at each time point were again conducted using three univariate ANOVAs. Consistent with findings for the percent of procedural items implemented, the P-GBG group implemented with a higher quality of fidelity at the initial observation (P-GBG: $X = 65\%$, $SD = 39\%$; NP-GBG: $X = 32\%$, $SD = 39\%$; F (1, 41) = 7.68, $p = < .01$, $d = .85$). There were no significant differences immediately following the 6-week consultation (P-GBG: $X = 67\%$, $SD = 28\%$; NP-GBG: $X = 58\%$, $SD = 34\%$; F (1, 42) = .92, $p = .34$, $d = .82$). At the final four-week follow up, teachers in the P-GBG group were again implementing with a significantly higher quality of fidelity (P-GBG: $X = 52\%$, $SD = 36\%$; NP-GBG: $X = 27\%$, $SD = 33\%$; F (1, 39) = 5.28, $p = .03$, $d = .53$).

Number of Actual Implementers
To determine if preference and self-selection was related to the number of actual implementers of the intervention at each of the three time points, a series of Chi-Square comparisons between the P-GBG and NP-GBG groups were conducted (Figure 1, Panel 1c). Significant differences were observed between groups at the initial observation session ($\xi^2 = 5.32$, $p = .02$, phi = .35), with 77% of the teachers in the P-GBG group implementing, compared to 43% of the NP-GBG group. No significant difference in the number of implementers between groups was found at the post 6-week consultation observation ($\chi^2 = .89$, $p = .31$, phi = .14) with 91% of the P-GBG group implementing and 81% of the NP-GBG group implementing. At the final follow up observation conducted four weeks later, there was a marginally statistically significant difference between groups demonstrating actual implementation at proportions similar to the initial observation (P-GBG = 76%, NP-GBG = 48%) ($\chi^2 = 3.64$, $p = .06$, phi = .29).

Conclusions:
This study provides preliminary evidence to suggest that offering teachers a choice of interventions might lead to higher implementation and sustained use of evidence-based practices. Teacher preference may be a vehicle in which to increase teacher motivation to comply with intervention protocols. This phenomenon has, in fact, been observed in the health care industry (Bradley, 1993; Janevic, et al. 2003; Wills & Holmes-Rovner, 2006). According to these health researchers, preference-driven trials may provide an opportunity to systematically explore process variables that may impact the diffusion of EBPs on a larger scale. The findings reported here provide evidence that one particular tailoring variable, intervention preference, was related to higher degrees of initial and sustained fidelity and quality of implementation as well as greater numbers of actual implementers across all participants. Future research is needed to examine the impact of a priori decisions to adapt either the actual EBP or the process by which the EBP is disseminated for adoption and implementation based on other tailoring variables (i.e. participant functioning, social-emotional needs, perceived benefit, outcome expectations, intensity/severity of targeted behaviors). In addition, future work should continue to explore the range of variables effecting teachers preferences for one intervention over another and whether those same variables similarly affect the quality of implementation. Designing studies to do so may be more difficult than 'first order' efficacy studies, but the promise of reducing the research to practice gap may more fully be realized the sooner we undertake the challenge.

Appendices
Not included in page count.

Appendix A. References
References are to be in APA version 6 format.

Abikoff, H. (2001). Tailored psychological treatments for ADHD: The search for a good fit.

    Journal of Clinical Child Psychology, 30, 122-125.

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper

    Saddle River, NJ: Prentice Hall.

August, G.J., Gewirtz, A., & Realmuto, G.M. (in press). Moving the field of prevention science

    to service: Integrating evidence-based preventative interventions into community practice

    through adapted and adaptive models. *Applied and Preventative Psychology*.

Barrera, M. & Castro, F.G. (2006). A heuristic framework for the cultural adaptations of

    interventions. *Clinical Psychology: Science and Practice, 13*, 311-316.

Berman, P., McLaughlin, M., Bass, G., Pauley, E., & Zellman, G. (1977). Federal programs

    supporting education change, Vol. VII: Factors affecting implementation and

    continuation. (Report No. R-15879/7-HEW). Santa Monica, CA: The Rand Corporation

    (ERIC Document Reproduction Service No. 140 432).

Bradley, C. (1993). Designing medical and educational intervention studies. *Diabetes Care, 16*,

    509–518.

Broughton, S.F., & Hester, J.R. (1993). Effects of administrative and community support on

    teacher acceptance of classroom interventions, *Journal of Educational & Psychological*

    Consultation, 4, 169-178.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression /*

correlation analysis for the behavioral sciences (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Collins, L.M., Murphy, S.A., & Bierman, K.L. (2004). A conceptual framework for adaptive preventative interventions. *Prevention Science, 5*, 185-196.

Cook, B.G. & Schirmer, B.R. (2003). What is special about special education? Overview and analysis. *Journal of Special Education, 37*, 200-205.

Durlak, J.A. & Dupree, E.P (2008). Implementation matters: A review of the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327-350.

Evers, W.J., Brouwers, A., & Tornic, W. (2002). Burnout and self-efficacy: A student on teachers' beliefs when implementing and innovative educational system in the Netherlands. *British Journal of Educational Psychology, 72*, 227-244.

Fagan, A.A., Hanson, K., Hawkins, J.D., & Arthur, M.W. (2008). Bridging Science to Practice: Achieving Prevention Program Implementation Fidelity in the Community Youth Development Study. American *Journal of Community Psychology, 41,* 235-249.

Fairbanks, S., Simonsen, B., & Sugai, G. (2008). Classwide secondary and tertiary tier practices and systems. *Teaching Exceptional Children, 40*, 44-52.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association, 32,* 675-701.

Fidell, L.S. & Tabachnick, B.G. (2003) Preparatory data analysis. In Schinka, J.A., Velicer, W.F, & Weiner, I.B. (Eds.). Handbook of Psychology: Research Methods in Psychology (Volume 2). Hoboken, NJ: John Wiley & Sons.

Gable, R.A., Hendrickson, J.M., & Van Acker, R. (2001). Maintaining integrity of FBA-based

interventions in schools. *Education and Treatment of Children, 24*, 248-260.

Gresham, F.M. (2005) Response to intervention: An alternative means of identifying students as

emotionally disturbed. *Education and Treatment of Children, 28*, 328-344.

Gresham, F.M. & Gansle, K.A. (1993). Treatment integrity of school-based behavioral

intervention studies : 1980-1990. *School Psychology Review, 22*, 254-273.

Guskey, T.R. (1988) Teacher efficacy, self-concept, and attitudes toward the implementation of

instructional innovation. *Teaching and Teacher Education, 4*, 63-69.

Janevic, M.R., Janz, N.K., Dodge, J.A., Lin, X., Pan, W., Sinco, B.R., & Clark, N.M. (2003).

The role of choice in health education intervention : a review and case study. *Social

Science and Medicine, 56*, 1581-1594.

Kealey, K.A., Peterson, A.V., Gaul, M.A., & Dihn, K.T. (2000). Teacher training as a behavior

change process: Principles and results from a longitudinal study. *Health Education and

Behavior, 27*, 64-81.

Kincaid, D., Childs, K., Blase, K.A., & Wallace, F. (2007). Identifying barriers and facilitators

in implementing schoolwide positive behavior support. *Journal of Positive Behavior

Interventions, 9*, 174-184.

Lane, K.L., Bocian, K.M., MacMillan, D.L., Gresham, F.M. (2004). Treatment integrity : An

essential –but often forgotten—component of school-based interventions. *Preventing

School Failure, 48*, 36-43.

Malecki C. K., & Elliott S. N., (2002). Children's Social Behaviors as Predictors of Academic

Achievement: A Longitudinal Analysis. *School Psychology Quarterly*. Guilford

Publications.

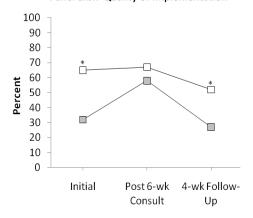Morrison, D.M., Hoppe, M.J., Gillmore, M.R., Kulver, C., Higa, D., & Wells, E.A. (2009). Replicating an intervention: The tension between fidelity and adaptation. *AIDS Education and Prevention, 21*, 128-140.

Reiners, T.M, Wacker, D.P., & Koeppl, G. (1987). Acceptability of behavioral interventions: A review of the literature. *School Psychology Review, 16*, 212-227.

Rogers, E. (2003). *Diffusion of Innovations, 5th Edition*. Free Press: New York, NY.

Smith, T., Scahill, L., Dawson, G., Guthrie, D., Lord, C., Odom, S., Rogers, S., & Wagner, A. (2007). Designing research studies on psychosocial interventions in Autism. *Journal of Autism and Developmental Disorders, 37*, 354 – 366.

Smith, S.W., Daunic, A.P., & Taylor, G.G. (2007). Treatment fidelity in applied educational research: Expanding the adoption and application of measures to ensure evidence-based practice. *Education and Treatment of Children, 30*, 121-134.

Stein, M.K., & Wang, M.C. (1988). Teacher development and school improvement: The process of teacher change. *Teaching and Teacher Education, 8*, 5-35.

Sterling-Turner, H.E., & Watson, T.S., & Moore, J.W. (2002). The effects of direct training and treatment integrity on treatment outcomes in school consultation. *School Psychology Quarterly, 17*, 47-77.

Sugai, G. & Horner, R.H. (2008). What we know and need to know about preventing problem behavior in schools. *Exceptionality, 16*, 67-77.

Sutherland, K. S., & Wehby, J. H. (2001). Exploring the Relationship Between Increased Opportunities to Respond to Academic Requests and the Academic and Behavioral Outcomes of Students with EBD. *Remedial and Special Education, 22*, 113-125.

Tapp, J., Wehby, J. & Ellis. (1995) . A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, & Computers,27,* 25-32.

Von Brock, M.B., & Elliott, S.N. (1987). The influence of treatment effectiveness information on the acceptability of classroom interventions. *Journal of School Psychology, 25*, 131-144.

Walker, B., Cheney, D., Stage, S., & Blum, C. (2005). Schoolwide screening and positive behavior supports: Identifying and supporting students at-risk for school failure. Journal of Positive Behavior Interventions, 7, 194-204.

Ward, E., King, M., Lloyd, M., Bower, P., Sibbald, B., Farrelly, S., Gabbay, M., Tarrier, N., & Addington-Hall, J. (2000). Randomised controlled trial of non-directive counseling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. British Medical Journal, 321, 1383-1388.

Warren, S.F., Fey, M.E., & Yoder, P.J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities, 13*, 70-77.

Weisz, J.R., Donenberg, G.R., Han, S.S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*, 688-701.

Weiss, B. & Weisz, J.R. (1995). Relative effectiveness of behavioral and nonbehavioral child psychotherapy. *Journal of Consulting and Clinical Psychology, 63*, 317-320.

Wills, C.E. & Holmes-Rovner, M. (2006). Integrating decision making and mental health interventions research: Research directions. *Clinical Psychology: Science and Practice, 13*, 9-25.

Appendix B. Tables and Figures

Figure 1. Mean percent of procedural items implemented (Panel 1.a.), quality of implementation (Panel 1.b.) and actual implementers (Panel 1.c.) by group across three observation intervals. An * above a marker on the graph indicates a statistically significant difference between groups.