5-2016

Effects of a Self-Management Procedure Using Student Feedback on Staff Members' Use of Praise in an Out-of-School Time Program

Cade T. Charlton
Utah State University

EFFECTS OF A SELF-MANAGEMENT PROCEDURE USING STUDENT FEEDBACK ON STAFF MEMBERS' USE OF PRAISE IN AN OUT-OF-SCHOOL TIME PROGRAM

by

Cade T. Charlton

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Disability Disciplines

Approved:

Richard P. West, Ph.D. Major Professor	Timothy A. Slocum, Ph.D. Committee Member
Benjamin Lignugaris/Kraft, Ph.D. Committee Member	Scott W. Ross, Ph.D. Committee Member
Mark S. Innocenti, Ph.D. Committee Member	Mark R. McLellan, Ph.D. Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2016

ABSTRACT

Effects of a Self-Management Procedure Using Student Feedback on Staff Members' Use of Praise in an Out-Of-School Time Program

by

Cade T. Charlton, Doctor of Philosophy

Utah State University, 2016

Major Professor: Richard P. West, Ph.D.
Department of Special Education and Rehabilitation

Out-of-school time (OST) programs are under increasing pressure to improve student achievement. However, OST program administrators face a number of challenges to improving program effectiveness including inconsistent student participation, inexperienced staff members, and a lack of high-quality professional development. OST program administrators can address these challenges by implementing staff development practices that encourage the use of effective instructional strategies.

Specific praise is a simple and effective instructional strategy that has been linked to improved student engagement, enhanced academic achievement, and stronger student-teacher relationships. Unfortunately, there have been very few studies examining the effects of interventions designed to increase OST staff members' use of specific praise. One staff development strategy that could be both feasible and effective is the use of self-management. Although there are a variety of approaches to self-management designed

for teachers, not all self-management strategies are effective.

One strategy that might increase the feasibility and effectiveness of self-management programs is the use of student feedback. The process of comparing self-evaluations against a third-party standard such as student feedback is called matching in the self-management literature. Students can be a useful source of feedback because they observe their teachers frequently and can report the use of instructional strategies like specific praise. The purpose of this study was to examine the effects of a self-management procedure using student feedback on OST staff members' use of specific praise. A multiple-baseline design across participants was used to examine the effects of the intervention on specific praise rates.

All participants increased their use of specific praise after implementing the self-management procedures. General praise rates became more variable throughout the study. These findings provide evidence for a functional effect on specific praise but not for general praise. Teachers reported high levels of satisfaction with the feasibility and effectiveness of the intervention. A statistically significant correlation was found between specific praise rates and student reports of specific praise.

(211 pages)

PUBLIC ABSTRACT

Effects of a Self-Management Procedure Using Student Feedback on Staff Members' Use of Praise in an Out-Of-School Time Program

Cade T. Charlton

Out-of-school time (OST) programs could be improved by increasing staff members' use of effective praise. Specific praise is a relatively simple, efficient, and effective practice for increasing student engagement, reducing problem behavior, and improving student-teacher relationships. Unfortunately, many of the programs that have been used to help traditional teachers use praise more effectively in the classroom may be too expensive, complex, or inefficient to be of use in OST programs. Self-management programs are designed to be implemented without extensive coaching and support. One feature of effective self-management programs matching self-evaluations with external feedback. To make matching feasible in self-management programs for OST programs, teachers could match their self-evaluations to student feedback.

The purpose of this study was to evaluate the effects of a self-management program using student feedback on teachers' use of specific praise. All participating teachers increased their use of specific praise with limited training and support. Teachers indicated that student feedback was useful, the procedures were simple to implement, and they noticed improvements in the quality of their instruction. As the teachers' use of specific praise increased so did the percentage of students reporting that they received specific praise from their teachers.

ACKNOWLEDGMENTS

I would like to thank my committee members, Dr. Richard West, Dr. Timothy Slocum, Dr. Benjamin Lignugaris-Kraft, Dr. Mark Innocenti, and Dr. Scott Ross. I am indebted to these individuals for their support and encouragement. I would also like to thank Mary Ellen Heiner for sharing her incomparable editing and formatting skills with me in the final stages of preparing this document. Finally, I also like to thank current and former students from the Disability Disciplines program who edited this document, gave feedback on the research design, and encouraged me to keep going during challenging moments.

I would also like to thank the program administrators, directors, and staff members who made this study possible. They willingly to provided technical assistance, adjustments to the schedule, and every other opportunity to make this study possible. The teachers who participated in this study did so because they cared deeply about the students with whom they worked.

Finally, I express gratitude and appreciation for the many sacrifices of my family. My siblings selflessly read multiple versions of this document and gave timely, critical feedback. Shawn took time away from his responsibilities at work and at home to provide 24-7 technical support. My parents struggled through personal challenges with the hope that this document would be done. Finally, my children and wife deserve to be acknowledged for their faith in me and for living a year without a father and husband. You are the reason this is all worth it. Thanks!

Cade T. Charlton

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The importance of out-of-school time (OST) programs in education has increased over the last decade. These changes are the result of a variety of social and political forces including the need for safe and productive supervision after school, greater emphasis on improving student achievement, and calls for greater integration between school and community resources (Lauer et al., 2006; Miller & Snow, 2004). Because OST programs supplement students' public school experiences and often use school resources, policy-makers and community stakeholders suggest that administrators should evaluate how OST programs contribute to the academic and social development of participating students (Afterschool Alliance, 2015). Researchers have found that participation in OST programs has only a small, positive effect on academic achievement (Durlak, Weissberg, & Pachan, 2010; Lauer et al., 2006; Taheri & Welsh, 2015; Woodland, 2008). The limited benefits to students of participating in OST programs may be the result of inconsistent student attendance (Lauver, Little, & Weiss, 2004), a dearth of qualified staff members (Asher, 2012), and limited funding to support high quality programs (Maynard, Peters, Vaughn, & Sarteschi, 2013).

To enhance the effects of participation in OST programs and address these challenges, one critical investment that OST program administrators can make is in improving the effectiveness of their staff members (Hirsch, Mekinda, & Stawicki, 2010). Specifically, administrators in OST programs need to identify and implement strategies to help staff members engage youth in quality programs, provide meaningful feedback on

performance, and positively manage problem behaviors (Sanders, Wright, & Horn, 1997; West, Smith, & Taylor, 2004). Cross, Gottfredson, Wilson, Rorie, and Connell (2010) found that students reported higher levels of satisfaction and greater participation rates in programs with well-trained and experienced staff members who created a positive learning environment.

One instructional strategy often used to encourage student participation and create a positive learning environment is specific praise. Specific praise refers to statements spoken by a teacher that express approval for a student's academic or social behaviors, which include a description of the praiseworthy behavior (Marchand-Martella, Lignugaris-Kraft, Pettigrew, Leishman, & Ross, 2010; Sprick, Knight, Reinke, & McKale, 2008).

Specific praise is a simple, efficient, and effective instructional strategy that can increase a wide variety of appropriate behaviors (Jenkins, Floress, & Reinke, 2015). Rathel, Drasgow, Brown, and Marshall (2014) demonstrated that when teachers doubled their specific praise rate the task engagement of their students increased from a baseline mean of 70% to over 90%. In a similar study, Andrews and Kozma (1990) showed that when the praise rate to students who were identified as consistently off-task was systematically increased the rate of on-task engagement for these students increased from below 50% during baseline to over 80%. In multiple studies, researchers have shown that praise can be used to encourage pro-social responding in children (Phillips, 1984; Serbin, Tonick, & Sternglanz, 1977).

The use of praise also improves the classroom learning environment and other

critical conditions associated with academic achievement. Sutherland, Copeland, and Wehby (2001) reported that the frequent use of praise improves students' relationships with teacher, perceptions of fairness in the classroom, and feelings of self-efficacy and competence. Mueller and Dweck (1998) reported that when teachers who used specific praise to recognize the efforts of their students, students were more likely to engage in challenging activities even when the risk of failure was high. Finally, Anderson, Evertson, and Brophy (1979) showed that specific academic praise correlated ($\beta = .37$, $p = .04$) with student achievement.

In addition to increasing engagement and achievement, researchers found that when teachers use praise frequently to encourage appropriate behavior the frequency of challenging or problem behaviors in the classroom is reduced (Mesa, Lewis-Palmer, & Reinke, 2005; Reinke, Lewis-Palmer, & Martin, 2007). Mesa et al. found that when the use of praise exceeded three praise statements per minute, the frequency of students' problem behaviors decreased by 50%. Reinke et al., in a similar study, found that the use of performance feedback increased teachers' use of specific praise to a specific target student and his peers. The increased use of specific praise was accompanied by reductions in disruptive behavior for the target students and the non-target peers in the classroom. The percent of intervals with disruptive behavior during the observation of the target students decreased from 14-30% during baseline to 8-20% after the intervention. For non-targeted peers in the classroom, the percentage of intervals with disruptive behavior decreased from 7-20% during baseline to 5-17% during the intervention phases.

Despite the benefits of praise, many teachers do not use praise frequently in their

classrooms (Alber & Heward, 2000; Beaman & Wheldall, 2000; Jenkins et al., 2015; Stokes, Fowler, & Baer, 1978). Over the last 30 years, researchers have documented consistently low praise rates in elementary schools (Anderson et al., 1979; Good & Grouws, 1977; Jenkins et al., 2015) and secondary schools (Beaman & Wheldall, 2000; Evertson, Anderson, Anderson, & Brophy, 1980; White, 1975). In observational studies, researchers indicated that at-risk students were the least likely to be praised in the classroom (Burnett, 2002; Deno, Maruyama, Espin, & Cohen, 1990; Latham, 1997; Shores, Gunter, & Jack, 1993). If these rates are found with experienced, credentialed teachers in K-12 schools, then staff members in OST programs, who do not have the same level of training and experience, may exhibit similarly low rates of praise and could benefit from targeted training and support to improve their use of specific praise.

Unfortunately, the professional development provided in many OST programs emphasizes brief, introductory workshops. Typically, coordinators introduce the specific curriculum or strategies used in the OST program, but do not provide modeling, coaching, or performance feedback on the implementation of the strategy or related skills like classroom management (Hirsch et al., 2010). In their seminal work on staff development, Joyce and Showers (1980) indicated that traditional workshops are ineffective strategies for professional development. They noted the need for focused approaches on a few instructional strategies, effective modeling, simulated practice opportunities, immediate feedback, and long-term coaching. Contemporary implementation science builds on our understanding of what it takes to establish and sustain new practices by adopting a broad, systemic view of implementation drivers

(Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). Best practices in professional development and the findings from implementation science emphasize the challenge of sustaining new practices in a context as diverse and multi-faceted as an OST program and highlight the need for teacher development strategies that are focused and provide regular performance feedback.

Researchers have identified a variety of effective approaches for increasing teachers' use of praise. These strategies include coaching, performance feedback, and self-management (Cavanaugh, 2013; Fallon, Collier-Meek, Maggin, Sanetti, & Johnson, 2015; Hattie & Timperley, 2007). While these strategies are effective in traditional school settings, there are only a few published studies on the effectiveness of these strategies in non-traditional settings like OST programs or with inexperienced instructors (Cavanaugh, 2013). For example, Briere, Simonsen, Sugai, and Myers (2015) used a within-school consultation process to help new teachers increase their use of specific praise. Rathel et al. (2014) demonstrated that emailing new teachers performance feedback effectively increased their use of specific praise. Keller, Brady, and Taylor (2005) demonstrated that a self-management procedure involving completing self-ratings, collecting performance data from audio recordings, and graphing performance increased student teaching interns' use of praise.

Unfortunately, many of the strategies for increasing praise rates may not be feasible in typical OST programs. In the aforementioned studies, the researcher or an experienced coach provided support for data collection, analysis, and goal setting. Most OST programs do not have access to specialists who can support the collection of

performance feedback or meet regularly with teachers in consultation (Asher, 2012) or experienced staff members prepared to implement programs with fidelity (Hartje, Evans, Killian, & Brown, 2007; Nee, Howe, Schmidt, & Cole, 2006). Furthermore, securing training, coaching, or regular performance feedback is prohibitively expensive for many resource limited OST programs (Asher, 2012).

One appealing alternative to resource—and personnel-intensive approaches to staff development—is self-management. Self-management procedures are designed to help teachers evaluate, modify, and reward their own behavior through self-monitoring, self-instruction, and self-evaluation (Hoff & Sawka-Miller, 2010; Young, West, Smith, & Morgan, 1991). While self-management is effective in some education contexts, these procedures have not been vigorously tested with inexperienced instructors, and in the few available empirical studies, the effects are inconsistent (Keller et al., 2005; Simonsen, MacSuga, Fallon, & Sugai, 2013; Sutherland & Wehby, 2001).

One explanation for the ineffectiveness of some self-management strategies is the lack of matching in many of these approaches. Matching is a critical component of many highly effective self-management programs (Peterson, Young, Salzberg, West, & Hill, 2006; Smith, Young, West, Morgan, & Rhode, 1988; Young et al., 1991) and it involves using feedback from an external agent to enhance the quality of self-evaluations. Although adding matching to the existing approaches to self-management for teachers may increase the effectiveness of these strategies, it also requires an external observer who can facilitate the program and support OST teachers. This additional burden reduces the utility and feasibility of these interventions.

Student feedback could be used as a source of external feedback for matching in self-management programs. Researchers have demonstrated that students' can accurately perceive and report on the quality of the learning environments in their schools and classrooms (Bell & Aldridge, 2014; Downer, Stuhlman, Schweig, Martínez, & Ruzek, 2015; Fauth, Decristan, Rieser, Klieme, & Büttner, 2014; Gray, Graubard, & Rosenberg, 1974; Measures of Effective Teaching, 2012a). These data can also be used to improve instructional conditions. For example, Gray et al. taught students to observe the instructional practices of their teachers and differentially respond to the behaviors they wanted to encourage and those they wanted to decrease. In other studies, students evaluated the quality of the classroom learning environment as a component of teacher evaluation and development (Bell & Aldridge, 2014; Downer et al., 2015; Fauth et al., 2014; Measures of Effective Teaching, 2012a). In each case, students were active participants in the procedures that produced a change in their teachers' instructional practices.

Given the need for efficient and feasible interventions in OST programs and the lack of empirical research on staff development in these programs, the purpose of this study is to examine the effects of a self-management program designed specifically for OST programs. To reduce the need for external coaching and oversight, student feedback was integrated into the matching component of the self-management program. There are few studies that address the needs of OST staff development and none that have employed self-management strategies coupled with student feedback.

The primary research question for this study asked to what extent will teachers'

specific praise rates change when they implement the self-management procedures? The study also considered the effects of the self-management procedure on general praise rates and teachers' perceptions of social validity. Finally, the relationship between the percent of students reporting specific praise and specific praise rates was examined to evaluate the sensitivity of student reports to changes in teacher behavior.

CHAPTER II

LITERATURE REVIEW

Out-of-School Time Programs

OST programs are supplemental student support programs operated outside the legally defined school day or school year. The Boys and Girls Club (B&G Club), the United Way, and 21st Century Learning Centers are all well-known examples of these programs. It is estimated that more than 7 million students participate in OST programs each year (Carver & Iruka, 2006). Since 2004, total appropriations from the U.S. Congress to support OST programs has totaled more than $10 billion (U.S. Department of Education, 2015). Many parents and policymakers have independently advocated for even more funding and support for OST programs, specifically programs that increase student participation and foster improvements in academic and social behavior (Afterschool Alliance, 2015). Community support and funding have made OST programs more widely available and have coincided with increased student participation. From 1995 to 2005, the percent of students in kindergarten through eighth-grade enrolled in an OST program increased from 5.6% to 20% (Beckett et al., 2009).

Historically, OST programs were established to provide safe, secure locations for appropriate social activities while parents or other caregivers were at work (Halpern, 2002). Contemporary OST programs, however, are much more focused on contributing to critical student outcomes including academic achievement and appropriate social development (Lauer et al., 2006; Miller & Snow, 2004). One of the driving forces behind

this emphasis on measuring the impact of OST programs is the 21st Century Community Learning Centers (21st CCLC) program. The 21st CCLC program has received the largest portion of the available federal funding for OST programs with consistent funding available since 2001 (James-Burdumy, Dynarski, & Deke, 2007). This level of support may reflect the program's strong emphasis on helping at-risk students achieve proficiency on state and local performance standards.

Out-of-School Time Program Outcomes

While significant funding has been directed toward OST programs, there is no clear consensus among researchers regarding the effects of participation in OST programs on student outcomes. In multiple meta-analytic and narrative reviews, researchers have examined the effects of OST participation on students' social behaviors (Durlak et al., 2010; Kremer, Maynard, Polanin, Vaughn, & Sarteschi, 2014; Scott-Little, Hamann, & Jurs, 2002; Taheri & Welsh, 2015) and academic achievement (Lauer et al., 2006; Scott-Little et al., 2002; Woodland, 2008). Durlak et al. reviewed the effects of 69 different OST programs on students' social and emotional outcomes after participation. Many of these studies (68%) were unpublished or non-technical reports. Where possible, effect sizes were reported as the average standardized mean difference at post-assessment. When too little data were reported to calculate an effect size or the study concluded with a non-significant result, the authors reported the effect size as zero. This approach was criticized by other researchers because it may artificially reduce the variability in the data and bias the estimation of the aggregate effect size (Kremer et al., 2014; Lipsey & Wilson, 2001). The average effect size for measures of students' self-

perceptions (e.g., self-esteem, commitment to learning, sense of belonging) was 0.34, school bonding 0.14, positive social behaviors 0.19, and school attendance 0.10. All of the effect sizes were statistically significant except school attendance. The magnitudes of these consistently positive effects are considered small to moderate effect sizes (Cohen, 1992).

Scott-Little et al. (2002) reviewed 43 evaluation studies conducted on the effects of OST programs. Scott-Little et al. examined the quality of the evaluation procedures, consistency of program implementation, and the overall effects on social and academic outcomes. In 15 of the 23 studies included in the review, the researchers measured the impact of program participation on academic outcomes. In eight of the studies researchers used a quasi-experimental or experimental design, but in only six of these studies the researchers reported sufficient data to calculate an effect size using Cohen's *d*. Based on this limited sample, the effects on reading were approximately $d = 0.19$ and $d = 0.14$ for math. These small, yet statistically significant effect sizes are consistent with the conclusions of Durlak et al. (2010). However, it should be noted that Scott-Little et al. found substantive methodological flaws and inconsistent reporting in a majority of the 23 studies they reviewed. The most common problems they noted were combining student outcomes across grade levels, failure to report pre- and posttest means for each group, and limited descriptions of the practices and procedures in comparison conditions

Lauer et al. (2006) limited their review of research on OST programs to quantitative studies with control/comparison groups that included a standardized assessments of student achievement in reading or mathematics. These criteria were used

to identify 17 studies with reading and math outcome measures, 13 with only reading outcome measures, and 5 studies with only math outcome measures. A single trained coder reviewed each study and coded the articles for details concerning program structure, implementation fidelity, and outcomes. Effect sizes were calculated using Hedges g. The average effect size for reading was $g = 0.05$ and g = 0.09 for math. The authors reported that these effect sizes indicate small, but consistently positive effects on academic achievement. However, the magnitudes of these effect sizes suggest that the effects of OST program participation on academic performance are near zero.

Finally, Kremer et al. (2014) reviewed 24 studies examining the effects of OST participation on school attendance and problem behavior. All 24 studies were experimental or quasi-experimental group studies. School attendance was defined as either the number of full school days attended and problem behavior was broadly defined to include student, parent, or teacher reports of students' participation in inappropriate, disruptive, dangerous, or illegal activities. In 16 studies, researchers included effect sizes for the school attendance metric, calculated using the standardized mean differences formula and adjusted for small sample sizes using the Hedges g. The results indicated that there was a very small, nonstatistically significant effect of OST participation on school attendance ($g = 0.04$). The effects on problem behavior were similarly evaluated. Sixteen studies with 49 effect sizes were examined to determine that OST participation has a small, non-statistically significant effect on problem behavior ($g = 0.11$). Kremer et al. attributed these smaller, nonsignificant effect sizes to their use of more conservative data analysis procedures relative to previous reviews (Durlak et al., 2010; Scott-Little et

al., 2002). The authors used Hedges *g* to correct for bias due to low sample sizes and all effect sizes were analyzed for each study using robust variance estimation.

These meta-analyses consistently point to small, positive effects of OST programs on student outcomes (Durlak et al., 2010; Kremer et al., 2014; Lauer et al., 2006; Scott-Little et al., 2002). Unfortunately, the small effect sizes and inconsistencies in the findings are enough to question the effectiveness of OST programs and the wisdom of continuing to allocate significant funding to their support. This finding applies across a wide variety of academic (e.g., math, science, reading) and social (e.g., delinquency, school attendance, pro-social behaviors, self-esteem, etc.) variables. These results seem to confirm Apsler's (2009) view on the origin of OST programs when he indicated "the rapid growth of after-school programming resulted from lobbying and grass roots efforts and was not based on strong empirical findings" (p. 2).

Challenges Contributing to Out-of-School Time Program Outcomes

One explanation for these small effects is poor student attendance and engagement in OST programs (Durlak et al., 2010; James-Burdumy et al., 2007; Lauer et al., 2006; Roth, Malone, & Brooks-Gunn, 2010). If students do not regularly participate in OST programs, then they cannot reasonably be expected to benefit from even the most robust and effective program. The Harvard Family Research Project estimated that young children attend only about 60% of the available sessions during the school year, whereas adolescents attend less than 20% of sessions (Lauver et al., 2004). Roth, Malone, and Brooks-Gunn (2010) suggested that increasing attendance is necessary but not sufficient

to improve outcomes. Roth et al. reviewed 35 studies of the effects of OST program participation on student outcomes. They reported that less than 25% of studies demonstrating increased student attendance also reported a significant improvement in academic achievement. The majority of studies reviewed by Roth et al. measured student attendance using a daily count of students attending the program. A better measure of student participation that may be sensitive enough to detect the relationship between student participation and performance is the length of time students were engaged in program activities and the extent of their participation.

Another variable that may help explain the limited benefits from OST participation is poor program implementation (Hirsch et al., 2010; Maynard et al., 2013). Maynard et al. reviewed 55 studies examining the effects of OST programs on academic and behavioral outcomes with at-risk students. They found that in 42% of the studies they reviewed, researchers indicated that a detailed program manual was provided to OST staff members and 33% indicated that training was provided on the core features of the OST program. Finally, in 24% of the studies researchers indicated that staff members received ongoing support and supervision to maintain high levels of treatment fidelity. Maynard et al. conclude that implementation fidelity could be improved by either employing and retaining staff members capable of implementing programs with high fidelity or developing strategies to systematically develop staff members to ensure quality implementation. Students report lower levels of program satisfaction, fewer positive interactions with staff members, and more frequent conflicts between students in programs with poor implementation fidelity (Cross et al., 2010).

Poor implementation fidelity is associated with inexperienced OST staff members (Hartje et al., 2007; Keller et al., 2005; Little, Wimer, & Weiss, 2007). Approximately 24% of OST staff members have only a high school diploma and 75% had no prior experience working in educational settings with children (Hartje et al., 2007; Nee et al., 2006). Of the roughly 50% of OST staff members who have completed a bachelor's degree or higher, nearly one fourth held a degree in an unrelated field (Nee et al., 2006). Because both Hartje et al. and Nee et al. used an online survey sent to program directors and site supervisors to examine the qualifications of OST staff members, the authors reported a concern that they might have overestimated the qualifications of OST staff members by systematically biasing the sample toward program leaders rather than the typical, front-line staff members and teachers. A review of OST programs associated with the B&G Clubs of America found that high staff turnover in many clubs contributed to a pervasive lack of experienced staff members with skills in relevant areas (e.g., mentoring, behavior management, writing, etc.) and a lack of institutional knowledge to sustain program quality (Arbreton, Sheldon, & Herrera, 2005). In a multi-site study conducted in Maryland, Cross et al. (2010) found that the more than 30% of the sites failed to fully staff their programs with a program director and three staff members prior to the start of the program. Nearly 50% of the sites participating in the study hired additional staff members after the start of the program to address critical staff shortages caused by higher than expected student enrollments.

In addition to inexperience, researchers examining OST program quality indicated that many staff members do not use appropriate child development and classroom

management practices (Asher, 2012; Larson & Walker, 2010; Scott-Little et al., 2002). For example, Larson and Walker reported that many staff members in OST programs struggle with maintaining fairness, clarity of expectations, and professionalism while trying to befriend and mentor struggling youth. Little et al. reviewed case studies of OST programs throughout the U.S. and found that punitive interactions between students and staff members were consistently associated with poor program outcomes including inconsistent student participation. Furthermore, Asher indicated that a majority of staff members lack basic training classroom management strategies. This is unfortunate because a third of OST program staff report that their program exclusively serves at-risk or vulnerable populations (Hartje et al., 2007). In national surveys researchers report that a higher proportion of students attending OST programs have a disability (approximately 16%) compared to the proportion of students in public K-12 schools (approximately 13%) who have a disability (Kena et al., 2015; Taylor, 2013). The fact that many at-risk students participate in OST programs and may not receive effective support is a cause for concern.

Developing Effective Out-of-School Time Staff Members

The importance of effective staff members to OST programs cannot be overstated (Cross et al., 2010). Staff members are directly responsible for supervising participants, providing feedback on academic activities, modeling appropriate social skills, and building relationships. Furthermore, the quality of the interactions between instructors and students is the best predictor of student achievement in schools and OST programs (Marzano, 2003; Sanders et al., 1997; West et al., 2004) and is significantly related to

engagement and participation in OST programs (Greene, Lee, Constance, & Hynes, 2013). Joseph Durlak indicated,

> If you want to influence real-world practices, you must make a concerted effort to inform and to collaborate with front-line providers and to support and to problem solve with them as new programs or practices are introduced into their setting. The eventual consumers of any after-school program, which means youth and their parents, should have some input. (Granger, 2008, p. 13)

Unfortunately, many OST programs are ill prepared to attractor develop high quality staff members (Asher, 2012). Researchers have repeatedly found that OST programs lack sufficient training and support for staff members (Hirsch et al., 2010; Maynard et al., 2013; Scott-Little et al., 2002). Limited staff support is often due to a lack of funding to recruit and retain experienced, well-trained site supervisors (Grossman et al., 2002). Experienced trainers are critical to effective staff development because they can provide modeling of effective instruction, onsite feedback, and ongoing support to encourage maintenance of the critical skills (Joyce & Showers, 2002; Showers, Joyce, & Bennett, 1987). In addition, resources also constrain the number of staff that can be hired in a support capacity. Asher reported that the relatively low wage and the limited number of hours available contributed to ongoing concerns over the availability of qualified instructors and support staff. As a result, when qualified staff are employed in OST programs they often support multiple sites and oversee too many staff members to provide effective support (Hirsch et al., 2010).

To improve the conditions for learning and student outcomes in OST programs, OST program administrators need effective strategies for staff development (Hirsch et al., 2010). Staff development is more effective when it is designed to fit the values, goals,

constraints, and resources of the target context (Horner, Blitz, & Ross, 2014). For OST programs, this means that low-cost, efficient staff development strategies that focus on specific, effective instructional practices that positively influence student engagement and achievement will likely be the most effective because this approach is consistent with the needs, values, and goals of these programs. One effective, low-cost strategy that could be readily used in OST programs is praise. Praise is an effective classroom management strategy that can be used to encourage student engagement and motivation.

Praise

Praise, in the most general terms, refers to expressing approval for someone or something ("Praise," 2016). As an instructional strategy, however, praise is more narrowly defined. Praise is defined as vocalizations made by a teacher to a student that communicate approval for a student's academic or social behavior (Madsen, Becker, & Thomas, 1968; Marchand-Martella et al., 2010). Praise has been used by effective teachers for generations. For example, in Plato's *Crito; Or, The Duty of a Citizen* (trans. 1870) Socrates converses with Crito about injustice. Socrates frequently praises Crito's participation and thinking with comments like "you say well" (Crito 9) and "you say truly" (Crito 10). Empirical research on praise as an instructional strategy has been ongoing since the early 1960s (Becker, Madsen, Arnold, & Thomas, 1967; Madsen et al., 1968; Zimmerman & Zimmerman, 1962). This interest in praise is related to its cost-effectiveness, simplicity, and intuitiveness as an instructional strategy that can be applied to a wide variety of instructional contexts as an integral part of effective feedback (Hattie

& Timperley, 2007).

Effects of Praise

The effectiveness of praise depends on how it is used in the classroom. Praise functions as a conditioned, social reinforcer when its contingent delivery increases the likelihood that the target behavior will be emitted under similar conditions in the future (Vollmer & Hackenberg, 2001). Researchers have experimentally manipulated the praise's rates in the classroom and demonstrated that contingent praise for on-task behavior increases engagement (Andrews & Kozma, 1990; Sutherland, Alder, & Gunter, 2003; Sutherland, Wehby, & Copeland, 2000). Similarly, when teachers praise students' use of appropriate self-referent statements, students report higher levels of social and emotional awareness and self-esteem (Phillips, 1984; Serbin et al., 1977). Praise has an indirect effect on the frequency of problem behaviors by systematically encouraging appropriate behavior and eliminating the opportunity and, in some cases, the need to engage in problem behavior (Mesa et al., 2005; Moffat, 2011; Reinke et al., 2007).

Cherne (2008) conducted a meta-analysis of the effects of praise on students' academic and social behavior. Cherne identified 18 experimental studies using single-subject designs that included a systematic manipulation of teachers' use of praise and measured appropriate social behavior, inappropriate social behavior, or academic behaviors. Aggregate effect sizes were estimated by calculating the percentage of nonoverlapping data (PND) and the percentage of all nonoverlapping data (PAND). Articles were also coded for the age of the students, disability status, type of praise (e.g., praise for ability, praise for effort, specific praise, or other), and whether or not praise

was used with other instructional strategies. The median effect of praise on academic behaviors was estimated at 78% using PND and 88% using PAND, a moderately strong effect. This finding means that there was less than 22% overlap on average between the data collected on academic behaviors during Baseline and the data collected during the praise or intervention phase. The median effect of praise on appropriate behavior was estimated at 58% using PND and 79% using PAND, a moderate effect. Finally, the strongest effect was observed in reducing the frequency of inappropriate behavior with a median effect size of 87% using PND and 92% using PAND. Thus, a strong effect was noted between the use of praise and reductions in problem behavior. Latham (1997) wrote, "The research is clear on this matter! Teachers have simply got to learn to be much more positive and encouraging than negative and discouraging" (p. 10).

Despite the research evidence, praise is not universally endorsed. One critic, Alfie Kohn (1999) argued that rewards, including praise, undermine a child's intrinsic motivation to learn. In response Kohn's argument, several researchers reviewed the research literature to find evidence to support or reject this criticism of the use of praise (Cameron, 2002; Cameron & Pierce, 1996). These reviewers found that Kohn's position relied heavily upon studies showing decreased task engagement in children who were offered a tangible reward non-contingently following engagement in a preferred task (Coe, Aloisi, Higgins, & Major, 2014; Lepper, Keavney, & Drake, 1996; Ryan & Deci, 1996). Cameron (2002) and Cameron and Pierce (1996) found support for the use of praise as an effective instructional strategy when it was used as a specific, contingent consequence for appropriate academic or social behavior. These studies highlight the

importance of using praise as a contingent response to appropriate behavior and further specify that the specificity (i.e., what is said in the praise statement) is a critical variable in predicting the effects of praise on students' behavior. For example, students who received praise for their efforts to accomplish difficult tasks were more likely to take risks, persist in difficult situations, and to pursue diverse learning opportunities than students who were praised for their abilities (Dweck, 2002; Mueller & Dweck, 1998).

Specific Praise

A praise statement must be functionally related to changes in student behavior to be considered a reinforcer. One critical feature of effective praise, or praise that functions as a reinforcer, is the contingency or dependency between the delivery of the praise statement and the occurrence of the target behavior. Praise that is contingent upon a select target behavior is more likely to produce immediate behavior change (Vollmer & Hackenberg, 2001). Unfortunately, several researchers have found that praise is often delivered following inappropriate behavior (Nafpaktitis, Mayer, & Butterworth, 1985; Strain, Lambert, Kerr, Stagg, & Lenkner, 1983).

One of the most effective ways of ensuring teachers are using contingent praise is to encourage them to use specific praise (Burnett & Mandel, 2010; Fisher, Ninness, Piazza, & Owen-DeSchryver, 1996; Sprick et al., 2008). Specific praise statements are verbal commendations, recognition, or affirmations spoken to a student that include a detailed description of the appropriate behavior (e.g., "Great job putting your coat away," "I like the way you walked in the hallway," "Yes! The sum of five plus five is ten," etc.). By contrast, general praise statements are positive verbal statements that lack the

behavioral description (e.g., good job, well done, thank you, etc.).

One explanation for the effectiveness of specific praise is its dual effect on students and teachers. For students, specific praise explicitly identifies the reinforcement contingencies in the classroom and helps clarify expectations (Fisher et al., 1996; Hattie & Timperley, 2007). For example, Fisher et al. conducted a study on the reinforcing effects of the content of verbal statements. They experimentally manipulated the content of verbal statements following instances of problem behavior. Fisher et al. found that statements that identified the problem behavior (e.g., "don't hit me" or "that hurts") were nearly twice as effective as unrelated statements (e.g., "it's sunny today"). The experimenters controlled for the intonation, facial expressions, and other features of the delivery of the statements by using different therapists in the descriptive and unrelated statements conditions These data support the idea that the content of verbal statements can alter the effects of verbal statements on student behavior. For teachers, the use of specific praise can help them maintain the contingency between their praise statements and student behavior (Anderson et al., 1979; Sprick et al., 2008).

Specific praise is more effective and preferred by students relative to general praise (Burnett, 2002; Burnett & Mandel, 2010). Burnett reported that students who received more specific praise in the classroom rated the quality of their classroom environment higher than students who received general praise. Burnett and Mandel reported that students in grades 6-12 preferred specific praise relative to other forms of feedback. Unfortunately, Wyatt and Hawkins (1987) reported that teachers in the younger grades were far more likely to use specific praise than teachers in the older grades.

Despite secondary students' preference for specific feedback, they may be less likely to receive specific praise than students in elementary schools.

Finally, it is important that teachers allocate more specific praise to struggling students in their classrooms (Andrews & Kozma, 1990; Brophy, 1981). Andrews and Kozma attempted to systematically alter the distribution of praise for on-task behavior in a target classroom to favor students who were frequently off task. The researchers used a prompt to praise students who were identified as frequently off-task and performance feedback to alter the distribution of praise in the classroom toward students in the off-task group. Although the students receiving the most praise prior to intervention still received a larger proportion of praise after the intervention, the increased use of praise with the off-task group was associated with a steady increase in on-task behavior. Thus, teachers should strive to increase their use of contingent praise, specific praise, and to ensure all students in their classrooms are frequently praised.

Observed Praise Rates

Brophy (1981) indicated that praise is used infrequently and often inappropriately in elementary and secondary classrooms. This critique has been echoed by other researchers who reported consistently low praise rates in elementary (Anderson et al., 1979; Baker & Zigmond, 1990; Good & Grouws, 1977) and secondary classrooms (Evertson et al., 1980; White, 1975). Beaman and Wheldall (2000) conducted a review of the research on naturalistic praise rates in K-12 classrooms. Their review discusses research across three decades starting with the 1970's. Observational research in the 1970s suggested teachers infrequently praised students for appropriate academic

responses and rarely praised for appropriate social behavior. For example, White (1975) found that praise rates in third grade and above were around 1-2 praise statements every ten minutes. Anderson et al. indicated that praise was used to encourage student responding after just 11% of correct answers.

Research in the 1980s, confirmed that praise was used infrequently in many classrooms and looked at the use of praise more systematically in a variety of educational contexts. For example, Brophy (1981) also found that teachers in elementary school classrooms were twice as likely to use praise relative to teachers in secondary schools. Beamann and Wheldall (2000) suggest that the finding that praise was used at a rate of 1.15 per minute in elementary schools indicates an increase over studies conducted in the 1970s, but praise in secondary schools was still infrequent and praise was still no more likely to be used to encourage social behavior.

Finally, in the 1990s, research on praise was conducted in international schools. Praise rates of 0.61 per minute in Australian primary schools and 1.15 in British primary schools (Merrett & Wheldall, 1987). In high schools, praise rates were lower with rates of 0.65 and 0.40 for Australian and British teachers, respectively (Wheldall, Houghton, & Merrett, 1989). Beaman and Wheldall (2000) are quick to point out that these apparent changes in teacher practice may be attributed to changes in measurement and the operational definition of praise. In general, they found definitions of praise becoming more general over time. They end their review by stating, "…it would appear that many teachers fail to take full advantage of [praise as a] potentially powerful behavior management tool" (Beaman & Wheldall, 2000, p. 443).

Equally discouraging is the distribution of praise among students. In observational studies researchers indicated that students with disabilities and students who engage in frequent disruptive behavior were the least likely to be praised in the classroom (Burnett, 2002; Deno et al., 1990; Latham, 1997; Shores et al., 1993; Sutherland, 2000). Sutherland (2000) concluded that praise rates with high-risk students were as low as 1.3-4.5 praise statements per hour and that the ratio of negative to positive interactions typically exceeded 2:1. Jenkins et al. (2015) reviewed the research literature to determine if praise was similarly employed in special and general education classrooms. They identified four studies examining the natural praise rates in special education classrooms. Based on their review of these studies and a comparison with studies conducted in general education classrooms, they concluded that praise rates were significantly lower in special education classrooms (i.e., fewer than five total praise statements per hour).

The limited use of praise may be attributed to a lack of naturally occurring reinforcement in schools and classrooms (Alber & Heward, 2000; Stokes et al., 1978) to sustain high rates of teacher praise. Regardless of the etiology, the available research confirms that praise is not abundantly used in classrooms generally and, perhaps more concerning, praise is not widely used with students with disabilities or those who may be at risk. The students who would benefit the most are often the least likely to be praised. Unfortunately, there is no clear consensus among researchers and practitioners on the optimal specific praise rates in elementary schools, secondary schools, or with at-risk students (Jenkins et al., 2015).

Increasing Praise Rates

The research literature confirms that the natural conditions in many schools do not support high praise rates. Unfortunately, there are no studies examining the use of praise in OST programs, but it is reasonable to conclude that staff members in OST programs, many of whom lack specialized training and access to high quality professional development, may not have high rates of specific praise. Thus, OST staff members are likely in need of high quality interventions and supports to help them achieve high praise rates. A variety of strategies to increase the use of specific praise have been examined by researchers including training, performance feedback, coaching, and self-management.

Training

Training is the simplest and most direct approach to increasing teacher's use of praise. Wheatley (2015) examined the effects of didactic instruction on OST program staff members' use of praise. Participants in the study were all staff members in an OST program operating in a northwestern U.S. public school. A research assistant observed each group and coded general praise, specific praise, and negative interactions including reprimands, error corrections, and all instructions to stop a current behavior. Finally, observers coded student behavior as engaged, off-task, or disruptive. During Baseline, staff members used praise at a rate of less than one statement per minute and students were off task for about 40% of the sessions and disruptive for 10% of the time, on average.

All participants attended a didactic training focused on increasing praise rates and reducing reprimands, or negative statements, during instruction. The experimenter

conducted the training, which included a rationale for increasing praise rates, a discussion of the findings from previous research on praise, and practice opportunities with specific strategies used to increase praise rates. Following the training, none of the teachers sustained improvements in their use of praise. For two of the teachers, the experimenter examined the addition of performance feedback, goal setting, and coaching. These additional procedures resulted in a substantial increase in the teachers' praise rates to over five praise statements per minute. A reduction in students' off-task behavior was noted in one of the two classrooms. Unfortunately, the use of an ABC design does not permit sufficient verification and replication of these effects to document a functional relationship between coaching and praise rates.

Dufrene, Lestremau, and Zoder-Martell (2014) conducted a study of the effects of didactic training on the use of specific praise in an alternative school serving students with behavioral disorders. Two certified teachers were observed by trained research assistants during instruction for 20 minutes each day. Observers noted the frequency of disruptive student behavior and the teachers' use of specific praise. Prior to training, one of the teachers had a specific praise rate near 0 while the other had a rate near 0.25 statements per minute. The training included the following components: a rationale for the use of specific praise, examples of specific praise, examples of student behaviors coupled with specific praise, and opportunities to practice using specific praise with immediate feedback from the experimenter. The experimenter provided the training to both participants one day apart. Following implementation of the training, one teacher's praise rate remained unchanged and the other increased slightly to approximately 0.5

specific praise statements per minute. There was no change in the frequency of disruptive behavior for the teacher whose praise rate did not change and a slight downward trend in disruptions for the other teacher. This study, and the work of Wheatley (2015), both illustrate that didactic training alone may be insufficient to produce substantial changes in the use of praise in the classroom.

In their seminal work on staff development, Joyce and Showers (1980) highlighted the need for clear presentation methods, effective modeling, simulated practice opportunities, immediate feedback, and long-term coaching to achieve sustained use of instructional practices in the classroom. Didactic trainings lack the ongoing, job-embedded feedback necessary to fully integrate explicitly taught skills into classroom practice (Joyce & Showers, 2002; Showers et al., 1987). Research on effective implementation echoes these recommendations. Effective implementation requires a broad, systemic view of skill development and emphasizes the importance of context-specific feedback and support from internal and external agents (Fixsen et al., 2005). One approach to enhancing the potency of training is providing access to coaching and feedback on performance.

Coaching and Performance Feedback

Coaching and performance feedback are integral components of effective professional development and training (Fixsen et al., 2005; Joyce & Showers, 1980; Showers et al., 1987). Coaching is defined as individualized support provided by a peer or an expert to assist the individual in acquiring a new skill or behavior (Ackland, 1991; Kretlow & Bartholomew, 2010; Lu, 2010; Sprick et al., 2008). Performance feedback is

"information provided by an agent (e.g., teacher, peer, book, parent, self, experience, etc.) regarding aspects of one's performance or understanding" (Hattie & Timperley, 2007, p. 81). Although performance feedback can be delivered independent of coaching, they are most frequently used and studied in combination (Cavanaugh, 2013).

There are two common approaches to coaching including expert or peer coaching (Kretlow & Bartholomew, 2010). Expert, or evaluative coaching, is most commonly used to support professionals (e.g., teachers, staff members, etc.) as a follow-up to inservice trainings. Peer coaching is accomplished by pairing two colleagues operating in similar assignments (e.g., two teachers, two preservice teachers, etc.) and providing them a structured process to discuss performance, review data, and plan for improvement. The use of peer coaching increased following the publication of Joyce and Showers' (1980) research, which increased awareness of the need for better coaching and support for preservice and inservice teachers. Peer coaching as a support strategy for preservice teachers is of interest here because typical OST staff members share much in common with preservice teachers (e.g., they lack specialized training and experience). Peer coaching involves a teacher or staff member observing a colleague, then using the results of that observation to collaboratively set goals for improvement (Hasbrouck, 1997; Morgan, Menlove, Salzberg, & Hudson, 1994). Lu's (2010) review of the literature on peer coaching identified eight studies examining the effects of peer coaching on preservice teachers' learning and performance. Peer coaching was linked to improvements in classroom management, assessment quality, and the implementation of effective instructional practices. Student teachers in a majority of the reviewed studies

indicated they preferred a peer coach over an expert because the participating teachers felt less anxiety, expressed a greater willingness to ask questions, and indicated that they felt greater equality with their coach. Although student teachers expressed a preference for peer coaching, in only one study researchers directly compared the effectiveness of peer coaching and expert coaching (Pierce & Miller, 1994). Pierce and Miller found that both approaches were effective at improving teacher behavior, and no significant differences were found between the effects of the two approaches on teacher behavior.

Finally, the researchers who conducted seven of the eight studies in Lu's (2010) review emphasized the importance of training peer coaches. Training on peer coaching was typically provided to students who had already completed the majority of their preservice training program. Although the training procedures varied from study to study, the average duration of the training coaches received was approximately five hours. The researchers trained coaches on valid data collection procedures, reporting performance feedback, and facilitating goal setting. In studies where technology was used to facilitate all or part of the peer coaching interactions, the researchers provided additional training and technical assistance on the use of the technology. Lu indicated that the additional cost and burden of this training was one of the reasons peer coaching has not been widely adopted.

Several recent reviews have been conducted to examine the effects of performance feedback and coaching on teachers' use of praise (Cavanaugh, 2013; Cornelius & Nagro, 2014). Cavanaugh examined the effects of performance feedback on praise in 22 experimental studies. Limitations in the designs or execution of four of the

22 studies make their conclusions difficult to interpret. For example, in one study researchers used a sequential components analysis (Sloat, Tharp, & Gallimore, 1977) similar to an ABCDE design without a baseline or control condition (e.g., reversal or withdrawal phase). In three studies researchers used a multiple-baseline design across two participants (Martens, Witt, Elliott, & Darveaux, 1985; Mesa et al., 2005; Rathel, Drasgow, & Christle, 2008). Of the 18 studies using experimental designs, 14 (78%) examined the use of performance feedback with coaching and four (22%) without coaching (Barton & Wolery, 2007; Kalis, Vannest, & Parker, 2007; Keller et al., 2005; Sutherland & Wehby, 2001). In the four noncoaching studies, teachers received performance feedback and engaged in structured self-reflection using a self-management procedure.

In five studies from Cavanaugh's (2013) review, researchers explored the effects of training with and without performance feedback (Cossairt, Hall, & Hopkins, 1973; Myers, Simonsen, & Sugai, 2011; Reinke et al., 2007; Simonsen, Myers, & DeLuca, 2010; Sloat et al., 1977). In all five studies, researchers used multiple-baseline designs and included a training only condition, which preceded the implementation of the full intervention in a performance feedback and coaching condition. Only two of the 18 participants from these studies exhibited an improved pattern of responding following the training only phase. These results confirm the conclusion that training alone may be insufficient to produce substantial and sustained changes in praise rates. Cavanaugh concluded that performance feedback is a more potent intervention than training, but the review fails to acknowledge that many of the performance feedback interventions

incorporate informal training into the procedures provided by the coach or other agent delivering the performance feedback.

Cornelius and Nagro (2014) also reviewed the performance feedback literature. Their review focused on the use of performance feedback in preservice teacher training programs. Of the seven articles they reviewed, three examined the effects of performance feedback on praise (Barton & Wolery, 2007; Capizzi, Wehby, & Sandmel, 2010; Rathel et al., 2008) and in the other four studies researchers focused on using performance feedback to improve implementation fidelity. Cornelius and Nagro concluded that the effects of performance feedback on implementation fidelity are consistent and robust across the four studies, but the effects on specific praise rates were inconsistent. In the reviewed studies, researchers reported that 33-80% of participants improved after receiving feedback. In other reviews, researchers concurred with the conclusion that performance feedback is an empirically supported treatment for improving implementation fidelity (Cornelius & Nagro, 2014; Fallon, Collier-Meek, Maggin, Sanetti, & Johnson, 2015; Kretlow & Bartholomew, 2010; Scheeler, Ruhl, & McAfee, 2004), but the effects of performance feedback on specific praise need to be clarified in future research.

The relevance of the literature on performance feedback and coaching to OST programs depends on the similarity between the contexts in which performance feedback has been studied and the typical OST program. The training and qualifications of the participants in these studies is an important concern when considering the applicability of performance feedback to OST programs because the effectiveness of performance

feedback may depend on the instructors' knowledge and familiarity with instruction. Table 1 includes all the performance feedback studies identified by Cavanaugh (2013) and an additional study identified through an updated literature search using the search terms and databases used by Cavanaugh in their literature search (Briere et al., 2015). In

Table 1

Summaries of Studies of the Effects of Performance Feedback on Instructors' Use of Praise in the Classroom

Citation	Participants	DV	Results
Briere et al. (2015)	3 induction-level teachers	SP (Frequency)	SP +3x
Capizzi et al. (2010)	3 preservice teachers	SP (Frequency)	SP +1.2x
Cossairt et al. (1973)	3 experienced teachers	GP (Frequency)	GP + 3.5x
Duchaine, Jolivette, & Fredrick (2011)	3 experienced teachers	SP (Frequency)	SP +1.1x
Fullerton, Conroy, & Correa (2009)	4 experienced teachers	SP (Frequency)	SP +2x
Hawkins & Heflin (2010)	3 experienced teachers	SP and GP (Frequency)	SP +1.4x GP +1.6x
Hemmeter, Snyder, Kinder, & Artman (2011)	4 experienced teachers	SP (Percent of intervals)	SP +10%
Mace, Cacelli, & Manos (1983)	1 experienced teacher 2 unlicensed teachers	GP (Frequency)	GP +1.4x
Moffat (2011)	1 experienced teacher	SP (Frequency)	SP +6x
Myers et al. (2011)	4 experienced teachers	SP and GP (Rate)	SP +1.2x GP +1.1x
Reinke et al. (2007)	3 experienced teachers	SP (Probability)	SP +3x
Reinke, Lewis-Palmer, & Merrell (2008)	4 experienced teachers	SP and GP (Frequency)	SP +2x GP +1.2x
Simonsen et al. (2010)	3 experienced teachers	SP (Frequency)	SP +8x
Stormont, Smith, & Lewis (2007)	3 experienced teachers	SP (Rate)	SP +2x
Sutherland et al. (2000)	1 experienced teacher	SP and GP (Frequency)	GP +1.5x SP +5x

Note. Dependent Variables (DV) include: Specific praise statements (SP); general praise statements (GP). Table adapted from Cavanaugh (2013).

the available studies participating teachers were experienced, induction-level, preservice, or unlicensed. Experienced teachers were licensed and had been teaching for two or more years. Induction-level teachers were licensed as well, but had been in the profession for less than two years. Preservice teachers were enrolled in a teacher preparation program whereas unlicensed teachers were not.

Forty-seven teachers were included in the studies summarized in Table 1. Of the 47 teachers, only 17% had been teaching for less than 2 years. The number of teachers who had advanced graduate training was approximately 65%. Two of the 47 (4%) teachers were identified as unlicensed. These individuals were both serving in a school but had not completed any post-secondary education programs. In a majority of the studies (80%), researchers included the frequency of specific praise statements as a dependent variable. In all of the studies included in Table 1, researchers demonstrated that performance feedback and coaching had a positive impact on the use of praise. Increases in the frequency of praise ranged from a 10-800%. Session duration was held constant in these studies. By contrast, the increases in praise rate were less dramatic, ranging from a 20-100% increase. Overall, the findings from these studies suggest that performance feedback for experienced teachers is an effective strategy.

Unlike the teachers in the research cited above, teachers in OST programs typically do not have advanced training and extensive experience in education so the closest comparison to their skills and knowledge are new teachers, preservice teachers, or noncertified staff members. Only three of the 16 studies in Table 1 were conducted with teachers with less than 2 years of experience (Briere et al., 2015; Capizzi et al., 2010;

Mace et al., 1983; Rathel et al., 2014). These studies provide the best estimate of the effectiveness of performance feedback and coaching on the use of praise with OST teachers.

Mace et al. (1983) examined the effects of performance feedback and coaching on the use of praise by two paraprofessionals. Both paraprofessionals provided small group, intensive services to a student with a history of academic and social problems. To increase praise rates, the researcher met after each instructional session for 10-15 minutes with each of the participants to review performance data and model the use of praise. These procedures increased the frequency of praise for both participants from near zero levels prior to the intervention to a mean of 0.5 praise statements per minute after performance feedback was implemented. A decreasing trend was evident in both participants' data suggesting the effects of the intervention did not sustain after the initial novelty of the training.

Capizzi et al. (2010) examined the effects of self-evaluation and consultation on preservice teachers' use of specific praise. Prior to meeting with the instructional consultant, a doctoral student trained to provide coaching on elements of effective instruction, each of the three preservice teachers recorded a 30-minutes video of themselves teaching a lesson. This video was reviewed with the instructional consultant during a one-hour consultation meeting. Immediately after the first consultation, two of the three teachers showed an immediate increase in their use of specific praise. Praise rates increased from a mean of 0.35 during Baseline to 0.80 specific praise statements per minute during the consultation phase. For the other teacher, her mean praise rate during

baseline was 0.83, equivalent to the post-intervention rates for her peers, and then 1.07 specific praise statements per minute after consultation. Unfortunately, her highest and lowest praise rates occurred on the first and second days of baseline, respectively. This resulted in 100% overlap between the data in the baseline and consultation phases. The high degree of overlap and insufficient evidence of an experimental effect (e.g., no clear change in level, trend reversal, or change in variability between phases) for this teacher contribute to the conclusion that consultation was not functionally related to changes in specific praise rates.

Briere et al. (2015) conducted a study of the effects of within-school consultation on new teachers use of specific praise. The three new teachers selected to participate in this study had been employed for less than three years and used praise at a rate below 0.4 specific praise statements per minute during an initial 15-minute observation. Each participant was matched with an experienced mentor who was trained on the consultation procedures by a member of the research team. The mentor trained their mentee to collect data on their use of specific praise by clicking a hand-counter each time they used praise. They were also taught to graph their specific praise data. The authors indicate that all trainings lasted approximately 15 minutes and were completed with perfect fidelity to the scripted procedures. The weekly consultation meetings included a review of the data from the previous week, the selection of an appropriate goal based on previous performance, and collaboratively identifying strategies to help the new teacher improve performance. The authors do not specify how long mentors and mentees took to complete each consultation meeting, but the meetings were held weekly throughout the duration of

the study.

A nonconcurrent multiple baseline design was used to examine the relationship between participation in the consultation model and use of specific praise. Participating teachers achieved substantial improvements in their specific praise rates following consultation with rates nearly tripling on average. These effects were maintained during a follow-up condition where teachers were observed once a week for four weeks by the researcher.

Of the studies discussed in depth, two demonstrated a functional relationship between the intervention and teachers' specific praise rates (Briere et al., 2015; Mace et al., 1983). Improvements in excess of two to three times the baseline praise rate were observed in the majority of participants in these two studies. Collectively, the preliminary evidence from the reviewed studies suggest that performance feedback and coaching could be effective with OST teachers, but further research is needed. The need for additional research is supported by the reviews on this topic, which all have concluded that more research is needed to determine how effective performance feedback is with a different types of instructors and in different educational contexts (Cavanaugh, 2013; Cornelius & Nagro, 2014; Lu, 2010).

We should not assume that the results from these studies will be mirrored in applications of performance feedback and coaching in OST programs because features of the context within a public school may moderate the effects of these interventions. For example, teachers in these studies had specific professional expectations as part of a school faculty and they had knowledgeable colleagues to help them benefit from access

to performance feedback. In addition, many teachers in OST programs lack the experience of even preservice teachers (Asher, 2012). These individuals may not be prepared to participate in a peer coaching or to analyze complex performance feedback. Research on the effects of performance feedback and coaching in OST programs is both warranted and needed.

The feasibility of performance feedback and coaching within OST programs is also a concern. All of the studies reviewed in this section included experienced coaches or coaches who received substantial training. Because OST programs operate with limited overhead and rarely have access to external university support, most programs do not have the resources to provide high-quality training and support (Hirsch et al., 2010). A specialized staff member with discretionary time is a luxury very few OST programs can afford. Thus, even if future researchers find support for the effectiveness of performance feedback and coaching on OST staff members' these programs may not be widely adopted because they are not feasible in this context.

Performance Feedback Only

Researchers have examined the effects of performance feedback without coaching. Barton and Wolery (2007) examined the effects of email performance feedback. In this study, the authors selected three undergraduate students majoring in special education. These individuals were selected based on their enrollment in the teacher preparation program, specific praise rates below 0.25 statements per minute, and less than three months of paid experience working with children. Participants received email feedback on their use of expansions, a more sophisticated restatement of a child's

response, after the researcher completed a classroom observation. No feedback on specific praise was provided. Unsurprisingly, the implementation of the email feedback procedure did not alter specific praise rates for any of the three participants. A follow-up study, published in the same article, was conducted with graduate students completing their master degrees in the special education program. In this study, feedback on specific praise rates was included in the weekly email. Praise rates for two of the three participants in this study increased from approximately 0.33 specific praise statements per minute during baseline to around one per minute during the feedback condition. There was no clear effect of the intervention in the third participant's data. A functional effect was not evident between email feedback and specific praise rates in either study.

Rathel et al. (2014) also examined the effects of performance feedback shared via email on the frequency of positive interactions, negative interactions, and behavior specific praise using a multiple-baseline design across teachers. Four certified teachers in their first year of teaching participated in the study. Each teacher was observed by the researcher, who also provided technical support and some coaching to participants, during 15 minutes of instruction. The results from this observation were then emailed to the teacher who was instructed to complete a brief self-reflection, set an improvement goal, and complete a journal entry after reviewing the data. The emailed feedback contained a graph of the teachers' ratio of positive to negative interactions, frequency of specific praise statements, corrective feedback, and praise for meeting the established criterion. The criterion for behavior specific praise was not specified, but teachers were instructed to maintain a ratio of five positive interactions to every negative interaction. A

brief training on the importance of positive feedback and the interpretation of performance feedback was delivered by the researcher prior to the delivery of the first email containing feedback.

After training and the initial performance feedback, three of the four teachers increased their use of specific praise for academic behaviors. All three teachers had near zero rates of specific praise for academic behaviors during baseline and then increased to approximately one specific praise statement per minute with feedback. The fourth teacher never met the positive to negative ratio of 5:1 during the intervention phase and her praise rate remained very near her baseline rate. Consistent level changes were evident in the data for three teachers, however one of these teachers was placed on administrative leave seven days after the start of the intervention phase. Her early withdrawal resulted in four data points during the intervention phase. The limited data collected from this teacher and the lack of a clear demonstration of the effect with the fourth teacher are reasons to cautiously accept these results. The authors indicated, "Although our results are promising, we urge caution interpreting them until further replication occurs using rigorous single-subject designs with intervention results apparent across at least three participants" (Rathel, et al. 2014, p. 229).

Another important concern is how frequently the researcher provided coaching and assistance to participants. Rathel et al. (2014) reported that the teachers were informed that they could ask questions via email or in person, but they do not report how often the researcher was contacted for help nor do they clarify if this support was offered before or after the daily observations conducted by the researcher. As discussed

previously, the amount of training and support that must be provided to collect and generate email performance feedback in addition to the training required to help staff members effectively interpret and respond to this feedback is not a trivial issue. More research is needed to determine the effectiveness of performance feedback and to evaluate its feasibility in OST programs.

Self-Management

One variation of coaching and performance feedback that may be particularly useful and feasible in OST programs is self-management. Self-management refers to strategies used by individuals to alter their own behavior (Shapiro & Cole, 1994; Young et al., 1991). The term self-management is often used interchangeably with self-regulation, self-control, and self-determination (Hoff & Sawka-Miller, 2010; Long & Williams, 1977). Interest in the application of self-management strategies to improving teaching has persisted for nearly 50 years (Kilbourn, 1991; Long & Williams, 1977), but no systematic reviews of the literature have been published and high-quality empirical studies are uncommon (Kalis et al., 2007; Silvestri, 2004; Simonsen et al., 2013).

The components of self-management programs include self-monitoring, self-instruction, and self-evaluation (Hoff & Sawka-Miller, 2010; West, 2005). Self-monitoring involves observing one's own behavior and documenting instances of a pre-selected target behavior. Self-instruction is an antecedent intervention employing the use of repeated vocal or sub-vocal scripts to prompt a target response. Finally, self-evaluation involves making a qualitative judgment about one's own performance during a specified time period. Teachers conducting self-evaluations typically review their use of target

behaviors during a lesson and then select a rating that summarizes their performance (e.g., My use of specific praise during math was a 6 out of 10).

Unfortunately, very few experimental studies have been conducted using the aforementioned self-management procedures. Studies that focus on the effects of self-management procedures on praise with experienced and inexperienced teachers are reviewed in the next section. Self-management strategies are uniquely appealing to OST programs because they do not require the extensive support and personnel employed to use coaching and performance feedback.

Self-monitoring. Self-monitoring strategies for teachers involve procedures where teachers observe their own behavior and document the frequency of a target behavior. Self-monitoring is the most common self-management strategy studied in the research literature (Hoff & Sawka-Miller, 2010) and its effects have been studied in relation to improving the use of curriculum-based assessment (Allinder & Beckbest, 1995; Allinder, Bolling, Oats, & Gagnon, 2000), enhancing data-based decision making (Belfiore & Browder, 1992; Browder, Liberty, Heller, & D'Huyvetters, 1986), improving the implementation of evidence based practices (Burgio et al., 1990; Lylo & Lee, 2013; Oliver, Wehby, & Nelson, 2015; Petscher & Bailey, 2006; Plavnick, Ferreri, & Maupin, 2010), and increasing teachers' use of praise (Kalis et al., 2007; Simonsen et al., 2013).

Simonsen et al. (2013) examined the effects of three different approaches to self-monitoring to improve teachers' use of specific praise. Five experienced teachers participated in the study. Three of the five teachers had earned master's degrees and their years of experience in the field ranged from 3-28 years. The three self-monitoring

strategies included in the study were making a tally mark on a clipboard, using a small handheld counter, and using a rating system to document the frequency of specific praise during instruction. The authors employed an alternating treatments design to identify the most effective intervention for each participant. The researchers then selected an optimal intervention for an extended intervention and maintenance phase. The handheld counter was selected as the optimal condition for 60% of the participants and the tally system was selected for the other 40%. The teachers' use of specific praise increased during the alternating treatments phase, but high variability and substantial overlap between the rates observed in this phase limited the conclusions that can be drawn from the study.

Two of the five participants emitted increased specific praise rates using the optimal self-monitoring method. For these teachers, their use of specific praise increased from near zero levels during a brief baseline phase to an average of around two specific praise statements per minute. These procedures have the benefit of being very applicable to OST program staff, but the lack of a clear consistent effect suggests that further study is warranted on self-monitoring strategies. In addition, the study featured experienced instructors with specialized training. It is unknown if the knowledge and skill they had acquired in their professional training were necessary to benefit from self-monitoring.

Kalis et al. (2007) evaluated the effects of self-monitoring on a new teachers' use of praise in a self-contained behavioral unit. The participant was a first year teacher with a certification in nonviolent crisis intervention. During baseline, the teacher's praise rate was zero with no variability in the data across four 10-minute observation periods. To introduce the self-monitoring intervention, the researcher provided a 20-minute training

on identifying general and specific praise and completing the self-monitoring procedure with a hand counter. Prior to the research session each day, the researcher met with the teacher to identify the goal for the day and then collected materials from her after the session. Her rate of specific praise during the self-monitoring procedure increased to approximately 0.6 statements per minute. These results should be considered with caution because the researcher did not employ an experimental design to control for other variables outside of the intervention that might have influenced the participant's responding. In addition, the researcher met each day with the participant before and after the research session to review performance and set goals. These procedures are more similar to coaching and performance feedback than they are to self-monitoring. Unfortunately, the lack of experimental control and the use of a packaged intervention including coaching makes it impossible to determine precisely what effect the self-monitoring component had on teachers' use of specific praise.

Finally, Silvestri (2004) examined the effects of self-monitoring on teachers use of specific praise. Four certified teachers participated in the study. One of the teachers had earned a master's degree and the others had all completed bachelor's degrees. On average, teachers worked at the school for 2 years prior to the study. During the baseline and intervention phases, teachers were audio or video recorded during 25 minutes every day. Audio recordings were coded for the frequency of general praise, specific praise, and negative statements. During the intervention phases, teachers were instructed to review a 5-minute audio segment and count the number of positive and negative statements they made during instruction that day. Next, they graphed their data and set a goal for the next

day. Baseline rates of specific praise were near 0 for all participants. For two teachers, their specific praise rates increased slowly after the start of the self-monitoring intervention. Both teachers ended the self-monitoring phase with specific praise rates of approximately one statement per minute. Unfortunately, these procedures did not affect the other two teachers' rates of specific praise. In an attempt to increase the potency of the procedure, the experimenter gave one of the teachers a MotivAider, a tactile prompting device, to prompt her use of praise during instruction. Prompts were delivered every 25 seconds; after each prompt, the teacher was instructed to scan the room and praise a student who was behaving appropriately. Prior to the use of the MotivAider, the teacher's use of specific praise was very near 0 and after implementation the rate increased to 0.5 specific praise statements per minute.

Similar to the effects found with email performance feedback, the three self-monitoring studies produced inconsistent effects on general and specific praise rates. These inconsistent effects and the lack of experimental control in some studies make it difficult to come to any firm conclusion about the effectiveness and feasibility of these procedures. However, the appeal of self-monitoring is its ease of implementation and potential applicability in a variety of contexts. This feature is particularly appealing in OST programs because personnel are often not available to sustain interventions that require extensive monitoring.

Self-instruction. Self-instruction has not been studied as an approach to helping teachers increase their use of specific praise. One strategy that has been explored and shares similarities with self-instruction as an antecedent intervention is the use of audio

or visual prompts. Van Houten and Sullivan (1975) used prompting to increase the praise rates of three certified teachers. On average, general praise rates during baseline were below one statement per minute. Teachers were instructed to praise a student every time they heard the audio prompt, delivered on a variable schedule every 30 seconds, over the intercom.

The teachers increased their praise rates to two or three statements per minute during the intervention phase. Rates remained at or above these levels when after fading the prompts to every 3 minutes and then completely fading prompts. The research assistants remained in the classroom every day to collect data after the audio prompts were discontinued. The presence of these individuals in the classroom may have provided a sufficient prompt to maintain high praise rates.

Although audio prompting was effective in this study, further research is warranted to determine if similar results would be found with inexperienced instructors and more complex behaviors like specific praise. While experienced educators are more likely to be familiar with specific praise and the related instructional skills (e.g., monitoring student behavior, stating expectations, describing behavior, etc.), inexperienced teachers may not have acquired these skills and may not respond as effectively to this type of intervention.

Self-evaluation. Two studies were conducted on the effects of self-evaluation on teachers use of praise (Keller et al., 2005; Sutherland & Wehby, 2001). Sutherland and Wehby randomly assigned 20 experienced elementary school teachers to either a self-evaluation or no-treatment group. Teachers in the self-evaluation group were trained to

create an audio recording of the instructional session and then review the recording to calculate their praise rate. Prior to reviewing the audio recording, teachers made a prediction about how many praise statements they used during the recorded session. Teachers compared their actual praise rate to their prior prediction and then set a goal for the next session. Every day the teachers updated their graphs with their new praise rate data collected that day. Finally, teachers were instructed to vocally or sub-vocally evaluate their performance at the end of the process and provide self-reinforcement when appropriate.

Sutherland and Wehby (2001) reported that the frequency of praise increased from pre-treatment to treatment for both the self-evaluation and control groups. The increase in praise observed in the self-evaluation group was significantly higher ($d = 2.0$) than the change observed in the control group ($d = 1.0$). These results apply only to the aggregate of general and specific praise data. The difference between groups for specific praise rate was not statistically significant.

Keller et al. (2005) replicated Sutherland and Wehby's (2001) self-evaluation strategy with interns placed in local schools while completing their teacher preparation program. The interns increased their specific praise rates from near zero in baseline to approximately one specific praise statement per minute during the intervention phase. During a follow-up maintenance phase occurring a month after the end of the intervention phase, one participant maintained a slightly higher specific praise rate with an increasing trend, while the other two participants exhibited declining trends in their use of praise.

In both of the studies reviewed here (Keller et al., 2005; Sutherland & Wehby, 2001), the researchers identified a relationship between the use of self-evaluation and praise. The Keller et al. study is particularly applicable to OST programs because the study was conducted with interns in typical education settings. Although the durability of the intervention effects is an area of concern in the Keller et al. study, the immediate change in specific praise rates suggest the collection of performance feedback, completing a self-evaluation of performance, and the graphing of these data were all critical components of the intervention.

Implementing these procedures in OST programs may prove to be challenging, however. Teachers in these self-evaluation studies spent considerable time evaluating their use of praise, setting goals, listening to their recorded audio tracks, graphing their data, and planning their strategies for the following day. This was facilitated by the flexibility in the participating interns' schedules. Teachers in typical schools or OST programs may not have the ability to allocate enough time for these activities every day after school. Teachers in OST programs may have even less flexibility since preparation time is not built into their schedules and many programs cannot afford to allow teachers to take time away from direct service with students for data analysis, self-reflection, and instructional improvement (Asher, 2012).

The self-management procedures used in the reviewed studies produced inconsistent effects on specific praise rates. In the self-monitoring studies (Kalis et al., 2007; Silvestri, 2004; Simonsen et al., 2013), none of the studies demonstrated a strong, functional relationship between the use of self-monitoring and specific praise rates. This

finding was true even though all of the teachers were experienced and credentialed. The results of the Van Houten and Sullivan (1975) study indicated that experienced teachers responded to the use of a frequent audio prompt to praise during instruction. Finally, the effects of self-evaluation on specific praise were not significant in an experimental group study (Sutherland & Wehby, 2001) and two of the three participants in the single-subject design study exhibited counter therapeutic trends during a maintenance phase (Keller et al., 2005). Although self-management procedures have the potential to be more feasible within the structures and constraints of typical OST programs, the current procedures examined in the empirical literature have produced inconsistent effects on teachers' use of specific praise. More research is required to determine whether or not these strategies can be effective with typical OST staff members.

Matching

Because of the inconsistent effects across the self-management studies reviewed above, it is worthwhile to examine how self-evaluation with teachers differs from self-evaluation with students. This comparison is worthwhile because student self-evaluation interventions have consistently produced significant effects on a variety of target behaviors and have been identified as an evidence based practice (Joseph et al., 2012; Maggin, Briesch, & Chafouleas, 2013; Sheffield & Waller, 2010). Student self-evaluation programs often employ self-observation, self-instruction, and self-evaluation. In addition, student self-evaluation programs use a matching procedure to encourage accurate self-evaluations (Hoff & Sawka-Miller, 2010; Peterson et al., 2006; Smith et al., 1988; Young

et al., 1991).

Matching involves comparing self-evaluation ratings against ratings from an external observer, typically a teacher. This procedure sensitizes students to the variables the teacher is using to rate their behavior. Achieving alignment between student and teacher ratings is a critical feature of many self-management programs; however, similar procedures have been altogether ignored in self-management programs for teachers. Although this may be justified with experienced teachers, accurate self-evaluation may be difficult for inexperienced instructors who need not only feedback on their performance but also evaluative feedback that helps them understand what they are doing relative to what they should be doing.

Smith et al. (1988) used the matching procedure with adolescents receiving special education services in a traditional junior high school. Direct observation was used to determine the percent of intervals participating students were off-task or disruptive using a 10-second partial interval data collection system during an 80-minute class period. A withdrawal design was used to investigate the effects of the intervention, which consisted of a class wide self-management program.

The self-management program involved the explicit teaching of expectations, students completing a self-evaluation card, teacher feedback, and a token economy in which students earned points for accurate self-evaluations and maintaining appropriate behavior in the classroom. To complete the self-evaluation, students rated their performance on a 6-point scale with ratings ranging from unacceptable (0) to excellent (5). Points were awarded in increasing amounts with each rating level but only if the self-

evaluations were also accurate. Accuracy was defined as agreement with the teacher. Self-evaluations were matched with teacher ratings periodically and all incentives were earned by achieving consistent alignment between the two raters. For example, if the two ratings were identical the student would receive a bonus point. If the two ratings were discrepant by just one point (e.g., a student rating of excellent [5] and a teacher rating of very good [4]), then the student would receive the number of points corresponding to the teacher's rating. During the intervention phase, the students and teachers completed the rating procedure every 10 minutes. The frequency of matching was reduced in subsequent fading phases from every 10 minutes to every 15 minutes and eventually to once per day.

The four target students emitted a high percentage of off-task and disruptive behavior during baseline, ranging from 62-74% of intervals. During the intervention, the behavior of target students was coded as off-task or disruptive during 0-15% of intervals, an average reduction in problem behavior of 55%. These performance improvements in the training classroom were maintained throughout the fading of the intervention procedures. Thus, after students learned to accurately evaluate their own behavior, based on the perceptions of their teachers, daily self-evaluations were effective at suppressing problem behavior.

In another study, Peterson et al. (2006) examined the effectiveness of the matching procedures with middle school students. Five students were taught to complete self-evaluations and then match these ratings with teacher evaluations in a special education classroom. Initially, all of the students completed the self-evaluation process

four times during the 45-minute class period. They earned points for high scores on self-evaluations and additional bonus points for matching. Points were exchanged for incentives from a reinforcement menu developed by the students prior to training. Once students consistently matched their self-evaluations with teacher ratings, they were asked to complete the same procedures in one of their general education courses. Points could then be earned for appropriate behavior and matching in both the special education and general education classroom. Each time consistently high student ratings matched similar ratings from their teachers on five consecutive days, a new classroom was added.

All five participating students increased their use of targeted social skills and reduced their off-task behavior in the training context. In addition, these improvements were replicated in at least four different general education classrooms for all five participants. While all five participants exhibited substantial reductions in problem behavior, for two of the five students their off-task behavior reduced to levels near zero. Pre-post teacher ratings concerning the effectiveness of the procedures indicated that general education teachers noticed positive changes in the social behavior, academic engagement, and attitudes toward school for all five participants. The consistency of these effects in a wide variety of contexts suggest that this approach to self-management is a potent strategy for behavior change.

The studies reviewed in this section used matching to encourage accurate self-evaluations. These studies incorporated matching as a critical feature of the independent variable and produced consistent and practically significant effects on the target behaviors in adolescents. To date, researchers have not examined the effects of self-

management with matching on teachers' use of praise. Given the need for more potent self-management procedures for teachers, the use of matching may be one way to enhance the effectiveness of these procedures.

It is important to note that these procedures have only been examined with students in traditional educational settings. There is no evidence to suggest that matching would be effective when incorporated into a self-management program for inexperienced OST teachers. The contingencies and controlling variables that were present in the self-management studies conducted with adolescent students are different than the variables that influence the behavior of OST staff members. For example, teachers control access to a wide variety of reinforcers. A teacher's ability to manipulate the explicit and implicit reward contingencies in a classroom magnifies the importance of their feedback to students. Similar contingencies may operate with a teacher and a supervisor, but it is difficult to predict how these contingencies would enhance or diminish the effectiveness of the matching procedure. Additional research is necessary to clarify these issues.

In addition, the matching procedure adds additional complexity to the self-management intervention. In the context of an OST program, the matching procedure requires a skilled evaluator to provide the external rating. If external management and coaching is required to make self-management effective, then all the practical benefits of self-management relative to performance feedback and coaching are removed. One possible solution to this problem would be to use students as external evaluators for OST staff members. Students are already in the classrooms with OST staff members and they have extensive experience observing the instructional practices of their teachers.

Student Feedback

Students are an important and underutilized resource for school or program improvement (Downer et al., 2015; Fauth et al., 2014; Fraser & Walberg, 1981; Gray et al., 1974). In a report from the Measures of Effective Teaching Project funded by the Bill and Melinda Gates foundation, the authors wrote,

> No one has a bigger stake in teaching effectiveness than students. Nor are there any better experts on how teaching is experienced by its intended beneficiaries. But only recently have many policymakers and practitioners come to recognize that—when asked the right questions, in the right ways—students can be an important source of information on the quality of teaching and the learning environment in individual classrooms. (Measures of Effective Teaching, 2012a, p. 2)

Fraser and Walberg (1981) echoed many of these ideas when they articulated the four benefits of collecting and responding to student feedback. First, the costs of collecting student feedback are significantly less than those associated with external observers. Second, student feedback is based on days and weeks of observation within the classroom whereas external observers see only a small portion of the instruction during brief observations. External observers have no way to know if the instruction they are witnessing during the observation is representative of typical classroom practices. Third, student feedback can be collected from an entire class, thus aggregating the experiences and perspectives of many observers as opposed to a single observer. Fourth, facilitating student learning and behavior change is the purpose of instruction, so if student perceptions of instruction and the classroom learning environment influence their behavior then these perceptions are worth studying.

Researchers and instructors in higher education, however, have expressed a

variety of concerns about the reliability, validity, and utility of student feedback (Aleamoni, 1999; Greenwald, 1997; Zhao & Gallant, 2012). The reliability of student feedback is a concern if students are unduly influenced by variables other than the quality of the instruction in their classrooms (Aleamoni, 1999). If students base their ratings on their final grade, the popularity of the instructor, the attractiveness of the instructor, the time of day the course is taught, class size, or the gender of the instructor then their feedback is potentially unreliable and biased. Concerns over the validity of student feedback often stem from the belief that students lack sufficient understanding of the content, life experiences, or appreciation of the discipline to provide useful feedback (Zhao & Gallant, 2012). This concern is problematic because it speaks to the value of the feedback to the instructor. In other words, do students really know what will help them learn? Does student feedback correlate with other indicators of effective teaching? Finally, many instructors question how to use student feedback to make meaningful, lasting improvements in their teaching (Greenwald, 1997). If student feedback just reflects the idiosyncratic values and interests of the current class, does it make sense to alter instruction for future classes based on these data? This concern addresses not only the substance of student feedback, but the procedures, timing, and availability of the data to teachers. Although these concerns have been put forth primarily by instructors and administrators in higher education there is no reason to believe that teachers in elementary or secondary schools would not share their concerns as well.

Reliability and Validity of Student Feedback

Research on the validity and utility of student feedback from students in K-12 has

only recently become available (Downer et al., 2015; Fauth et al., 2014; Follman, 1992, 1995; Measures of Effective Teaching, 2012b). Prior research on student feedback was almost exclusively conducted in higher education (Balch, 2012). Follman (1995) wrote a narrative review of the sparse literature on student feedback in elementary schools. Six studies were conducted on student feedback concerning specific instructor behaviors in the five decades prior to Follman's review. In two of these studies researchers compared administrator, expert, and student feedback and found that student feedback was related to other more traditional measures of teaching (Cook & Leeds, 1947; Tuckman & Oliver, 1968). For example, Cook and Leeds (1947) reported that student feedback correlated with principal's ratings of teachers (r = .39) and with expert raters (r = .33).

Contemporary interest in the use of student feedback may be the result of high-profile projects like the Measures of Effective Teaching (MET) project funded by the Bill and Melinda Gates Foundation. The MET project has engaged district and state partners in a comprehensive effort to examine the validity, reliability, and utility of measures of teaching quality including student feedback (Measures of Effective Teaching, 2012a). The MET team conducted an analysis of just over 2,500 classrooms with student feedback and student achievement measures. Student feedback was collected using the Ferguson's (2012) Tripod survey. The Tripod survey is organized into seven domains including care, control, clarify, challenge, captivate, confer, and consolidate. At least two questions of the 36-item survey are mapped to each of these domains. The overall interrater reliability of the aggregate Tripod score for all teachers in the sample was 0.67. This statistic was calculated by comparing the relationship between student feedback

collected from different students in multiple classes taught by the same instructor.

The MET research team and many contemporary researchers focused on evaluating the quality of student feedback as it relates to student achievement (Downer, 2015; Follman, 1995). Student achievement in this study was measured by the SAT-9 for reading and the Balanced Assessment of Math (BAM). A value-added measure was calculated for both the reading and math assessments to determine how much growth individual students made based on their pre- and posttest scores. The correlation between the Tripod summary score and the aggregate value-added measures was 0.23 for reading and 0.11 for math. These correlations were both statistically significant, but the small magnitude of the correlation suggests the relationship between student feedback and student achievement was small. Although the MET project concluded that student surveys were valid, reliable predictors of student achievement, the relationships they reported are not practically significant for teachers.

One possible explanation for the small magnitudes of the correlations identified in the MET studies is the assessment of constructs (i.e., care, clarify, control, consolidate, challenge, captivate, etc.) rather than discrete, specific instructional practices. Follman (1994) indicated;

> Students can properly, objectively, reliably, and perhaps, validly report on such descriptive matters as events which occur in their class as well as their teacher interactions. On the other hand, most students would not be qualified to evaluate their teacher's content knowledge, choice of pedagogy, except, perhaps, for outlier, extreme behaviors or positions. (p. 59)

Although the MET project assessed student perceptions of constructs rather than specific teacher behaviors, the MET researchers also made several valuable recommendations

about the process for collecting student feedback (Measures of Effective Teaching, 2012b). First, the process for collecting student feedback should be transparent. Teachers will be more likely to accept and benefit from these data when they understand how the data are collected and reported. Second, student surveys need to be designed to allow students to make age-appropriate discriminations. For example, student surveys should be short, include age-appropriate items, and ask questions about things students would have experienced in the classroom (Measures of Effective Teaching, 2012a).

Downer et al. (2015) conducted a study examining the reliability and validity of a student survey used to measure the quality of the classroom instruction. Over 500 students in the fourth and fifth grades in a mid-Atlantic public schools completed the survey. The CLASS-SR survey contained 62 items assessing students' experiences relative to the availability of emotional support, classroom organization, and instructional support. The vast majority of the items asked students to report on their personal experiences with specific instructional events in the classroom, while a smaller number of items focused on general perceptions of the quality of instruction. The authors indicated that this was important because "at this developmental level, students may be cognitively predisposed to focus on their own individual experiences with the teacher and classmates and are not able to take a "bird's eye view" of the classroom and provide a more abstract report on general features of the classroom interactions outside of their personal experience" (Downer et al., 2015, pp. 744-755).

Prior to the collection of the student feedback, the participating teachers recorded several lessons. Trained observers then coded these videos using the CLASS-UE

observation protocol, a parallel form of the CLASS-SR. This observation protocol allows observers to note specific events related to emotional support, classroom organization, and instructional support. Teachers also submitted a summary of their students' reading achievement scores and the number of office discipline referrals in their classrooms. hierarchical linear modeling (HLM) was used to examine the associations between student feedback, observer evaluations, and academic and behavioral data. Although individual student feedback scores were not consistently related to any of the dependent variables, student feedback aggregated at the classroom level was significantly associated with observers' ratings of classroom interactions for emotional support, classroom organization, and instructional support (unstandardized estimates = 3.59, 1.11, and 2.14, respectively, p = .001, .05, and .001, respectively). Aggregate student feedback regarding classroom organization was associated with student achievement (unstandardized estimates = 158.69, p = .01) and a reduction in office referrals (unstandardized estimates = -1.6, p = .01). Aggregate student feedback scores for emotional support were similarly related to achievement and discipline data.

Fauth et al. (2014) conducted a similar study that addressed the predictive validity of student feedback in elementary schools. Fauth et al. gathered data from 1,556 German students in the third grade. These students completed a brief 21-item assessment of the instructional conditions in their classrooms. This assessment addressed instructional quality in three areas including creating a supportive climate for learning, functioning classroom management program, and system to encourage cognitive activation. The students also completed a measure of teacher popularity along with a pre- and posttest

measure of content knowledge on a science unit assessment. The relationships between these variables were examined using HLM. In the final model, teacher popularity was included in the model to control for its potential influence on student feedback. Teacher popularity (standardized estimate = .30, $p < .05$) and student feedback on classroom management (standardized estimate = .41, $p < .05$) were both significantly related to the posttest scores for the science unit. The authors note that it is difficult to separate many characteristics of effective teaching (e.g., praise, clarity of instructions, supportive feedback, etc.) and popularity. It is also important to note that the classroom management questions were more discrete and specific than the items focusing on cognitive activation and supportive climate, a finding in line with prior work suggesting student feedback should be event-based rather than qualitative (Follman, 1995).

Kunter and Baumert (2006) examined the construct and criterion validity of student feedback. They compared student feedback with teachers' self-evaluations of instructional quality. These authors used the same conceptualization of instructional quality as Fauth et al. (2014). Nearly 300 teachers participated in the study from elementary and secondary schools in Germany. Teachers and students completed a 28-item assessment of the quality of their mathematics instruction. Student feedback, aggregated at the classroom level, was moderately correlated with teacher self-evaluations on items associated with classroom management ($r = 0.65$, $p < .05$), experiences of cognitive autonomy ($r = .30$, $p < .05$), and perceived social support ($r = .38$, $p < .05$). However, significant correlations were found in less than half of the comparisons between teacher and student ratings. There were no statistically significant

relationships between teacher self-evaluations or student feedback and the PISA measure of student achievement used in the study. There is no clear methodological problem that fully explains the lack of a relationship between student achievement and student feedback found in this study. The finding that student reports and teacher reports do not consistently correlate is consistent with prior research indicating that teacher reports and that student feedback are differentially related to one another and to student achievement (Balch, 2012; Mayer, 1999; Porter, 2002).

Finally, De Jong and Westerhof (2001) compared student feedback from 1,084 middle school students with classroom observations conducted by trained observers in 49 teachers' classrooms. A measure of student motivation was collected along with a measure of mathematics achievement to assess validity. Observers rated participating teachers using a 40-item checklist mapped to specific direct instruction strategies. Ratings were completed during three different 45-minute lessons. Student feedback was collected using an instrument developed for the study and aligned with the tool used by the observers. The authors noted that they designed this tool because, "we were not only interested in how students perceive the learning environment as such, but in the question of how they perceive the actual behavior of teachers" (De Jong & Westerhof, 2001, p. 58). The students and external observers both noted the occurrence and frequency of specific teacher behaviors. For analysis purposes, these variables were aggregated at the classroom level and a factor analysis was used to group items into scales.

The validity of the observer scales was compared to that of the student feedback scales using a variety of tests. An analysis of discriminant validity was conducted by

examining the intercorrelations between the student feedback scales and comparing them to the intercorrelations between the observer scales. The student feedback scales had fewer significant intercorrelations with lower magnitudes than the observer scales, thus indicating higher discriminant validity. Congruent validity was examined by comparing student feedback with observer ratings for the same instructor. The authors found significant correlations between student reports on the motivation and clarity items and observer scores on promoting insight ($r = .38$, $p < .01$), between students' reports on the Teaching Skills items and observers reports of promoting positive behavior ($r = .29$, $p < .05$), and student feedback on group control, or classroom management, and observers' ratings of promoting positive behavior ($r = .34$, $p < .05$). Student feedback on group control was also significantly correlated with the calling out items ($r = .37$, $p < .01$). Items in the calling out group included questions about the teachers' use of student engagement strategies like high rates of group choral responding or providing corrective feedback for incorrect responses.

Finally, the authors evaluated the predictive validity of the observer and student scales by correlating these scores with a measure of student motivation to learn and a measure of mathematics achievement. All of the student feedback scales, except the checking homework scale, correlated significantly with student motivation (range = .29 - .82). Two student feedback scales, motivation and clarity ($r = .30$) and checking homework ($r = .40$), were significantly correlated with mathematics achievement. By contrast, none of the observer scales correlated with student motivation and only the "clear, task-oriented and positive behavior" scale correlated with mathematics

achievement. A multi-level model was calculated to further examine predictive validity and revealed that the observer and student feedback scales for clarity of instruction were both significantly associated with mathematics achievement.

Several interesting themes regarding student feedback are evident from the available research. First, student feedback is most relevant and predictive of critical school outcomes when it is aggregated at the classroom level. Second, researchers who requested student feedback on more discrete teaching behaviors observed stronger, more consistent relationships between student feedback and outcomes (De Jong & Westerhof, 2001; Downer et al., 2015). Asking young students to provide feedback on more global domains of instructional quality may violate the recommendation from the Measures of Effective Teaching project to ask students to make age-appropriate discriminations. Finally, in three studies, researchers found statistically significant relationships between student feedback on classroom management and student achievement (De Jong & Westerhof, 2001; Downer et al., 2015; Fauth et al., 2014). Although there is emerging evidence that students in elementary schools, specifically grades 3-5, can provide valid and meaningful feedback for teachers the research reviewed in this section does not answer the question of whether or not teachers can improve their performance with access to student feedback.

Utility of Student Feedback

To date, there are no research studies in which researchers systematically examine the use of student feedback within the context of an intervention to change teacher behavior. That is not to say that students have not been integrally connected to an

intervention to change teacher behavior. Gray et al. (1974) in their seminal article, "Little Brother is Changing You," described a series of case studies where students altered the behavior of their teachers by managing the consequences of their teacher's behavior. For example, Jess was taught to look away and not to speak with his teacher when his teacher said or did something that Jess perceived as critical or negative. Similarly, Jess smiled and talked with his teacher when his teacher praised him, acted in a friendly way, or did something he felt was positive. By engineering the consequences of his teacher's behavior, Jess experienced more frequent positive contacts and felt more connected to his classroom. Gray et al. reported that students initially reported praise rates well below the praise rate reported by trained data collectors. Thus, the researchers had to define praise and model its use in an effort to reduce this discrepancy. After training, the discrepancy between reported praise rates disappeared.

The only study to address the effects of student feedback on teacher behavior was conducted by Gaertner (2014). Gaertner's study examined the effects of student feedback on self-reported changes in instruction. Over 300 teachers were recruited to participate in the study from a pool of over 1,100 teachers who had collected student feedback using a voluntary online survey tool available in two districts in Berlin and Brandenburg, Germany. Teachers who volunteered to participate in the study completed a survey about their analysis and use of the student feedback. The results from the survey indicated that teachers noticed a high degree of alignment between student feedback and their own self-evaluations of their teaching. They also reported only slight agreement with statements about changes in their classroom practice. For example, teachers reported agreement with

the statement, “I strengthened methods that proved to be successful,” and “I pay more attention to specific behaviors (e.g., be more structured).” The finding that self-evaluations and student feedback were aligned across a majority of teachers is inconsistent with previous research (Mayer, 1999; Porter, 2002; Wilkerson, Manatt, Rogers, & Maughan, 2000). This may be a result of the use of self-report as the measure of alignment and the desire to appear more capable as an instructor. It may also reflect the use of convenience sampling and self-selection to identify participants for the study. Many of the teachers elected to use student feedback frequently prior to the study. As a result, the pool of participants may have been limited to those who received positive feedback from their students on their initial survey. These individuals would be more likely to report alignment between their feedback and their instruction and to have experienced less critical feedback. Less critical feedback may also have negatively impacted their perceptions of the changes they experienced in their classroom practices.

These researchers suggest that students can play a valuable role in shaping the learning environments they need to be successful. Previous researchers confirmed that students can discriminate between good and bad instruction and that their perceptions of teacher quality relate to student achievement and engagement. Student feedback is more accurate when these discriminations are aggregated and reported at the classroom-level as opposed to viewed as individual student responses. Unfortunately, too few studies have been conducted on the effects of student feedback within interventions designed for teachers. The lack of empirical intervention studies may reflect the reality that many traditional schools have sufficient resources to provide coaching and performance

feedback to their teachers so they do not have a need to explore more experimental programs. This is not true for OST programs, however. These programs need low-cost, high-efficiency approaches to professional development and student feedback has yet to be systematically incorporated into effective staff development interventions, including self-management programs accessible to OST programs.

Summary

One critical instructional strategy that enhances both participation in classroom activities and student performance is specific praise. OST programs could benefit from support strategies designed to help inexperienced instructors increase their specific praise rates. Unfortunately, the strategies that have been widely studied and found effective with experienced teachers, coaching and performance feedback, have not been studied in OST programs nor have they been consistently applied to helping novice instructors improve. Coaching and performance feedback may not be practical in many OST programs because these strategies require access to additional personnel, extensive training, and knowledgeable staff to provide data collection, analysis, and recommendations.

Self-management is an appealing alternative to these procedures because these strategies are designed to be implemented by the target individual with very little external oversight and support. In addition, in many self-management studies, researchers employed very little specialized training prior to implementation (Kalis et al., 2007; Silvestri, 2004). Although there are very few applications of self-management that have been studied with inexperienced instructors and the results from these studies are often

mixed, these strategies have the potential to be both feasible and effective through increasing staff members' ability to self-evaluate accurately. The immediate yet often not maintained performance improvements highlighted in several of the reviewed studies (Kalis et al., 2007; Silvestri, 2004; Van Houten & Sullivan, 1975) suggests that self-evaluation interventions may require additional features to increase their potency.

One possible strategy for increasing the potency of self-evaluation programs is to incorporate a matching component to increase the accuracy of self-evaluation ratings, a strategy that has been well documented in the research literature and shown to be effective with adolescents (Hoff & Sawka-Miller, 2010; Young et al., 1991). There are no studies in which researchers examine the feasibility and effectiveness of matching as a component within a self-evaluation strategy designed for adult instructors, however there is very little evidence to suggest that the current approaches to self-management with adults have been consistently effective either. The feasibility of matching within a self-management program designed for OST programs could be enhanced if students were used as the external rater. As the primary consumer of OST programs, students have an incentive to encourage teachers to use praise and there are studies supporting the accuracy and value of student feedback for instructors (DeJong & Westerhof, 2001; Fauth et al., 2014; Follman, 1995).

Despite the promise of self-management for OST staff members, the reality remains that there have been no experimental studies of these approaches and only preliminary evidence to support their feasibility and effectiveness. The purpose of this study is to examine a novel approach to self-management using student feedback to

increase praise in the classroom. The benefit of this approach is providing a cost-efficient, potentially effective strategy that may help OST teachers increase their praise rates. If effective, this approach could also provide teachers an alternative to coaching or external observation to get feedback on their performance in the classroom, thus extending the growing literature on the utility and value of student feedback.

Primary Research Question

To what extent will teachers' specific praise rates change when they are trained to use specific praise and receive feedback on the degree to which their self-evaluations match students' reports of specific praise?

Secondary Research Questions

1. To what extent will teachers' general praise rates change when they are trained to use specific praise and receive feedback on the degree to which their self-evaluations match students' reports of specific praise?
2. To what degree will teachers indicate that the intervention was feasible and effective?
3. To what degree are students' reports of specific praise associated with teachers' use of specific praise?

CHAPTER III

METHOD

Participants

Teachers

Teacher recruitment. In consultation with the administration of the B&G Club, the experimenter recruited teachers from the FANZ program to participate in this program. The FANZ program is a summer program for students entering grades three through five during the following school year. This program was targeted for this study because it typically had the most consistent student participation of the B&G Club summer programs and the B&G Club administrators indicated most of the students had sufficient reading skills to participate in the intervention procedures. FANZ program administrators hire college students or early college graduates between 18 and 23 years old to serve as teachers during the summer program. None of the teachers hired for these positions have formal teacher preparation.

All five teachers in the FANZ program attended a recruitment meeting to learn more about the study. During the recruitment meeting, the experimenter explained the goals of the project, data collection procedures, expected benefits of participation, and risks. Teachers asked clarifying questions and expressed concerns prior to reviewing the letter of informed consent (see Appendix A). All five teachers consented to participate in the study.

Teacher selection. Since the intended outcome of the study was improving

teachers' specific praise rates, only teachers with specific praise rates below 0.4 statements per minute were selected to participate in the study. To determine how frequently the prospective participants used praise, teachers audio recorded their interactions with students during academically-focused activities over a period of three days. This selection process is identical to the procedures used by Briere et al. (2015). All five of the prospective participants had specific praise rates below the target rate. Unfortunately, shortly after completing the selection process one of the five prospective participants quit his job with the B&G Club.

Selected teachers. The four remaining teachers selected to participate in the full study included Brendan, Ginny, Lilli, and Kate. Although pseudonyms are used here, a random participant number was assigned to each teacher during the study to protect their anonymity during data collection. Brendan was 20 years old and the only male teacher who participated in the study (see Table 2). He was one of only three male teachers employed by the B&G Club throughout the summer program. Brendan had prior experience working for the B&G club and was seeking a dual secondary education and sports science degree, but had not yet begun teaching-oriented coursework in his

Table 2

Summary of Teacher Demographics

Name	Age	Race/ethnicity	Gender	Educational attainment
Brendan	20	White, non-Hispanic	Male	Some college, no degree
Ginny	18	White, non-Hispanic	Female	High school diploma
Lilli	19	White, non-Hispanic	Female	Some college, no degree
Kate	22	White, non-Hispanic	Female	Bachelor of Science Family, Consumer, and Human Development

program. Ginny was the youngest participant at age 18. She graduated from high school one month prior to the start of the B&G Club summer program. She planned to attend college in the fall after the end of the summer program, but had not declared a major. Lilli was 19 years old and had worked for the B&G Club during the after-school program the previous school year. In addition, she was a full-time student at a local university preparing to apply for the elementary education program. Finally, Kate was the oldest teacher at age 22. She recently completed a college degree in Family Consumer and Human Development, and had no prior training in classroom management or instruction. All four teachers identified their race and ethnicity as White, non-Hispanic.

Students

Students participated only as members of classroom groups and were not identified individually during data collection. Students are described in this section because they played an active role in the implementation of the intervention and their characteristics are relevant to understanding the context of the study. Parents of the students enrolled in the FANZ program received a letter of information regarding the study including a description of the intent of the study, the procedures in which their child(ren) would be invited to participate, and the data collection procedures. Parents were given one week from the date they received the letter of information to express their desire not to have their child(ren) participate in the study. In addition, parents could request to have their student removed from the study at any time. The B&G Club staff members did not receive any parent requests to have their students removed from the study.

Three hundred fifty-four students were enrolled in the FANZ program, but only a small group attended the program regularly. On average, the total number of days attended per student was 20 days (standard deviation = 19 days). Approximately 40 students were on site each day. Thus, the average student attended only two of the five days each week or 40% of the entire program.

Other Staff Members

The program coordinator and site supervisors at each site had a small role in collecting implementation fidelity data and providing technical assistance to the teachers. The program coordinator was responsible to oversee the site supervisors, ensure that specific programs were implemented, and coordinate transportation between sites. The program coordinator was hired in 2012. She had a bachelor's degree in child development from a local university and had prior experience working with local non-profit agencies. Due to unforeseen circumstances, the program coordinator also served as a site supervisor. Site supervisors were responsible for overseeing the daily operations of the summer program, establishing the weekly schedule, running student activities, and communicating with the program coordinator.

Setting

The B&G Club operated summer programs at three elementary schools in Northern Utah. Each site served students in the FANZ program, had at least one participant in the study, and was located in a different city. The sites were separated by approximately 25 miles. The South site was housed in a former elementary school. The

school was converted into a community center approximately 10 years prior to the start of the study. Half of the building housed a local museum and community resource center and the other half was exclusively used by the B&G Club. A locked door separated the museum from the B&G Club. The portion of the school used by the B&G Club included the gym, the playground, a hallway with eight small classrooms, two sets of bathrooms, and two offices. The classroom assigned to the FANZ program at the South site was approximately 20 feet by 35 feet. The room contained two small group tables, a cabinet for materials, a couch, and a television. Brendan was assigned to the South site.

The summer program at the East site was held in a local elementary school. Activities at the East site were held in the gym, library, music room, and two large classrooms. The school's playground was on a city park adjacent to the school with multiple basketball courts, a volleyball court, and a pavilion. The classroom assigned to the East site FANZ program was a large classroom measuring approximately 40 feet by 40 feet. The room contained five large group tables with six chairs at each table. A white board was attached to the north wall and a sink on the east wall. Lilli and Kate were assigned to the East site.

The summer program at the West site was operated in local public schools with the permission of the school district. Staff members held programs at the West site in the gym, a couple of classrooms, outside on the playground, and in a large common area adjacent to the cafeteria. The FANZ program at the West site used the common area adjacent to the cafeteria for research sessions. This space was approximately 90 feet by 35 feet, contained 10 small group tables with six chairs at each table. Because the room

opened up to a small amphitheater it was frequently reconfigured to accommodate various activities. Unlike the dedicated spaces for the FANZ programs at the East and South sites, other programs at the West site used this space for their activities during research sessions. Ginny was assigned to the West site.

With 144 students, the East site had the largest FANZ program followed by the South and West sites with 121 and 89 students, respectively (see Table 3). Attendance and demographic data were collected by the B&G Club staff members during the enrollment process. Unfortunately, many of the parents chose not to complete the intake forms or failed to complete some sections. Approximately 77% of the parent forms collected during enrollment were complete on average across all three sites. Parents at the West site completed 92% of the intake forms and parents at the South site completed 83%. Parents at the East site completed approximately 60% of the forms.

Based on the data parents reported, the students attending the B&G Club were primarily Hispanic/Latino and White, non-Hispanic/Latino. More of the students at the East and West sites identified with multiple racial and ethnic groups than those at the

Table 3

Demographic Summary of Students Enrolled in the B&G Club FANZ Program

Site	Enrolled students	# of complete forms	% female	% Asian	% Hispanic/ Latino	% Black or African American	% White	% two or more races	% Low SES
East	144	89	59	5	23	4	80	11	10
South	121	100	49	1	30	2	66	7	8
West	89	82	39	4	22	0	85	20	19

Note. Low socioeconomic status (SES) was defined as students eligible for free or reduced lunch at their local public school.

South site. Thus, these sites may serve a more diverse population than is reflected by the limited demographic information collected on the enrollment form. A larger percentage of the students at the East site were female, approximately 60%, compared to only 40% at the West site. There was no clear programmatic explanation for these differences. Finally, the percent of students from low income homes was highest at the West site with nearly 20% of the families completing the survey eligible for free or reduced lunch at their local school. The percent of families eligible for free and reduced lunch served by the West and South sites was approximately 10%.

Research Sessions

Research sessions were held in rooms assigned to the FANZ program at each site. Sessions were scheduled every day for 30 minutes. During these sessions, the teachers taught activities from Summer Brain Gain, a common curriculum designed by the National Boys & Girls Club. The Summer Brain Gain curriculum was created to provide youth with daily opportunities to practice reading, math, and science skills during interactive, teacher-led activities (see Appendix B). Research sessions started at 10:00 a.m. at the East site, 11:00 a.m. at the South site, and 1:30 p.m. at the West. Occasionally, the research sessions were canceled due to field trips, staff absences, or other circumstances.

Equipment Used During Research Sessions

Amazon Kindle Fire tablets were available during the research sessions at each site. The Kindle Fire features an 8.9-inch touch screen interface, Wi-Fi, weighs approximately one pound, and was fitted with a thick rubber protective covering to make

it easier to handle and distribute the devices to students. Each tablet was updated to Android 4.4.2 (KitKat) and contained the Amazon Silk web browser as well as a variety of apps (e.g., Minecraft, Temple Run, etc.). The tablets were used in some of the Brain Gain activities to look up information or view instructional videos. In addition, staff members occasionally offered free access to the tablets as a reward for active participation in the lesson or good behavior. The teachers were responsible to plug in and charge the tablets daily and the site supervisor verified that all tablets were secured at the end of the day.

Dependent Variables

The dependent measures for this study included specific and general praise rates. Teachers' perception of the social validity of the procedures was also measured. The following sections present the operational definition of specific praise, operational definition of general praise, recording procedures, coding procedures, and social validity. A summary of the dependent variables described in the following sections is included in Table 4.

Specific Praise

Specific praise was defined as an affirmative statement spoken by a teacher to a student or students in response to an academic or social behavior that includes a description of the praiseworthy behavior (Marchand-Martella et al., 2010). For example, teacher statements including "James, thanks for following instructions! I appreciate when you look at me, say ok, and go and do what you've been asked to do" or "I like when

Table 4

Operational Definitions and Data Sources for Dependent Variables

Variable	Definition	Source
Specific praise	An affirmative statement spoken by a teacher to a student in response to a student behavior that includes a description of the praiseworthy behavior	Audio recording
General praise	A global or broad statement that indicates a positive response to a student behavior	Audio recording

everyone gets their materials quickly and quietly" were coded as specific praise. These examples were coded as specific praise because they are positive, affirming statements that clearly identify the praiseworthy behavior. However, positive statements made to other staff members (e.g., "I really enjoyed your cooking lesson yesterday") would not be coded because it was not made in response to a student behavior. Statements indicating agreement with student comments unrelated to the Brain Gain lesson (e.g., "I agree, the weather is beautiful today") would not be coded because they do not approve or affirm a student behavior. Comments about the Brain Gain lesson would be coded as specific praise if the statement was affirmative, described the correct response, and was made in response to an academic behavior (e.g., "Yes, a fireman is someone who helps people in our town").

A statement was defined as a continuous vocalization made by the teacher to a student, without pausing or being interrupted, that contained at least one word. The first example of specific praise presented above is a single statement even though there are two distinct phrases. The first phrase is "James, thanks for following instructions!" and the second phrase is "I appreciate when you look at me, say ok, and go and do what

you've been asked to do." These two phrases would be coded as a single specific praise statement because the teacher spoke them as a continuous statement without pausing for more than two seconds or being interrupted by another speaker. If the teacher paused for two or more seconds between the first and second phrases, then the two phrases would have been coded as two separate specific praise statements.

An interruption was defined as any instance where the teacher paused speaking while another person spoke. Interruptions could be of any length. For example, if the teacher started to praise a student (e.g., "Sierra, I like your colorful picture and…") and was interrupted by the student's response (e.g., Sierra says "Thanks") before finishing his or her statement (e.g., "the details you've included on her hair are awesome!"), then the two statements would be coded as two separate specific praise statements.

General Praise

General praise was defined as a global or broad positive statement spoken by a teacher to a student or students in response to an academic or social behavior (Marchand-Martella et al., 2010). General praise statements could be directed toward an individual student or a group of students. For example, teacher statements including "Good job," "Nice work," "Excellent," or "Thanks everybody" were coded as general praise. As described in the previous section, the definition of a statement, a continuous vocalization without pausing or being interrupted, was used to determine the start and end of each general praise statement. For example, if a teacher made a broad positive statement (e.g., "Good!") and then paused for two or more seconds before saying another broad positive statement (e.g., "Yes, I like it!") then these two statements would be coded as two general

praise statements. If a teacher used a non-word to respond to a student behavior (e.g., "mhm!") the response was not coded because it failed to meet the definition of a statement. Finally, if a teacher said, "That's great!" and then paused for two seconds before finishing with, "That's great quieting down so we can move on" then the first statement would be coded as a general praise statement and the second statement would be coded as a specific praise statement. Both statements indicate a positive or affirmative response to a student social behavior, but only the latter specifies the praiseworthy behavior.

Recording Procedures

To quantify the number of specific and general praise statements, teachers made an audio recording of their research session each day. These recordings were uploaded to a cloud storage service and then coded for instances of general and specific praise by trained data collectors.

Recording equipment. Teachers recorded their vocal interactions with students during research sessions using an LG Realm mobile phone and a wired microphone. The LG Realm is equipped with a 4.5-inch touch screen interface, a 2100 mAH lithium-ion battery capable of 17 hours of continuous use, and Wi-Fi. A wired microphone was attached to the LG Realm using the audio jack. The microphone was a Neewer Hands-free omnidirectional lapel clip microphone with a wind muff.

The LG Realm operates using version 4.4.2 (KitKat) of the Android operating system. This version of Android provides increased security and a guest access mode. The phone was setup with restricted guest access to limit the functionality of the phone to

only the audio recording app. Thus, teachers could not send messages, browse the Internet, or make changes to the available applications. To access the recording app, teachers entered a five-node pattern lock. Audio recordings were made using the Skyro Pro audio recorder application. Skyro Pro provided teachers with a secure, reliable platform to record each research session. Audio files were recorded using a 192 kilobits per second bit rate and then stored as compressed mp3 files. In addition to recording and storing the audio files, Skyro Pro also automatically transmitted a backup to Box.com, a cloud storage solution authorized for data collection and storage by Utah State University. Data security was further enhanced by requiring a unique five item pattern to unlock the phone.

The LG Realm was connected to the Wi-Fi networks at the three sites. The Wi-Fi connections at all three schools were fast and reliable. However, connecting to the Wi-Fi network at the West site was problematic. The Wi-Fi network required authentication using a username and password every time the device was inactive for more than 30 minutes. Thus, Ginny, the teacher at the West site, had to manually enter the password and enable the Internet on all of the devices that were used each day to complete the intervention procedures. On average, she connected five devices to the Internet each day.

Recording procedures. To record their interactions using the LG Realm, teachers opened the Skyro Pro application and initiated the recording by tapping the "Start Recording" button. They were then free to lock the phone and store it in their pockets or on their lanyard for the duration of the session. At the end of the session, teachers unlocked the phone and stopped the recording. The Skyro Pro application immediately

transmitted the recording to Box.com, a secure cloud storage solution. On average, uploads to Box.com required an hour to complete. Through the transmission and storage of the audio recordings on Box.com, the experimenter was able to keep the research assistants blind to the identities of the participants and their progression through the research study. The experimenter instructed teachers not to record themselves as they conducted the self-management procedures at the end of the research sessions. When teachers inadvertently recorded these activities, the experimenter trimmed the audio track to delete these conversations from the audio recordings before assigning them to the research assistants for coding.

Coding Procedures

Research assistants coded the frequency of general and specific praise statements using the Audio Coding Form (see Appendix C). To complete the Audio Coding Form, research assistants first noted the participant number, date of the recording, date the track was coded, and the length of the track. The participant number and date of the recording, were included in the filename of the audio track. The length of the recording was displayed after accessing the audio track using the preview feature on Box.com. Research assistants used preview to pause, rewind, and fast forward the audio tracks. All research assistants used Google Chrome to access Box.com.

To code the frequency of general and specific praise, the research assistants used a three-step process. First, a research assistant listened to the track to identify praise statements. After identifying a praise statement, the research assistant rewound the audio recording to identify exactly when the statement began. The exact time the statement

started was noted from the audio track timeline and then recorded on the Audio Coding Form using a decimal to separate minutes and seconds (e.g., 10.36 for 10 minutes and 36 seconds). Next, the research assistant recorded the praise statement verbatim on the Audio Coding Form. Third, the research assistant reviewed each statement, independent of the audio track, and coded it as either specific or general praise.

The Audio Coding Form was an excel spreadsheet that was programmed to calculate the specific and general praise rates based on the length of the audio track and the frequency of each type of praise. The rate of specific and general praise statements was calculated by dividing the total frequency of each type of praise statement by the duration of the session in minutes.

Research assistants completed a 10-hour training prior to coding audio tracks from the teachers in this study. This training included didactic instruction, guided practice, and independent practice. The didactic and guided practice sessions lasted 4 hours. During the didactic training, the experimenter presented examples and non-examples of general and specific praise statements to the research assistants. Research assistants then classified the examples and non-examples as general praise, specific praise, or neither. To move onto the guided practice phase of the training, research assistants had to identify general, specific, and non-praise statements with 100% accuracy.

During the guided practice phase of the training, the experimenter and the research assistants listened to a sample audio track to identify general and specific praise statements. The sample audio track was recorded during the B&G Club after-school

program the previous school year and used with permission from the teacher and directors of the B&G Club. Each research assistant took a turn pausing the audio track when she heard a praise statement and then demonstrated to the group how to correctly record the timestamp, transcribe the praise statement, and enter the code for the identified statement (i.e., G = general praise, S = specific praise) into the Audio Coding Form. Any disparities between the experimenter and the research assistants were resolved by referring to the operational definitions and the coding procedures.

The independent training took approximately 6 hours to complete. To complete the independent practice session, each research assistant listened to a sample audio track and independently completed the Audio Coding Form. Eight different sample audio tracks were available. The sample audio tracks were an average of 7 minutes long (range = 5-9 minutes) and contained 4-18 praise statements per track. Research assistants coded their assigned track and then submitted the completed Audio Coding Form to the experimenter who reviewed the data and calculated the percent agreement between the experimenter and the research assistant. Instances of agreement were identified using the time-window method (MacLean, Tapp, & Johnson, 1985). An agreement was noted between two independent observers when they coded the same type of praise statement within plus or minus 3 seconds. Percent agreement was calculated by dividing the number of agreements by the number of opportunities to agree, or the total number of agreements plus disagreements (Cooper, Heron, & Heward, 2007). Training was continued until the research assistant achieved 80% agreement with the experimenter on at least three sample audio tracks. On average, research assistants completed seven of the

eight tracks independently (range = 3-8) before reaching the criterion.

Social Validity

Social validity is integral to applied research (Baer, Wolf, & Risley, 1968; 1987; Wolf, 1978). Wolf (1978) suggested, "if we aspire to social importance, then we must develop systems that allow our consumers to provide us feedback about how our applications relate to their values, to their reinforcers" (p. 213). Horner et al. (2005) provided further clarification of social validity by expanding its definition to include four essential features. These features include (1) the social importance of the dependent variable, (2) the social significance of the magnitude of change in the dependent variable, (3) the cost-effectiveness of the independent variable, and (4) the quality of implementation achieved by typical change agents in typical contexts.

The social validity of the procedures and results of this study were assessed using an adapted form of the *Intervention Rating Profile-15* (IRP-15; Martens et al., 1985). The IRP-15 is an abbreviated form of the Intervention Rating Profile (IRP) that contains statements concerning the general treatment acceptability of the target intervention, effects of the intervention on the target individual's behavior, and positive and negative impact on the implementers' behavior. The IRP-15 has been frequently used in school-based intervention research (State and Harrison, 2016) and researchers have shown evidence of convergent validity (r=-.86) between the IRP-15 and the evaluative sub-scale of the Semantic Differential Scale (Martens, Witt, Elliott, & Darveux, 1985). The significant negative correlation between these measures of the feasibility and effectiveness of interventions was expected since higher scores on the IRP-15 indicate

greater social validity of the procedures and lower scores on the evaluative sub-scale indicate greater social validity.

The IRP-15 measures teachers' perceptions of the social importance of the dependent variable (1) using a single indicator concerning the value of the procedures to students generally. The social significance of the magnitude of change in the dependent variable (2) is measured by three indicators that address the effectiveness of the procedures in changing the teacher's behavior and positive or negative effects on the participating students. The cost-effectiveness of the procedures (3) is measured by four indicators examining the teacher's perceptions of the feasibility of implementation in the classroom, another teacher's classroom, other educational settings, and with a variety of instructors. Finally, the IRP-15 addresses the quality of implementation (4) using a single indicator focused on how much teachers' liked the implementation process.

The standard form of the IRP-15 was modified for this study by removing questions focused on specific intervention outcomes for students. Researchers used a similar procedure to adapt the IRP-15 for use in a teacher intervention study similar to the current study (Briere et al., 2015). Six open-ended questions were also added to gather specific feedback about the components of the intervention (see Appendix F. These questions were an indirect measure of their perceived level of implementation. They also provided teachers an opportunity to share their recommendations for improving the procedures.

Evaluation of Student Feedback

Student responses during the study were collected as part of the self-management

intervention. To evaluate the validity of the reports collected from students, the correlation between these reports and teachers' use of specific praise was calculated during the intervention phases. If the percent of students reporting specific praise increased as teachers increased their specific praise rates, then it would indicate that the changes in teacher behavior were detected and reported by their students. Unfortunately, student perceptions of the importance or value of these changes in their teachers' use of specific praise was not measured.

To quantify the relationship between teacher behavior and student perceptions, a non-parametric rank-order correlation coefficient was calculated between teachers' specific praise rates and the percent their students reporting specific praise. A three-step procedure was used to calculate the rank-order correlation. First, the specific praise rates recorded for each teacher during all intervention phases of the study were independently ranked from highest to lowest. Second, the corresponding percent of students reporting specific praise were similarly ranked. Third, the correlation between these ranks was calculated using the following formula (Kachigan, 1986)

$$r_0 = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)}$$

where *n* is the number of pairs of ranks, D_i is the difference between the *i*th pair of ranks, and $\sum_{i=1}^{n} D_i^2$ is the sum of the squared differences between paired ranks.

Independent Variable

The independent variable included (a) the teacher self-evaluation, (b) the student feedback, (c) visual performance feedback, and (d) teacher and student trainings. Two

additional intervention procedures are described: prompting and rewards. These procedures were used to increase the potency of the matching procedure for teachers who did not respond to the basic self-management procedure.

Teacher Self-Evaluation

Teachers rated themselves using a simple, web-based teacher rating form (see Appendix D). At the top of the form was written, "Please indicate your rating for today." Directly below this statement was a series of five radio buttons vertically positioned. Adjacent to each button was presented one of five ratings: "1 = Poor—I specifically praised 0-20% of my students," "2 = Ok—I specifically praised 20-40% of my students," "3 = Good—I specifically praised 40-60% of my students," "4 = Very Good—I specifically praised 60-80% of my students," and "5 = Outstanding—I specifically praised 80-100% of my students." Teachers selected the rating that they believed best reflected their use of specific praise during the preceding research session.

Student Feedback

Student feedback was collected using a web-based Student Feedback Form (see Appendix D). The Student Feedback Form contained two questions about students' experiences during the research session. The first question posed to each student completing the form was, "Did [teacher name] praise your behavior today?" Students responses to this question were indicated by tapping a radio button adjacent to one of the following options "No," "Yes, one time," and "Yes, more than one time." The second question read, "Did [teacher name] tell you which behavior she/he was praising?" Two response options, "Yes" or "No," were presented below this question.

The data from these questions were used to calculate the percent of students reporting specific praise. Students who answered yes to both of the questions on the form were identified as students reporting specific praise. The percentage of students who reported specific praise was calculated by dividing the number of students reporting specific praise by the total number of respondents to the survey. If a student answered yes to only one of the questions or no to both, then he or she was not identified as a student reporting specific praise. To make the units of student feedback comparable to the teacher self-evaluation rating, the percent of students reporting specific praise was converted into a 5-point scale. If the percentage of students was between 0-20% then the student rating was a one, 20-40% a two, 40-60% a three, 60-80% a four, and 80-100% a five. These percentages directly align with those presented to teachers on the Teacher Rating Form.

Visual Performance Feedback

Visual performance feedback was used to present the data collected via the Teacher Rating Form and Student Feedback Form (see Appendix D). The layout of the figures and tables emphasized the match, or lack of a match, between teacher and student ratings. This emphasis was consistent with the design and intent of the feedback used in previous self-management studies (Young et al., 1991). The most prominent element of the visual performance feedback was a graph presenting student and teacher ratings from each of the last 5 school days. The line representing the teacher's self-evaluation rating showed the exact rating selected by the teacher using the Teacher Rating Form on each of the last 5 school days. The line representing the students' ratings showed student rating based on the percent of students reporting specific praise for each of the last 5 days.

To provide feedback on the degree of correspondence between the student and teacher ratings a table was displayed below each date on the line graph with the words "Match," "Next-door," or "No Match." The match score was determined based on a comparison between the teacher and student ratings. If teachers rated their performance as "5 = Outstanding…," meaning that they specifically praised 80-100% of students in the session, and the percent of students reporting specific praise was between 80-100%, resulting in a student rating of five, then the label "Match" was displayed. A next-door match was awarded if the discrepancy between teacher and student ratings was no more than one category (e.g., a teacher rating of "5 = Outstanding…" and a student rating of 4 based on 60-80% of students reporting specific praise).

The space behind the matching score was colored to aid interpretation and encourage "Match" scores. The background behind the "Match" label was colored green. The letters in the label "Next-door" were displayed in black over a yellow background. Finally, the background behind "No Match" was colored red. These specific colors were used to aid interpretation by drawing upon teachers' common experience with traffic lights. That is, the colors were associated with meanings similar to traffic lights: green means keep doing what you are doing; yellow means caution or change what you are doing; and red means stop. Analogous changes in practice were suggested during the initial training.

Additional Intervention Procedures

Matching + prompts. Prompting was added to increase teachers' specific praise rate during research sessions. Teachers transitioned into the Matching + Prompts phase if

their data during the Matching phase revealed a counter therapeutic trend or if there was greater than 50% overlap between the data in the Matching and Baseline phases. Silvestri (2004) suggested that prompting was an appropriate strategy to increase the potency of a self-monitoring intervention because it helped participants remember to use praise when they were busy teaching. An auditory and tactile prompt was issued every 5 minutes via the MotivAider app on the LG Realm. The auditory prompt was a single chime presented at half volume. The audio component was presented at half volume to reduce the likelihood that it could be heard by students. The tactile prompt consisted of three bursts of intense vibration. A slight buzzing sound accompanied the tactile prompt, but this sound was barely audible over the corresponding chime.

Teachers were taught to use specific praise every time they felt or heard the prompt. Specifically, they were instructed to scan the classroom to identify one or more students who was currently meeting their academic or behavioral expectations for the current activity. Once they identified a student, they were instructed to use specific praise to recognize the student's appropriate behavior. After the research session, students completed the Student Feedback Form, teachers completed the Teacher Rating Form, and then teachers viewed the visual performance feedback to determine if their ratings matched with their student feedback.

Matching + prompts + rewards. A reward contingency was added to the matching and prompting procedures as needed to provide a positive consequence for perfect matches combined with student ratings of 4 or 5. Together, the experimenter, site supervisor, and teacher developed a reward contingency. The group met in the morning

on the first day of the matching + prompts + rewards phase to review the teacher's daily feedback and to select a reasonable performance goal for the next 5 days. There were no specific protocols used to select this goal, but the decision had to be unanimous. Next, the teacher selected a reward that would encourage him/her to meet the performance goal. The reward could be anything that cost less than $30 total, was agreeable to the site supervisor, and could be started or delivered on the next day after the goal was met. The contingency included a performance goal specifying a target student rating and a level of matching that must be maintained to gain access to the selected reward. For example, if a teacher received an average student rating of 4 and achieved a perfect match or next-door match on four of the five days then lunch would be provided for the staff members at the site.

Training

Teacher training. The teacher training included four components including (a) didactic training on the use of praise, (b) activity planning to identify appropriate behaviors to praise, (c) a demonstration of how to complete the Student Feedback Form and the Teacher Feedback Form, and (d) practice interpreting visual performance feedback.

Didactic training. The didactic training was used to help teachers understand the benefits of high specific praise rates in the classroom and how the procedures could help them increase their use of praise. The experimenter defined general and specific praise and provided opportunities for teachers to practice saying specific praise statements. The experimenter also instructed teachers on the importance of delivering praise so it is

recognized and acknowledged by students. The experimenter used a PowerPoint displayed on a laptop computer to display summaries of key ideas from the training and present examples or non-examples of specific praise.

To assess understanding of the concepts from the didactic training, teachers completed a brief worksheet to test their ability to correctly identify different types of praise. The worksheet contained a mix of 25 general, specific, and nonpraise statements. Teachers reviewed each statement and identified whether the statement was a general praise statement, a specific praise statement, or a nonpraise statement. The experimenter required that teachers correctly identify 80% or more of the statements before moving on to other training activities. If the teacher scored below 80% on the worksheet, the experimenter reviewed the examples and nonexamples again, discussed novel examples, and then asked the teacher to complete the worksheet again. This pattern was repeated until the criterion of 80% accuracy was achieved.

Activity planning. Teachers prepared an action plan to guide their use of praise during the next research session. They were asked to state their expectations for a variety of activities and then create contextually appropriate specific praise statements for academic and social behavior that met their expectations. A completed action plan included a description of the activities planned for the next research session, a summary of the teacher's academic and behavioral expectations for each activity, and three or more specific praise statements they could use to recognize appropriate student behavior. The action plan was considered complete once it contained at least six correct specific praise statements for two distinct instructional activities.

Complete student feedback and teacher self-evaluation forms. Teachers were shown how to complete the Teacher Rating Form and Student Feedback Form using the LG Realm and Kindle Fire, respectively. The teachers verbally described each step and then demonstrated how to complete the process using the tablet or phone.

Interpret visual performance feedback. The experimenter showed teachers multiple examples of visual performance feedback, provided strategies to interpret the data, and discussed how to use their data to modify their use of praise. To interpret the data, the experimenter demonstrated how to identify the student rating, identify the teacher rating, and then locate the match score. In addition, the experimenter discussed with teachers how the student and teacher data related to one another and pointed out longitudinal trends. The experimenter instructed teachers to select one of three strategies in response to the data. If a green "Match" was presented on the visual performance feedback displayed on the phone, then the teachers should keep doing what they had been doing. If a yellow "Next-door" match was presented on the visual performance feedback, then the teachers should make an incremental change in their use of specific praise or their self-evaluations. Finally, if a red "No Match" was presented on the visual performance feedback, then the teachers should make a substantial increase in their use of specific praise or make their praise more noticeable to their students by using the student's name when delivering a praise statement or increasing the number of students who received multiple specific praise statements. Teachers demonstrated how to respond to visual performance feedback by correctly identifying each component of the graph and table. They were also required to select an appropriate strategy based on the hypothetical

visual performance feedback.

Student training. Students received training on the purpose of the research project, how to discriminate between general and specific praise, and how to complete the Student Feedback Form. First, the students' teacher explained the purpose of the study and identified the potential benefits to the students in their classroom. Second, the experimenter defined general and specific praise and then gave examples of each type of praise. The experimenter asked a student to model a specific behavior (e.g., sitting quietly, doing jumping jacks, raising his or her hand, etc.) while the teacher read scripted general praise, specific praise, or non-praise statements in reference to the student's behavior (e.g., "Thank you for sitting quietly," "That's great," "The color is blue," etc.). After the teacher read each statement, the experimenter asked a student to identify the statement as praise or not praise. If the student correctly identified the statement as praise, the experimenter asked another student to indicate whether the statement was general or specific praise. This process was continued until at least five different praise statements were correctly identified by the first student selected by the experimenter.

To teach the students how to complete the Student Feedback Form, each student was given an opportunity to select answer and submit their response on the electronic form using the Kindle Fire. First, the experimenter and the teacher distributed tablets to two students, the experimenter then told the students to read each question and to remember all the activities they participated in during the session. Second, the experimenter showed the students how to select answers on the electronic form. The experimenter explained that the students should select "Yes" on the first question if their

teacher praised their behavior during the session. If the students could not remember being praised, they should select "No." Next, the students were told to select "Yes" on the second question if their teacher told them which behavior was praiseworthy or to select "No" on the second question if they were not sure whether the teacher described their behavior. Finally, the experimenter showed the students how to submit their responses. The students were then told to pass their tablet to another student after they submitted the Student Feedback Form. The experimenter and teacher repeated this process with the next pair of students until all the students in the classroom completed an electronic form. During subsequent sessions, the teacher explained the purpose of the form to the students and distributed approximately one Kindle Fire for every five students. Teachers monitored students to make sure they were using the tablet to complete the survey and to ensure all students completed a form.

Research Design

A multiple-baseline design across teachers was used to evaluate the effect of the independent variables on teachers' specific and general praise rates (Cooper et al., 2007). Experimental control within a multiple-baseline design across teachers is achieved by systematically introducing the independent variable one participant at a time. The effects of the introduction of the independent variable can then be compared with the participant's baseline data and against the data collected from the other participants. In this study, experimental control was maintained across participants during the implementation of the Matching phase, but not the implementation of the additional

intervention phases, Matching + Prompts and Matching + Prompts + Rewards phases. Thus, between subject comparisons of the effects of the additional intervention phases may not be as useful as within subject comparisons.

The following sections describe the various phases of the research study. The Baseline phase preceded all other phases. All teachers participated in two phases: Baseline and Matching. Brendan, Lilli, and Ginny also participated in the Matching + Prompts phase. Brendan was the only teacher to participate in the Matching + Prompts + Rewards phase, the fourth and final phase of the study. The frequency of completing teacher self-evaluations and collecting student feedback was held constant throughout all intervention phases. The critical features of each experimental phase are presented in Table 5.

Table 5

Research Procedures by Phase

Baseline	Matching	Matching + Prompts	Matching + Prompts + Rewards
Daily • Record target session	Daily • Record research session • Complete Teacher Feedback Form • Collect Student Feedback Forms • View performance feedback	Daily • Record research session • Complete Teacher Feedback Form • Collect Student Feedback Form • View performance feedback • Prompts every 5 min	Daily • Record research session • Complete Teacher Feedback Form • Collect Student Feedback Form • View performance feedback • Prompts every 5 min • Review progress with supervisor
	One-time • Attend teacher training • Conduct student training	One-time • Attend prompts training	One-time • Setup contingency • Receive rewards

Baseline

During the baseline phase, teachers and students participated in all regularly scheduled instructional and social activities planned at the B&G Club. Participating teachers audio recorded their research sessions each day. Audio recordings were collected using the LG Realm and the Skyro Pro app. During the first three days of Baseline, the experimenter met with the teachers to resolve any problems they encountered using the audio recorder and thanked them for participating in the study. Thereafter, the experimenter provided technical support via text messaging or onsite consultation, as needed.

Teachers advanced from baseline to the matching phase only after a stable pattern was evident in their specific praise rates. A stable pattern of responding was defined as a clear pattern in the level, trend, and variability of the data. Although all three variables were considered carefully to determine stability, the trend in the Baseline data was used to determine when to start the Matching phase to ensure that no trend or a counter-therapeutic trend was evident during Baseline. Stability had to be evident for no less than five data points.

Matching

On the first day of the matching phase, teachers completed the training in the morning prior to the research session. The experimenter trained Brendan, Lilli, and Ginny; whereas, the B&G Club Program Coordinator trained Kate. Trainings were held onsite in an otherwise unoccupied classroom for approximately one hour. The site coordinators sat in on the training to complete the implementation fidelity checklist for

Brendan, Ginny, and Lilli. The program director trained Kate and a clerical staff member completed the checklist.

After completing the didactic training, activity planning, and demonstrating an understanding of how to interpret the visual performance feedback, the experimenter and the teacher discussed their roles in delivering the student training. The experimenter described each activity that needed to be completed and addressed any concerns about the logistics of conducting the training. They also enabled access to the Student Feedback Form on the Kindle Fire tablets. To enable access, teachers opened the Silk web browser, navigated to the bookmarks screen, and tapped a link to the Student Feedback Form labeled with their participant number (e.g., #31, #22, etc.). The teacher placed the tablets in the materials for the research session and returned to her regularly assigned responsibilities until the last 15 minutes of the research session when the student training was held.

At the end of the research session, the experimenter and the teacher conducted the student training, which lasted approximately 20 minutes. After all students completed the Student Feedback Form the experimenter helped students transition to the next activity and the teacher completed the Teacher Rating Form on the LG Realm. To complete the Teacher Rating Form, teachers tapped a link on the phone's homepage to open the form in the Chrome web browser. Teachers completed the form by tapping the radio button next to the rating that best described their use of praise during the research session and then tapped the submit button. The teacher downloaded the visual performance feedback by tapping a link on the "Thank You" screen displayed after submitting the Teacher

Rating Form. The visual performance feedback was stored as a Portable Document Format (i.e., .pdf) on the LG Realm. It was displayed immediately after the download was complete and archived on the phone.

The experimenter provided technical assistance where necessary and then discussed the visual performance feedback with the teacher. This consultation ensured each teacher understood the data collection process, could interpret the data, and felt comfortable making decisions based on the data prior to completing these activities without support. During subsequent research sessions, the experimenter provided technical assistance upon request from the program coordinator or via text message. The teacher collected the Student Feedback Form, completed the Teacher Rating Form, and reviewed the visual performance feedback every day during the Matching phase.

Matching + Prompts

Fifteen minutes prior to the research session on the first day of the matching + prompts phase, the experimenter enabled the Motivaider app on the LG Realm assigned to the teacher transitioning into the new phase. The experimenter adjusted the settings to provide a prompt to praise every 5 minutes using the speaker and vibration motor on the phone. In addition, the experimenter taught the teacher how to adjust the settings on the app and told each teacher to use specific praise following each audio signal and vibration. Teachers continued to collect student ratings, complete self-evaluations, and review their visual performance feedback every day.

Matching + Prompts + Rewards

In addition to the matching and prompts intervention components, the supervisor, experimenter, and participating teacher identified a goal and a reward that the teacher thought would help increase the teacher's use of specific praise during the research sessions. The teacher and the site supervisor then met daily to review the visual performance feedback, discuss progress toward the goal, and develop strategies to improve performance. The site supervisor reported progress to the experimenter after each meeting. If the site supervisor informed the experimenter that the teacher met the identified goal for five consecutive sessions, then the experimenter made arrangements to deliver the reward on the next day. For example, if a teacher selected lunch for himself and the staff members at the site as the reward, then the experimenter brought lunch to the staff on the next day after the contingency was met.

Interobserver Agreement and Implementation Fidelity

Interobserver Agreement

Interobserver agreement (IOA) was collected on 82% of the audio tracks collected for this study. IOA was collected in every phase of the study and on both dependent variables. Two research assistants were assigned to independently code selected audio tracks. Agreement between coders was assessed using the same time window procedure employed during training. Specifically, an agreement was scored when both research assistants identified the same type of praise statement, general or specific, within a time window of plus or minus three seconds. Percent agreement was calculated by dividing the

number of agreements by the total number of agreements plus disagreements and multiplying by 100 (Cooper et al., 2007).

IOA across all participants and all phases of the study averaged 85%. The IOA scores for Brendan and Ginny were 86% (range = 75-100%) and 87% (range = 75-100%), respectively. Kate and Lilli had slightly lower IOA percentages with 84% (range = 72-100%) and 84% (range = 71-100%), respectively. The percent agreement was slightly higher for specific praise (range = 83-90%) than it was for general praise (range = 86-92%). A detailed breakdown of the percent agreement data by phase is presented in Table 6.

Table 6

Summary of the IOA Results for Each Participant During Each Phase for Each Type of Praise

	Baseline		Matching		Matching + prompts		Matching + prompts + rewards	
Participant and type of praise	*M* (%)	Range (%)	*M* (%)	Range (%)	*M* (%)	Range (%)	*M* (%)	Range (%)
Brendan								
Specific praise	82	50-100	84	54-100	96	89-100	96	96-96
General praise	82	50-100	73	33-89	94	89-100	89	88-90
Ginny								
Specific praise	92	50-100	90	71-100				
General praise	90	50-100	91	86-95				
Lilli								
Specific praise	88	57-100	93	81-100	87	78-97		
General praise	96	79-100	82	54-98	88	79-95		
Kate								
Specific praise	86	64-100	84	64-100	91	91		
General praise	82	68-100	80	68-100	100	100		

After computing IOA for the initial coding, research assistants reviewed each instance of disagreement. First, each research assistant opened their original coding form and reviewed how they coded the disputed statement. Next, each research assistant reviewed the operational definitions of each type of praise and then listened to the audio track at the identified time to determine if they should revise their codes. Research assistants made any revisions they deemed necessary to their Audio Coding Sheet and then submitted the revised data to the experimenter. Any disagreements that were not resolved by this process were reviewed by the experimenter and the two research assistants to decide how to resolve the disagreement.

Implementation Fidelity

A summary of the evaluation of implementation fidelity is presented in Table 7. The quality of the implementation of the critical features of the self-management procedure were documented using information from checklists and electronic records.

Table 7

Implementation Fidelity Measures and Collection Procedures

Critical feature	Source	Product
Teacher training	Direct observation	The percent of completed steps on the teacher training checklist
Student training	Direct observation	The percent of completed steps on the student training checklist
Teacher self-evaluations	Digital records	The percent of sessions during intervention with a completed teacher rating form
Student feedback	Digital records	The percent of sessions during intervention with completed student feedback forms
Performance feedback	Digital records	The percent of sessions on which the file displaying the performance feedback was downloaded

The checklists were completed by B&G Club staff who attended the student or teacher trainings and were asked to observe and document critical features of the training. Electronic records were used to document the collection of information from student and teachers.

Implementation fidelity for the participant training was measured using the Teacher Training Checklist (see Appendix E). The checklist was used to calculate the percent of completed steps by dividing the number of steps marked competed by the total steps listed on the checklist. The site supervisors at each location completed the checklist while the experimenter conducted the training for Brendan, Ginny, and Lilli. Based on the site supervisors' records, the percent of completed steps for these participants was 100%. For Kate, the program coordinator conducted the training and the Teacher Training Checklist was completed by a staff member. The fidelity of implementation for Kate's training was 89%. The trainer failed to have Kate complete the action plan and generate six unique specific praise statements. In addition, Kate's initial score on the worksheet, used to verify each teacher's ability to discriminate between general and specific praise statements, was 52% correct and the trainer failed to provide remediation to improve her score until the next day. By comparison, Brendan, Ginny, and Lilli scored 88%, 84%, and 96% correct, respectively.

The Student Training Checklist was used to assess the implementation fidelity of the student training (see Appendix E). A staff member at the B&G Club was asked to complete the checklist during the student training conducted by the teacher and his or her trainer. The percent of completed steps was calculated by dividing the number of steps

marked competed by the staff member observing the training by the total steps listed on the checklist. The implementation fidelity of the student training was 100% for all teachers.

Fidelity of implementation for the collection of the Student Feedback Form, the completion of the Teacher Rating Form, and the viewing of the visual performance feedback was evaluated using electronic records. On 100% of the research sessions the Teacher Rating Form and Student Feedback Form were accurately completed and logged on Google Forms. The best available evidence to document that teachers viewed the visual performance feedback was the archived copies of the pdf files stored on the LG Realm. Unfortunately, these data were not actively backed up to prevent data loss. This lack of redundancy increases the likelihood that individual files may have been deleted due to user error, malfunctioning hardware, or other factors unrelated to viewing these data. Across all teachers and phases of the study, 77% of the pdf files were archived on the LG Realms. The device used by Brendan contained 87% of the pdf files for each day he collected data compared to 92% for Lilli and 88% for Kate. Ginny's device contained the fewest overall records with only 42% of the pdf files archived.

CHAPTER IV

RESULTS

The primary research question concerned the effects of the self-management procedure on teachers' use of specific praise. Each Teacher's rate of specific praise during the Baseline and intervention phases is presented in Figure 1. On average, research sessions were held on 80% of the days the B&G Club was open. Brendan's research session was canceled by his supervisors 11 times during the course of the study compared to only five times for Ginny, five times for Lilli, and six times for Kate.

Primary Research Question: Specific Praise

Baseline

All teachers emitted low, generally stable specific praise rates during the Baseline phase. Brendan's specific praise rate averaged 0.25 statements per minute (range = 0.05 to 0.56). Ginny's baseline data were less variable. She averaged 0.12 specific praise statements (range = 0.0 to 0.38). Higher variability was present in the data from Kate and Lilli. Their baseline data could be characterized as generally stable and punctuated by multiple rate spikes. These spikes occurred more frequently in the early stages of the baseline phase for Lilli. Lilli's specific praise rate averaged 0.40 statements per minute (range = 0.16 to 0.88) whereas Kate's specific praise rate averaged 0.25 statements per minute (range = 0 to 0.66). It should be noted here that Kate was hired several weeks after the start of the study and was recruited for the study on her second day on the job.

There were no clear trends in the specific praise data collected during the Baseline

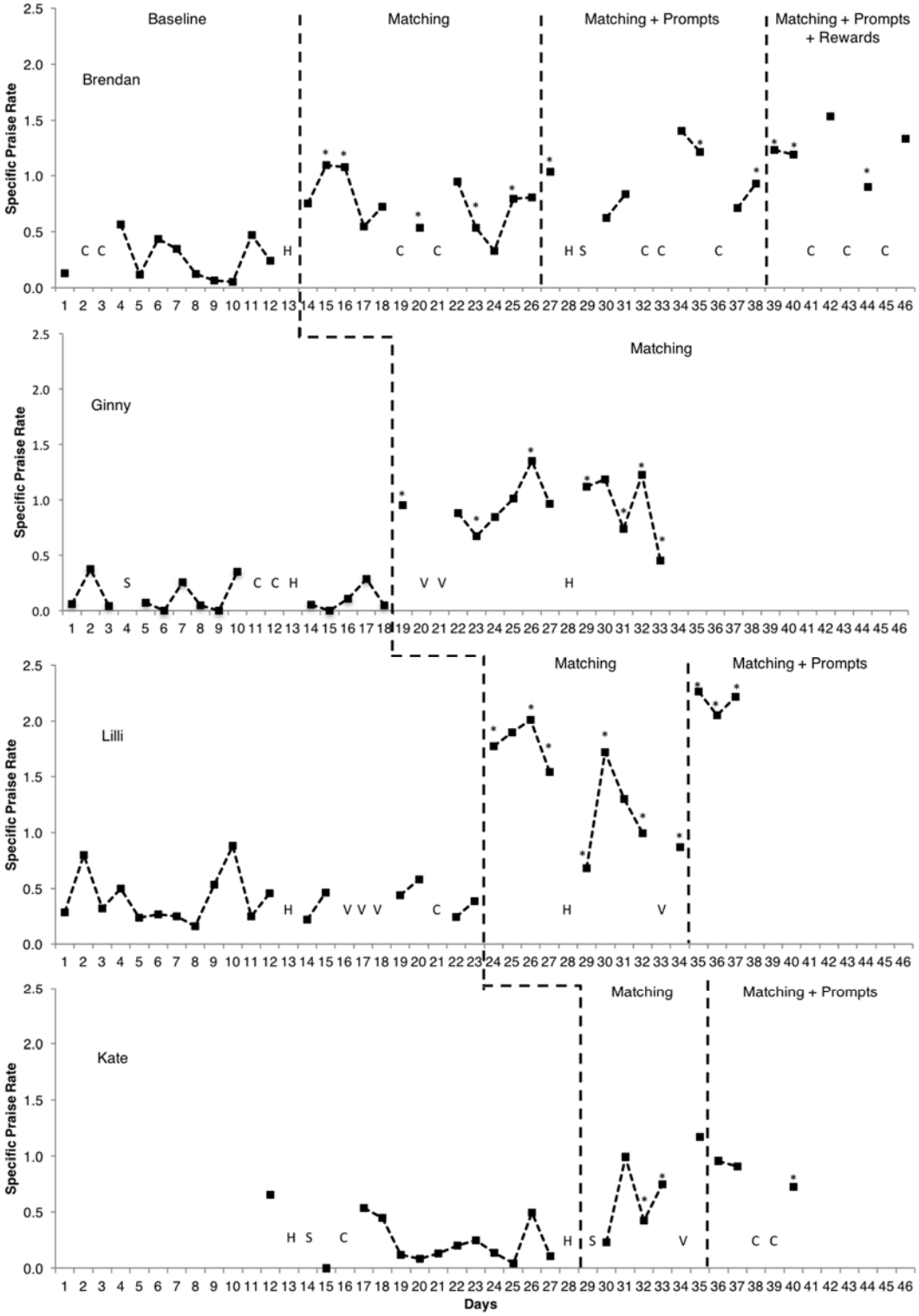


Figure 1. Specific praise rate for all participants across all phases. H = Holiday break; S = Sick day; V = Vacation day, C = Change in the schedule by administrators. * denote days when teacher's self-evaluations perfectly matched student ratings.

phase. To quantify these trends, a line was fit to the data using the least-squares method (Parsonson & Baer, 1978) and revealed a small negative slope for Brendan, Ginny, and Lilli. This analysis for Kate's data indicated a negative slope of 0.2. The consistently negative slopes may suggest initial reactivity to the audio recordings. The data from Lilli and Kate are both characterized by higher specific praise rates and greater variability during the first half of the Baseline phase. More consistent and lower specific praise rates were observed for both participants during the second half of the Baseline phase.

Matching

After three data points during the Matching phase, a level change was evident in Brendan's specific praise data. Brendan's mean specific praise rate during the Matching phase was 0.74 statements per minute (range = 0.33 to 1.10). Unfortunately, a counter therapeutic trend was also evident in his data. The regression line had a slope coefficient of -0.021. By comparison, the regression line for Brendan's baseline data had a slope coefficient of -.0041. The clear level change in Brendan's data should be considered in light of the overlap between his data during the Baseline phase and all his data from the Matching phase. Approximately 36% of the data points during the Matching phase overlapped with his data during Baseline.

The changes evident in Brendan's use of specific praise were not mirrored in the data from his peers during this same period. No level or trend changes were evident in the baseline data for Ginny during days 14 through 18, the first 5 days of the Matching phase for Brendan. Although only two data points were collected during this period for Lilli, her pattern of responding appeared very similar to her data during the previous 12 data

points. Kate's data were highly variable during this period but similar variability during the first days of active recording was observed in Brendan's and Lilli's data. By contrast, Kate's specific praise rate was zero on day 15 when Brendan's specific praise rate was at its highest point.

An immediate effect of the intervention was evident in Ginny's data. Ginny's use of specific praise during the Matching phase increased to a mean of 0.95 (range = 0.45 to 1.35) statements per minute. For comparison, her mean rate during the Baseline phase was 0.12. There was no clear trend in Ginny's data during the Matching phase. The regression line for Ginny's baseline data had a near zero slope (m = -0.0001). The regression line fit to her data from the Matching phase was similarly flat (m = -0.0016). Finally, there was no overlap between her specific praise rates during Baseline and those observed during the Matching phase.

Changes were not observed in Kate's and Lilli's specific praise rates immediately after Ginny started the Matching phase. Ginny's praise rate increased immediately during the first 5 days of the Matching phase, days 19 through 23. Four data points were collected during this period for Lilli. Her mean specific praise rate across these four data points was nearly identical to her mean throughout baseline. Kate collected data every day during the interval of day 19 through 23. Her data were consistently below her baseline average throughout this interval.

Lilli's mean specific praise rate increased to 1.42 (range = 0.68 to 2.01) specific during the Matching phase. Her three highest praise rates occurred during the first three data points of the phase and then decreased steadily thereafter. A line was fit to her data

to quantify changes in trend between the Matching and Baseline phases. The line fit to her data during the Matching phase had a slope coefficient of -0.105 compared to a slope coefficient of -0.001 during Baseline. Thus, Lilli's data exhibited a level and trend change from the previous phase. Unfortunately, the trend during the Matching phase was contrary to the intended, therapeutic direction. Only 15% of Lilli's data during the Matching phase overlapped with her baseline data.

Finally, Kate's use of specific praise increased from a mean of 0.25 specific praise statements per minute during baseline to 0.71 statements (range = 0.23 to 1.17) during the Matching phase. Unlike the decreasing trends noted for Brendan and Kate during the Matching phase, Kate's data are best characterized by an increasing trend. The trend line fit to her data had a slope coefficient of -0.02 during baseline compared to 0.141 during the Matching phase. It is difficult to pinpoint exactly when this trend started because Kate's data on the second day after the start of the Matching phase for Lilli spiked upward above her baseline average. The other two data points during the period following the start of the intervention for Lilli were near 0 and in line with the previous six data points.

In addition, there was substantial overlap (40%) between Kate's data in the Matching and Baseline phases. However, the substantial overlap between her data in during the Baseline and Matching phases should be considered in light of the positive trend in her data. Greater overlap is less of a concern with a clear reverse in the trend. The lack of data points during the Matching phase makes it difficult to determine whether this positive trend was the result of an experimental effect or an artifact of high

variability across a limited number of data points. The slight level change, potential change in trend, increased variability, and less than 60% nonoverlapping data points between the Matching and Baseline phases make it difficult to conclude that there was an effect of the intervention on Kate's use of specific praise.

Matching + Prompts

Three participants were included in the Matching + Prompts phase. Participants were selected for this phase if no clear level, trend, or variability change was evident after no less than five data points, or if any of these changes were in a nontherapeutic direction. Clear level changes required near zero overlap between data in adjacent phases.

Brendan's mean specific praise rate increased to 0.97 statements per minute (range = 0.63 to 1.40) from a mean of 0.25 statements per minute during Baseline and 0.74 statements per minute during the Matching phase. The line of best fit indicated a slight positive trend (m = .006) in his data during this phase. However, considerable overlap (66%) was observed between his data during the Matching + Prompts phase and the adjacent Matching phase.

Lilli's mean specific praise rate during the Matching + Prompts phase was 2.18 statements per minute (range = 2.05 to 2.26) relative to mean rates of 0.47 and 1.42 during the Baseline and Matching phases, respectively. Because only three data points were collected during this phase, a line of best fit was not calculated. There was no overlap between Lilli's specific praise rates during the Matching + Prompts phase and her rates during the Matching phase. In addition, there was very little variability in Lilli's performance during this final phase, but the lack of additional data makes it difficult to

ascertain whether this was a long-term effect of the addition of prompts or the short-term result of additional training.

Kate's rate increased to 0.86 specific praise statements per minute (range = 0.73 - 0.96) relative to mean rates of 0.41 and 0.71 during the Baseline and Matching phases, respectively. Given her highly volatile praise rates throughout the study, it is interesting to note that Kate's praise rates were very consistent during this phase. A line was fit to Kate's data during the Matching + Prompts phase to quantify any trend in her data. The resulting line had a slope coefficient of -0.22. Considerable overlap (i.e., 66%) was noted between her data during the Matching and Matching + Prompts phases.

Matching + Prompts + Rewards

Brendan was the only participant who experienced the final phase: Matching + Prompts + Rewards. Teachers moved into this phase if no clear change in level, trend, or variability was evident after no less than five data points during the Matching + Prompts phase, or if any of these changes were in a nontherapeutic direction. Kate's data during the Matching + Prompts phase justified her movement into this final phase, but a scheduling conflict precluded further participation in the study.

Brendan's mean specific praise rate did not substantially improve during the Matching + Prompts + Rewards phase (M = 1.24 specific praise statements per minute), but his rate of specific praise during this phase (range = 0.90 to 1.53) was less variable than rates observed in other phases (Matching range = 0.33 to 1.10; Matching + Prompts range = 0.63 to 1.40). Overlap with specific praise rates observed during the Matching + Prompts phase was 66%.

Secondary Research Questions: General Praise

The first secondary research question focuses on the effects of the intervention on teacher's general praise rate. Figure 2 displays the general praise rates for each teacher during each phase of the study. Marginal increases in Brendan's mean general praise rate were evident across all four phases with the mean rate increasing from 0.26 (*SD* = 0.14) to 0.43 (*SD* = 0.28) to 0.55 (*SD* = 0.25) and to 0.69 (*SD* = 0.28) general praise statements per minute during the Baseline, Matching, Matching + Prompts, and Matching + Prompts + Rewards phases, respectively. During Baseline, Brendan's highest general praise rate was 0.43 statements and his lowest rate was 0.04 statements per minute. The range in all other phases was nearly identical with a minimum value around 0.20 and a high around 1.00 statements per minute. Overlap with baseline rates of general praise was observed in every phase for Brendan.

Ginny's average general praise rate was 0.12 during Baseline (range = 0.13 to 0.39; *SD* = 0.12) compared to an average of 0.66 during the Matching phase (range = 0.17 to 1.41; *SD* = 0.39). Her data during the Matching phase are characterized by greater parity between general and specific praise rates that gradually decreases throughout the phase with general praise rates decreasing steadily toward the baseline mean. The percentage of nonoverlapping data between Ginny's general praise rates in Baseline and those in the Matching phase was 75%.

Lilli's general praise rate was highly variable throughout the study. During Baseline, her average general praise rate was 0.78 statements (range = 0.13 to 1.49; *SD* = 0.35). Her average general praise rate increased to 1.17 statements (range = 0.41 to 2.37;

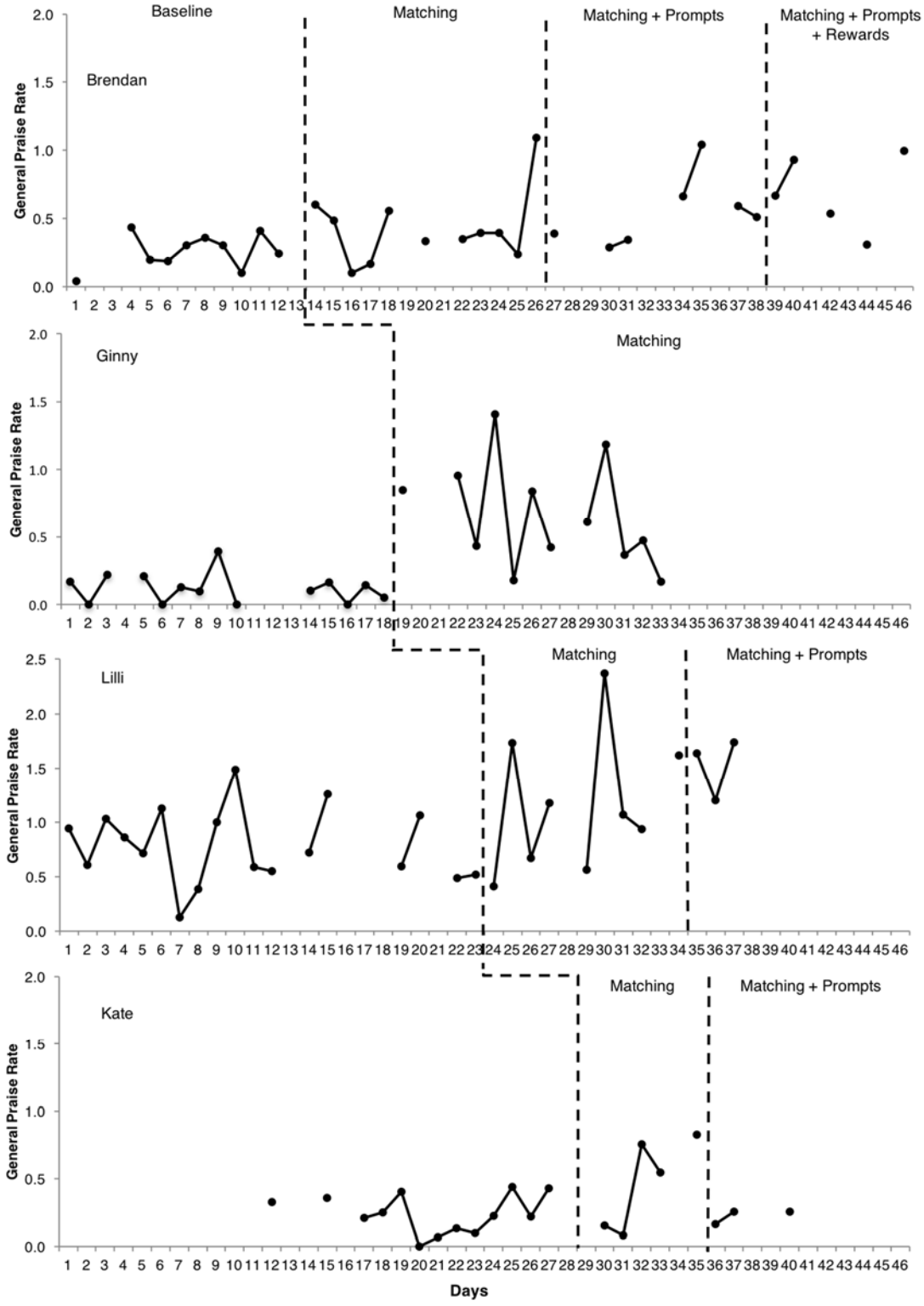


Figure 2. General praise rates for all participants. Rates are displayed using a solid line with a circular marker.

SD = 0.63) during the Matching phase and then to 1.52 statements (range = 1.20 to 1.74; *SD* = 0.28) during the Matching + Prompts phase. Lilli's highly variable general praise rates created a high degree of overlap between phases. The percentage of non-overlapping data between her general praise rates during Baseline and the Matching phase was 33%.

Finally, Kate's use of general praise increased during the Matching phase (*m* = 0.47; range = 0.08 to 0.83; *SD* = 0.13) relative to baseline rates (*m* = 0.24; range = 0 to 0.44; *SD* = 0.33) and then rapidly decreased during the Matching + Prompts phase (*m* = 0.22; range = 0.17 to 0.25; *SD* = 0.06). The percentage of nonoverlapping data between Kate's general praise rates during the Baseline and Matching phases was 60%.

Distribution of Specific and General Praise

An additional analysis was run to examine the effects of the intervention on the teachers' relative use of specific and general praise. Figure 3 displays the percent of specific praise (frequency of specific praise divided by the sum of the general and specific praise) plotted with the corresponding specific praise rate for each teacher. The specific praise rates and primary scale on the left are identical to those in Figure 1.

During Baseline, the percent specific praise for most participants was highly variable. For example, Ginny's and Kate's percent specific praise ranges from 0% specific praise to 100%. Brendan's data range from 17- 76% and Lilli's from 19-66%. The mean percent specific praise ranged from a low of 35% for Lilli's data and a high of 51% for Kate. The mean percent specific praise for Brendan and Ginny during Baseline was approximately 45%.

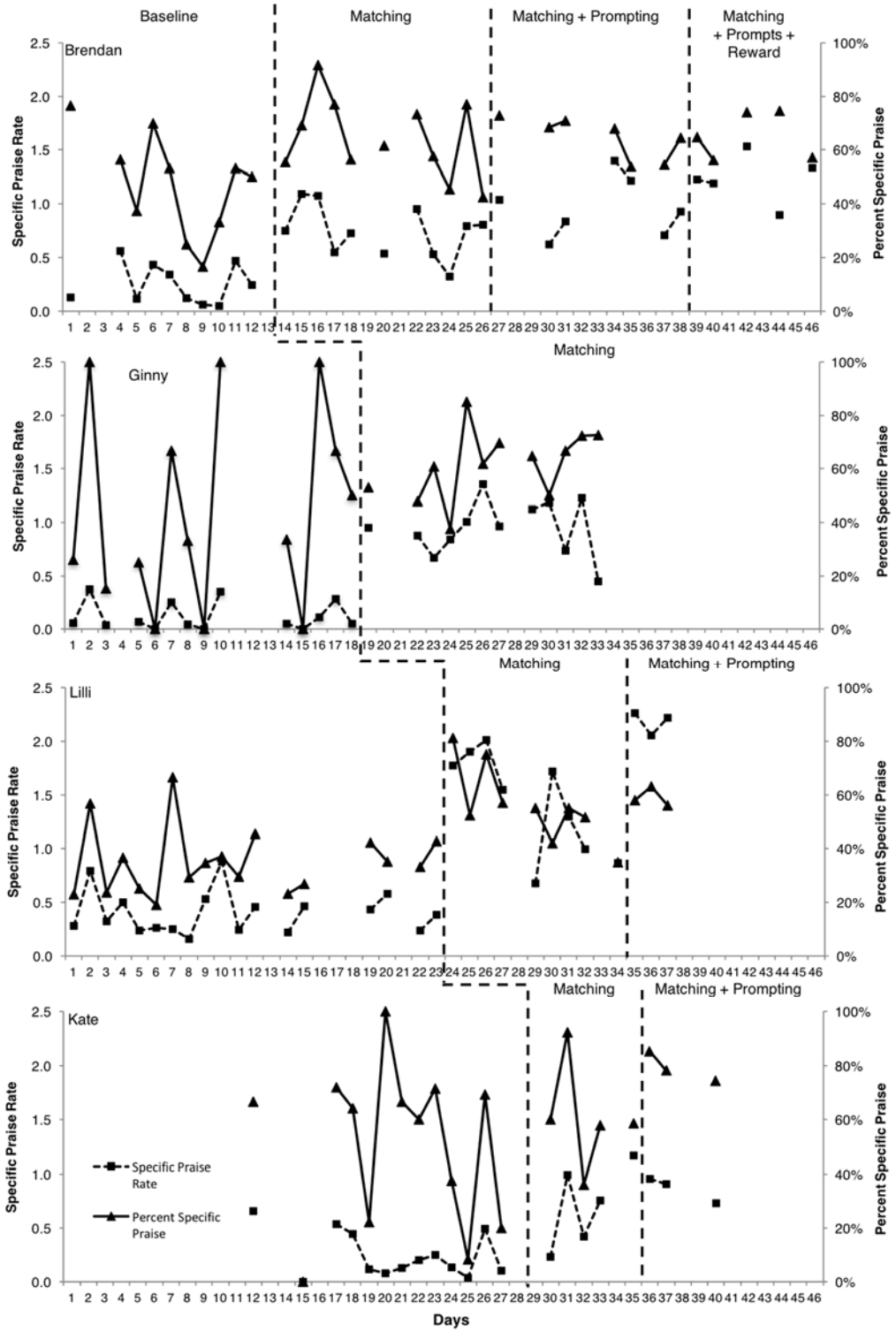


Figure 3. The percent of praise coded as specific praise plotted with the specific praise rates for all participants. The percent specific praise is represented using a triangle symbol and a solid line. These data are plotted using a secondary axis to the right. The specific praise rate is represented using a square and dashed line on the primary axis.

After the start of the intervention phases, Brendan's percent specific praise data were consistently higher than his baseline data and less variable. Brendan's mean percent specific praise increased to 64% during the Matching phase and remained at about this level through the remaining phases. His data during the Matching phase ranged from 42-92% specific praise. Further reductions in variability were noted in the Matching + Prompts (range = 54-73%) and the Matching + Prompts + Rewards (range = 56-75%) phases. It is also interesting to note that nearly every data point after intervention exceeds 50% specific praise, with notable exceptions occurring on days 24 and 26. It is also interesting to note that Brendan's specific praise rate exceeded his general praise rate during 40% of the Baseline sessions. This number increased to 82% during the Matching phase, and 100% during the Matching + Prompts and Matching + Prompts + Rewards phases.

During the intervention phases similar changes were noted in Ginny's data. Ginny's mean level of percent specific praise increased from 44% during Baseline to 62% during the Matching phase. The most dramatic difference, however, was the reduction in variability from the Baseline phase (range = 0-100%) to the Matching phase (range = 38-85%). A small positive trend was associated with her data across both phases. Only 36% of data points exceeded 50% specific praise during Baseline whereas 75% of data points exceeded this threshold during the Matching phase. It is also interesting to note that the Ginny's specific praise rates exceeded her general praise rates on 36% of the Baseline sessions. During the Matching phase, this percentage increased to 83%.

Lilli's data during the Matching phase are more difficult to interpret. Lilli used specific and general praise more frequently than the other participants. During the intervention phases, her use of general and specific praise increased. Lilli's percent specific praise increased from a mean of 35% during Baseline to approximately 56% after the start of the intervention. No distinct change in variability was noted until the start of the Matching + Prompts phase (range = 56-63%). However, the percent of data points exceeding 50% specific praise during Baseline was only 11% for Lilli. Following the introduction of the intervention procedures this percentage increased to 78% during the Matching phase and then 100% during the Matching + Prompts phase. The percent of specific praise rates that exceeded the corresponding general praise rates was 5% during Baseline, 78% during the Matching phase, and 100% during the Matching + Prompts phase.

Finally, there were no clear level changes in Kate's percent specific praise until the start of the Matching + Prompts phase. Her mean percent specific praise during Baseline was 51% (range = 0-100%). During the Matching phase her mean percent specific praise increased to 61% (range = 36-92%) and then 79% (range = 74-84%) during the Matching + Prompts phase. Similarly, the percent of data points exceeding 50% specific praise was consistently high (range = 67-100%) across all phases for Kate. The percent of specific praise rates that exceeded the general praise rates during Baseline was 62% compared to 80% and 100% in the Matching and Matching + Prompts phases, respectively. However, the limited number of data points during the Matching + Prompts phase makes it difficult to determine the effects of the intervention on the percent specific

praise during this final phase. Table 8 summarizes the descriptive statistics and the percent of data points above 50% specific praise for each participant during every phase.

Generally, the four teachers in this study were more likely to use specific praise after the start of the intervention than during Baseline. This pattern was consistent throughout the study even as specific praise and general praise rates increased throughout the duration of the study. Unfortunately, the magnitude of this effect for Lilli and Kate is insufficient to justify stronger conclusions regarding the effects of the intervention on specific praise relative to general praise.

Table 8

Descriptive Statistics for the Percent Specific Praise for Each Participant by Phase

Participant and variables	Baseline	Matching	Matching + prompts	Matching + prompts + rewards
Brendan				
Mean (%)	47	64	65	65
Range (%)	17-76	42-92	54-73	56-75
SD	19	14	8	9
Percent above 50%	50	82	100	100
Ginny				
Mean (%)	44	62		
Range (%)	0-100	38-85		
SD	37	13		
Percent above 50%	36	75		
Lilli				
Mean (%)	35	56	59	
Range (%)	19-66	35-81	56-63	
SD	12	14	4	
Percent above 50%	11	78	100	
Kate				
Mean (%)	51	61	79	
Range (%)	0-100	36-92	74-85	
SD	30	20	6	
Percent above 50%	67	80	100	

Note. Percent above 50% = the percentage of data points exceeding 50% specific praise.

Social Validity

Teachers responded to a modified version of the Intervention Rating Profile-15 (IRP-15). The IRP-15 was modified for use in this study by removing questions that addressed specific changes in students' behaviors as a result of participation in the study. Six free response items were also added to gather participant feedback on the various components of the intervention. Ginny, Lilli, and Kate completed a paper version of the IRP-15 after their last research session. Brendan completed the IRP-15 online using a google form.

Table 9 summarizes teachers' responses to selected IRP-15 questions. Participants responded to brief questions regarding their personal experience during the intervention

Table 9

Participant Responses to the Modified Intervention Rating Profile (IRP-15)

Question	Brendan	Ginny	Lilli	Kate	Mean
This intervention was effective in changing my behavior.	5	5	5	5	5
I would suggest the use of this intervention to other teachers.	6	4	5	5	5
I would be willing to use this intervention in the classroom setting.	5	6	6	6	5.75
This intervention would *not* result in negative side effects for the child.	5	3	5	4	4.25
This intervention would be appropriate for a variety of teachers.	5	5	6	5	5.25
This intervention improved students' behavior.	4	4	5	5	4.5
The time and effort required to participate in this intervention is reasonable.	5	4	5	6	5
I liked the procedures used in this intervention.	4	3	5	5	4.25
Overall, this intervention was good for our students.	5	4	6	6	5.25

Note. 1= Strongly disagree; 2 = Disagree; 3 = Slightly disagree; 4 = Slightly agree; 5 = Agree, 6 = Strongly agree.

phases and their perception of the social value of the intervention outcomes using a 6-point scale ranging from strongly disagree (1) to strongly agree (6). Questions were worded such that higher ratings indicated higher levels of satisfaction with and confidence in the effects of the intervention procedures. Overall, participants indicated high levels of satisfaction with the intervention procedures with average scores above four on every item. The highest average rating was given on the item, "I would be willing to use this intervention in the classroom setting." In addition to universal agreement on the item, "This intervention was effective in changing my behavior" there appears to be consensus among participants that the intervention was effective and efficient. Lilli indicated she agreed or strongly agreed with all statements on the IRP-15.

The two lowest rated questions were "This intervention would *not* result in negative side effects for the child" and "I liked the procedure used in this intervention." Ginny and Brendan gave their lowest ratings to the latter statement indicating they enjoyed the intervention procedures less than the other two participants. Ginny and Kate gave their lowest ratings to the statement, "This intervention would not result in negative side effects." In light of Kate's strong agreement with the item, "Overall, this intervention was good for our students," it is difficult to interpret her perception that these procedures may have negative side effects. Other than lower scores on the aforementioned items, Kate and Brendan agreed or strongly agreed with a majority of items on the survey.

Ginny's responses were the least positive of the four participants. She expressed slight disagreement with two statements and expressed less positive views relative to her peers on several other items. Unfortunately, Ginny did not indicate why she felt the

intervention might have a negative effect on students in the open-ended section of the social validity survey. She also did not specifically identify features of the intervention that were ineffective. She did, however, express frustration with the technology used in the study. She specifically mentioned problems with the Wi-Fi connection at her site. The West site's Wi-Fi security settings did not permit devices to remain connected to the network when they entered standby mode. This meant that every day Ginny had to connect the device to the network by entering an email address username and a 14-character password. Although incidental to the core features of the intervention, this inconvenience may have been enough to frustrate Ginny and negatively bias her overall perceptions of the procedures.

Teachers also responded to open-ended questions regarding what they liked and did not like about the self-evaluation, student feedback, and visual performance feedback components of the intervention. Teachers all identified the simplicity and responsiveness of the procedures as benefits across all three intervention components. For example, they reported that they spent less than 5 minutes distributing the Kindles to their students and helping them complete the online survey. Two teachers also noted that they wanted greater specificity in the reporting of the student feedback. Specifically, they wanted to see the percent of students reporting specific praise rather than the student rating. They indicated this would be a more sensitive measure of the changes they were making in their use of praise.

However, only one teacher mentioned the match score in her description of what she liked about the visual performance feedback. Two teachers mentioned the student

feedback component as a specific benefit of the information they received in the visual performance feedback. These teachers also indicated that they wanted more control over the data they were shown and more specificity in the percentage of students reporting specific praise. It is unclear if the participants were using the match score to guide their decision-making or focusing solely on the student feedback. Additional clarification is needed to determine if this is an area for improvement in the design of this intervention.

Evaluation of Student Feedback

Students were asked to indicate whether their teachers praised them and if they were told why they were being praised on the Student Feedback Form. **S**tudents' reports of general and specific praise are summarized in Table 10. Overall, the percent of students reporting general and specific praise increased along with praise rates for all participants except Kate. The percent of students in Brendan's group who reported specific praise was on average 47% (range = 25-71%) during the Matching phase, 56% (range = 32-75%) during the Matching + Prompts phase, and 78% (range = 64-91%) during the Matching + Prompts + Rewards phase. The average percent of student's

Table 10

Summary of the Percent of Students Reporting That They Were the Recipients of Specific Praise During the Intervention Phases

	Matching		Matching + prompts		Matching + prompts + rewards	
Participant	M (%)	Range (%)	M (%)	Range (%)	M (%)	Range (%)
Brendan	47	25-71	56	32-75	78	64-91
Ginny	65	35-86				
Lilli	81	64-100	88	67-100		
Kate	80	55-100	73	53-85		

reporting specific praise in Ginny's group during the Matching phase was 65% (range = 35-86%). On average, 81% of students (range = 64-100%) in Lilli's group reported receiving specific praise during the Matching phase and 88% (range = 67-100%) during the Matching + Prompts phase. During the Matching phase, 80% of Kate's students (range = 55-100%) reported receiving specific praise. Whereas only 73% (range = 53-85%) reported receiving specific praise during the Matching + Prompts phase. Kate's students were the only group whose reports of specific praise did not increase during the Matching + Prompts phase.

A rank order correlation was computed between the percent of students reporting specific praise and teachers' specific praise rates throughout the study. The results from this analysis are presented in Table 11. The correlations between specific praise rates and the percent of students reporting specific praise were statistically significant for all teachers except Ginny. The magnitudes of the significant correlations were all in the moderate to high range (Kachigan, 1986). However, for three of the four teachers, the correlations were below .60.

Table 11

Rank Order Correlations Between Specific and Total Praise Rates and Student Ratings

Participant	*n*	Correlation w/specific praise
Brendan	23	0.50*
Ginny	13	0.16
Lilli	12	0.57*
Kate	8	0.79*

Note. n represents the number of days with student and teacher ratings.
* $p < .05$.

CHAPTER V

DISCUSSION

The primary research question asked about the extent to which the use of the self-management procedure influenced teachers' specific praise rates. One of the secondary questions asked about the effects on general praise. The final two research questions focused on teachers' perceptions of social validity and evaluated the relationship between student reports and teachers' use of specific praise. Teachers evaluated the effectiveness and feasibility of the procedures using a survey completed after their participation in the study was complete. Student feedback was evaluated by comparing the percent of students reporting specific praise during each intervention phase and then correlating those percentages with their teachers' specific praise rates.

Effects on Specific Praise

All four participants increased their use of specific praise after they implemented the self-management procedures in their classrooms. To demonstrate a functional relation between the implementation of these procedures and increased specific praise rates, the experimenter would need to produce three demonstrations of the experimental effect at three different points in time (Horner et al., 2005). A demonstration of an experimental effect requires a clear change in level, trend, or variability associated with the implementation of the intervention. In a multiple-baseline design across participants, these demonstrations need to occur after the implementation of the intervention for each teacher and to occur without coinciding changes in the data from teachers who have yet

to experience the intervention.

An immediate level change in the specific praise rates for Brendan, Ginny, and Lilli were evident following the implementation of the self-management intervention. Brendan was the first teacher to implement the intervention and his specific praise rate increased by nearly three times his mean rate during Baseline. Coinciding changes were not evident in the data for Ginny and Lilli. Thus, these changes were most likely the result of implementing the self-management procedures. Ginny's specific praise rate increased immediately after she started using the self-management procedure and, again, neither Lilli's nor Kate's data showed corresponding changes. Finally, Lilli's use of specific praise increased following her implementation of the intervention. Kate's data on the second day of the Matching phase increased, but her specific praise rate returned to a near zero level the next day. An effect, based on a comparison of the level of specific praise rates during Baseline and the Matching phase, was demonstrated for three of the four participants at three different points in time.

A clear change in trend was also evident for Brendan and Lilli, unfortunately these trends were counter-therapeutic. Both participants had a slight decline in their specific praise rates during Baseline that may be the result of initial reactivity to the data collection procedures. Ginny's specific praise rates during Baseline and the Matching phase were both characterized by a slight downward trend, but the trends were very similar between the two phases. The most distinct change in trend occurred with Kate's data. Her Baseline data were characterized by a slight downward trend whereas her data during the Matching phase had a positive trend.

Finally, there was very little overlap between the praise rates observed during Baseline and those during the Matching phase for Ginny and Lilli. Approximately 36% of Brendan's specific praise rates during the Matching phase overlapped with his data during Baseline. Based on the results from this study, there is evidence of experimental control between the self-management procedure and teachers' specific praise rates.

The results from the additional intervention procedures suggest that these procedures positively influenced specific praise rates. Brendan's use of specific praise during the Matching + Prompts phase was more consistent and had a slightly higher level than his previous performance in the Matching phase. His mean specific praise rate during the Matching + Prompts + Rewards phase was higher than his mean rate in any of the previous phases as well. Kate and Lilli both exhibited more consistent specific praise rates during the Matching + Prompts phase.

Unfortunately, the introduction of these conditions was not systematically controlled so it is difficult to determine if these changes were the result of the introduction of these new procedures or simply the result of continuing to use the self-management procedures. Improved performance was associated with the addition of prompting. The addition of these procedures required little training and no external supervision or coaching. Previous researchers reported that prompts effectively increased specific praise rates (Silvestri, 2004; Van Houten & Sullivan, 1975), but the results of this study in regard to the effects of the addition of prompts are unclear.

Effects on General Praise

Secondary research question one addressed the effects of the intervention on participants' general praise rates. This question was a secondary research question because the focus of the training and visual performance feedback was on participants' use of specific praise. No effort was made to train, practice, or evaluate participants' use of general praise at any point in the study. The content of the didactic training used general praise statements as nonexamples of the target behavior, and teachers received no instruction to use general praise with their students. The content of the visual performance feedback contained no reference to general praise.

Despite the lack of training or feedback on general praise rates, the teachers' mean general praise rates increased during the Matching phase for all teachers except Kate. Kate's mean general praise rate increased during the Matching phase, but decreased to a level below her baseline mean during the Matching + Prompts phase. Although the average percentage of all praise statements that were specific increased during the intervention phases, the percentage of general praise statements accounted for an increasingly lower percentage of between 20-40% of all praise statements. There was also an increase in the variability of general praise rates after the start of the Matching phase for all participants. Despite these changes in general praise rates, there was insufficient evidence to conclude that the self-management procedures were functionally related to general praise rates.

This conclusion was no unexpected because the intervention was not designed to target general praise. The lack of support for general praise makes the overall conclusion

expected, but does not help explain the changes in variability noted during the intervention phases. Conceptually, the research on generalization suggests that behavior does not spontaneously occur in novel contexts without direct training and support (Stokes & Baer, 1977). The frequency and variability of general praise, however, increased without direct support in this study. One potential explanation of this result is that several features of the independent and dependent variables may have implicitly encouraged the use of general praise. For example, the experimenter used examples of specific and general praise throughout the training. Greater familiarity with general praise and its association with the target behavior, specific praise, may have been enough to increase general praise rates.

Alternatively, general praise may have increased because it is a subcomponent of specific praise. Teachers who intended to use specific praise but were disrupted, hurried, or otherwise prevented may have intentionally or unintentionally used general praise instead. In addition, general praise rates might have increased because general praise can be used in situations when it is difficult or impossible to use specific praise. The ease of expressing general praise may help explain why researchers report that the natural rate of general praise is higher than specific praise in many elementary and secondary classrooms (Burnett & Mandel, 2010; Jenkins et al., 2015; Reinke et al., 2007; 2008).

The effects of the self-management procedure on general praise rates were unclear. The data contained in Figure 3 and Table 8 suggest that specific praise accounted for the largest part of the changes in teachers' use of praise during the study. However, the use of general praise was not discouraged and there is no evidence in the literature

suggesting that general praise rates should be reduced categorically. The results from this study showed that participants were more likely to use specific praise relative to general praise while the intervention procedures were in effect. The impact of this change, however, is unclear because the study did not include a direct measure of students' academic or social behavior.

Social Validity

Another secondary research question addressed the social validity of the intervention procedures from the perspective of the teachers. Overall, teachers were overall very positive about their experiences during the study. Their responses on the items addressing the feasibility of the intervention indicate agreement or strong agreement across all items. These data are of particular importance because the intervention was designed with the goal of being simple to implement and easy to maintain. To be useful at the B&G Club and within other OST programs, teachers will need to perceive the procedures as simple and effective since they will be primarily responsible for fidelity of implementation.

Although there is little evidence that the teachers disliked any of the procedures, there is also only limited data on the procedures that were preferred by the teachers. In fact, multiple participants commented that they wished they had more access to the data and more detailed feedback from students. Only one teacher identified the matching procedure as a key part of the visual performance feedback. More teachers specifically identified the student feedback process and data in their descriptions of what they liked

and what they would like to see improved in the future. Additional data will need to be collected regarding user experiences to determine exactly how the matching procedure influences teacher perceptions of the procedures and changes in instruction.

The teachers indicated agreement or slight agreement for all items addressing the effectiveness of the intervention. These scores were slightly lower than the scores on the feasibility items. The lower scores for effectiveness should be interpreted with caution for several reasons. First, the teachers had limited experience with classroom management strategies. This lack of experience may have contributed to unreasonable expectations about the immediacy of the effects of the intervention. Even seasoned educators can struggle to maintain high praise rates because the effects of praise are rarely immediate and pronounced (Alber & Heward, 2000).

Second, the B&G Club did not have a system in place to support teachers dealing with problem behavior. The teachers were not required to use specific classroom management strategies such as explicit teaching of classroom expectations, prompting appropriate behavior, active supervision, or the use of predictable routines (Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). When egregious problem behaviors occurred the teachers notified the site supervisors and the students were removed from the activity. The lack of a supportive system of classroom management may have limited the effects of increased praise and made it more difficult for teachers to notice changes in student behavior. Despite these issues, all but one teacher indicated they believed the intervention had a positive impact on their students.

Evaluation of Student Feedback

Finally, this study examined the correlation between the percent of students reporting specific praise and the observed specific praise rates of their teachers. If a greater percentage of the students' reported receiving specific praise after their teachers' specific praise rates increased, then the procedures produced noticeable changes in the classroom environment that could be reported by students. This analysis confirmed that students' reports of specific praise were related to teachers' specific praise rates for a majority of the participating teachers. Three out of four correlations showed a statistically significant relationship. The correlations between student reports of specific praise and teachers' specific praise rates ranged from .16 to .79 in this study. The magnitude of these relationships was higher than those reported by researchers comparing external observer ratings with student feedback (range = -.17-.38; DeJong & Westerhof, 2001).

The results reported in this study provide additional support for the validity of student feedback (Alber & Heward, 2000; Gray et al., 1974; Measures of Effective Teaching, 2012a). In regard to the utility of student feedback, this study is unique because very little research has been conducted on the effects of student feedback on teacher behavior using an experimental research design.

However, the strength and consistency of this relationship needs to be examined systematically in a variety of educational settings and with different student groups. One area of concern that should be considered by future researchers is the training students received prior to completing the Student Feedback Form. Given the poor attendance at all sites, it is very unlikely that all students received the initial training prior to

completing the feedback procedure for the first time. In addition, no booster trainings were held to provide support to new students and to encourage consistency with the operational definitions of specific praise used in the training. The effects of poor student training should have diminished the magnitude of the correlations between specific praise rates and the percent of students reporting specific praise, but the observed correlations in this study were mostly large and statistically significant. It is possible, however, that even higher correlations would have been observed had more students received training initially and throughout the study.

Implications

The results from this study contribute to the research literature beyond improving conditions for the B&G Club. First, it extends the available literature on staff development strategies tested in OST programs. This unique context is applicable to an ever-increasing number of students participating in these programs. Second, this study represents the first time the matching component has been incorporated into a self-management program for teachers. Matching has the potential to address the limitations of self-evaluation by providing a mechanism to calibrate and enhance self-evaluations through student feedback, a cost-effective source of feedback that has been shown to be significantly related to critical student outcomes. Finally, this study extends the literature on student feedback by using student feedback within the context of an intervention. There have been no studies to date, on the effects of student feedback on teachers' use of praise. This intervention package provides an opportunity to examine the feasibility and

effectiveness of student feedback. The results of this study may inform all of these applications, but their ultimate value to the field will depend on replication.

Increasing Praise Rates

A variety of strategies including training, performance feedback, coaching, and self-management are used to help teachers increase their use of praise in the classroom (Ackland, 1991; Cavanaugh, 2013; Lu, 2010), but very few researchers have focused on the effects of these procedures with OST staff members. This study is one of only a few experimental studies in which researchers examined the utility of a self-management program without expert coaching. The B&G Club needed effective strategies for staff development that were efficient, cost-effective, and required few staff to implement. These strategies needed to be efficient in their implementation because the B&G Club has high staff turnover. Over the course of the summer program, two of the five participants originally recruited for the study chose to leave their positions early.

The results from this study suggest that the intervention procedures used in this study had an impact on inexperienced teachers' use of specific praise. Briere et al. (2015) showed similar changes in specific praise rates with new teachers. Briere et al. paired new teachers with experienced mentors who met periodically to review performance feedback and set goals with their mentees. The specific praise rates for two of the new teachers after the start of consultation with their mentor increased to approximately 1.5 specific praise statements per minute. The third teacher's specific praise rate steadily climbed from near 0 during baseline to a steady level near one specific praise statement per minute. Lilli and Ginny both achieved specific praise rates near 1.5. Brendan and

Kate emitted specific praise rates approaching one statement per minute. Similar to the third teacher in the Briere et al. study, Kate did not immediately respond to the training and her praise rate steadily increased until the start of the Matching + Prompts phase. Her final data point during the Matching phase exceeded one specific praise statement per minute for the first and only time during the study.

The changes in specific praise rate achieved by the participants in this study were comparable to those from previous research (Briere et al., 2015; Keller et al., 2005; Mesa et al., 2005), but the intervention procedures were not. The behavioral changes in this study were the result of a self-management procedure largely managed by the participating teachers and their students. Although the experimenter provided a brief training and limited ongoing technical assistance, experienced coaches, mentors, data collectors, and performance feedback systems were not required to alter specific praise rates. Compared to the extensive coaching and support programs detailed in the literature (Briere et al., 2015; Keller et al., 2015) the procedures involved in this study required very little training and support to initiate and were effective with instructors with no prior classroom management training.

In addition, the self-management procedures may benefit students as well. Specific praise has been shown to increase student engagement, reduce problem behaviors, and enhance student-teacher relationships (Mesa et al., 2005; Rathel et al., 2014; Sutherland, 2000). The teachers indicated on the modified IRP-15 that the procedures had a noticeable impact on the behavior of their students. Unfortunately, no direct measures of student academic or social behaviors were collected during this study

so it is difficult to determine exactly how student behavior changed as a result of increased specific praise.

Students also learned to identify praise during the study. Prior to training, many students could not define, identify, or describe the purpose of praise in their classrooms. Some students indicated they believed their teachers did not praise them because they liked other students more. No data were collected on student responses during the training so it is impossible to say how many students accepted this idea of why their teachers praised some students and not others, but no student during the four student trainings linked the teachers' use of praise with good behavior. Thus, one immediate benefit to students was learning to identify praise statements. Greater awareness of praise may have improved relationships between teachers and students.

Gray et al. (1974) reported that students had inaccurate perceptions of the use and purpose of praise in their classrooms. Gray et al. provided training to these students to help them accurately identify the frequency of praise. After training the students' perceptions of the use of praise aligned with the data collected from external observers in the classroom. Once the students could accurately identify praise they expressed greater levels of satisfaction in the classroom and an improved sense of self-efficacy. The data from this study indicate that a higher percentage of students reported receiving specific praise during the intervention phases, but we do not have a measure of their satisfaction with this outcome or their self-efficacy with the process.

Self-Management Procedures for Teachers

The self-management procedure used in this study mirrored effective strategies

used with struggling adolescents (Young et al., 1991). Although these procedures were effective with students, their effectiveness with teachers was untested. The results from this study suggest that self-management programs utilizing matching between teachers' self-evaluations and student feedback can be effective. The strength of the experimental effect and the use of an experimental design to evaluate these results lends support to these findings and extends the existing literature base on self-management programs (Keller et al., 2005; Simonsen et al., 2013; Sutherland & Wehby, 2001).

The use of the matching feature to align teachers' ratings with those from students is the most unique component of this approach to self-management. Matching was included as a means of increasing the strength of the intervention relative to previous studies using only self-monitoring (Simonsen et al., 2013). The results from the Matching phase confirm that this is a viable strategy for teachers, but the need for additional intervention procedures suggests that further research is needed to identify stronger treatment packages. The use of a treatment package also makes it difficult to ascertain to what degree the use of matching improved the effectiveness of the intervention relative to the contributions of visual performance feedback, self-evaluation, and student feedback. Future research will need to disentangle the effects of the various procedures used in this study.

However, there is some evidence to suggest that more consistent matching was associated with better performance. Perfect matches appeared more frequently later in the Matching phase for Lilli and Ginny, but not for Brendan and Kate. Kate and Brendan were also the only participants to receive a "No Match" score, indicating that their self-

ratings and the student ratings deviated by more than two rating levels. Given the stronger response to the treatment for Lilli and Ginny, this may be an important variable for future research. However, it is difficult to determine exactly why Lilli and Ginny were better at matching than their peers. They may have been more engaged in the intervention procedures, had stronger relationships with their students, or felt more confident altering their own behavior to increase the likelihood of a perfect match.

The Utility of Student Feedback

To increase the feasibility of the self-management procedure the intervention was designed to be self-contained within a classroom. In other words, teachers could implement the self-management procedure without a peer coach or external observer to collect performance feedback. This was accomplished by using students as the source of feedback in the classroom. Student focused self-management procedures have used teachers as secondary observers with whom the student self-evaluations are compared, but little attention has been given to the value of student feedback in self-management programs designed for teachers (Young et al., 1991).

This study contributes to the research that demonstrates that young students can be taught to implement strategies that increase preferred teacher behaviors (Alber & Heward, 2000; Craft, Alber, & Heward, 1998; Gray et al., 1974). These studies demonstrate that students can be directly involved in efforts to improve the classroom learning environment. The data, however, do not support the conclusion that these procedures have a direct benefit to students. These results confirm that student feedback can be useful to improve teachers' use of specific praise. More research is needed to

determine under what conditions students can provide feedback to improve the learning conditions in their classrooms and the effects of these procedures on the academic and social behavior of participating students.

Barriers to Collecting Student Feedback

Several reasons might explain why these procedures are not widely adopted. First, teachers face a variety of demands on their time including lesson preparation, collaboration time, grading, communicating with stakeholders, delivering instruction, and developing new interventions. Even more constrained is their instructional time with students, time often disrupted by interventionists, specialists (i.e., music, P.E., etc.), school wide events, and other activities. Therefore, it is not surprising that teachers hesitate to dedicate time to teaching students to complete a feedback procedure and the time to regularly collect the data. To address these concerns, the student feedback and evaluation system in this study was designed to collect and report student feedback very efficiently. These data were immediately processed and available in a report that could be viewed on a mobile device immediately after class or anytime throughout the day. Every participant noted the ease and simplicity of these procedures. However, there were also costs associated with securing and maintaining the technology that made this study possible.

Second, some teachers may object to the premise of asking students to influence or shape their behavior on philosophical grounds. Some teachers may reject the notion that students should have a more active role in shaping their behavior. Teachers may also lack confidence in the validity and reliability of student feedback. One way the

procedures in this study addressed this concern was restricting student feedback to very specific instructional events. This was intended to restrict the role of the student in teacher development and to increase teachers' confidence in the quality of the feedback they received from their students. Students were not asked to rate the quality of the instruction they were receiving. The fact that multiple teachers asked for more detailed reports of student feedback suggests that the teachers wanted to better understand the feedback they were receiving from their students and were not dismissing the feedback.

Potential Confounds in Student Feedback

One important issue that has yet to be discussed is the potential confound between student attendance and student feedback. In classroom environments with relatively stable attendance, student feedback may be more reliable and valid because students are more familiar with the process, more capable of making discriminations because they have more experience observing their teacher, and fewer confounding variables are present to influence the data. Unfortunately, the student attendance at the B&G Club could not be characterized as stable. This is primarily a concern because student reports were aggregated and reported to teachers as a rating based on the percent of students reporting specific praise. Thus, it is possible that volatile student attendance explained more of the variance in students' reports than the teachers' use of specific praise.

Two sources of attendance data were used to explore this relationship. First, the B&G Club staff recorded the names of all students who officially checked in with the program each day. Based on these records, the average daily attendance for the South site was 38 students (*SD* = 7.00; range = 16 to 51), the West site was 30 students (*SD* = 7.36;

range = 15 to 42), and the East site was 59 students (*SD* = 9.96; range = 41 to 90). The number of students in attendance was consistently higher on Tuesday and lower on Friday, Monday, and everyday adjacent to a holiday. Highly variable and cyclical attendance patterns were documented at every site, but these daily student counts do not reflect the actual number of students attending the research session each day because of alternative activities at the site and students were not removed from the overall count if they left prior to the research session.

Based on the records maintained by the B&G Club staff members, there is no way to determine the exact number of students who attended a specific teachers' research session on a given day. The best estimate of this number was the survey response counts collected during the intervention phases of the study. The average daily attendance based on these data is lower than the aforementioned attendance rates because it reflects only those students who chose to attend the research session, stayed until the end of the session, and completed the Student Feedback Form. The average number of completed surveys at the South program was 14 (*SD* = 6, range = 6 to 30). On average, 17 students completed surveys in Lilli's group (*SD* = 6, range = 6 to 27) and 12 in Kate's group (*SD* = 4; range = 4 to 17) at the East site. Ginny's groups at the West site averaged 13 completed surveys (*SD* = 3, range = 7 to 17).

A rank order correlation was computed to examine the potential confounding relationship between attendance and student reports. The number of student surveys submitted was used as the measure of student attendance in the research session. The rank-order correlation was computed using the same procedures outlined in Chapter II.

The correlations were -0.16, -0.18, -0.45, and -0.14 for Brendan, Ginny, Lilli, and Kate, respectively. None of these correlations were statistically significant.

All correlations between attendance and student reports were negative. This suggests that as the number of students in the session increased, the percent of students reporting specific praise decreased, or vice versa. The consistent negative correlations between attendance rates and the percent of students reporting specific praise makes logical sense because as the number of students in the classroom increases it becomes more difficult to praise every student. However, none of these correlations were statistically significant so the directionality and magnitude of the relationships are not significantly different from zero. Factors including praise rate, distribution of praise, and other attributes of instruction offer a more parsimonious explanation of the variability in student feedback than attendance.

It is interesting to note that the magnitude of the correlation between student reports and attendance in Lilli's classroom was nearly double those of her peers. Lilli was the only participant with a percentage of students reporting specific praise consistently above 80% and she had the largest student groups of any teacher, on average. Her relatively high specific praise rates during the intervention phases suggest that despite her larger group size she was capable of providing lots of specific praise to a high percentage of her students, but the larger correlation ($r_s = .45$) indicates that class size was related to student reports of specific praise. However, the relationship between specific praise rates and the percent of students reporting specific praise ($r_s = .57$, $p < .05$) was significant and of greater magnitude than the relationship between attendance and student reports. This

suggests that although Lilli's use of praise was uniquely sensitive to the number of students in her classroom, Lilli's student reports were influenced more by her use of specific praise than attendance.

The student feedback system used in this study was part of a treatment package that produced changes in teachers' specific praise rates. It was cost-effective, efficient, and provided meaningful feedback to participating teachers. The teachers reported that they would use the procedures in future classrooms and that they generally considered the procedures effective for themselves and their students. The percentage of students reporting specific praise increased as teachers' specific praise rates increased. This provides preliminary support for the validity of the student feedback collected through the Student Feedback Form. However, more research is needed to determine if these strategies are feasible in other OST programs and classroom settings.

Overall, these finding are not unexpected. Students are uniquely positioned to report on instructional events because they are present in the classroom every day. Furthermore, students are the primary consumers of educational services. Their perceptions of teacher behavior and the quality of instruction are vitally important to developing a productive classroom learning environment.

Limitations

The behavior changes documented in this study were not as dramatic or as consistent as expected. Specific praise rates between 1.0-1.5 statements per minute were achieved by Ginny and Lilli during the Matching phase. To put this in perspective,

consider the change in Lilli's data from Baseline to the Matching phase. An increase from a mean specific praise rate of 0.4 to 1.42 translates into an increase from 12 specific praise statements to over 50 statements during a 30-minute research session. With an average class size of 17 students, the ratio of specific praise statements per student was less than one during Baseline, assuming an equal distribution among students. During the Matching phase, this ratio increased to nearly three specific praise statements per student. Kate and Brendan had higher rates of specific praise during the Matching phase relative to their baseline data, but neither achieved the same degree of change evident in the data collected from Ginny and Lilli. Specific praise rates over one per minute are commensurate with the rates observed by new teachers and other professionals in similar studies (Barton & Wolery, 2007; Briere et al., 2015; Simonsen et al., 2013) as are the volatility and counter therapeutic trends evident in three of the four teachers' data. These trends suggest that more research is needed to find strategies that are equally cost-effective but produce sustainable improvements.

There are several possible explanations for the volatility in the data. First, the self-management procedure did not include explicit reward contingencies as implemented by Young et al. (1991). These authors used implicit and explicit incentives to encourage participation and performance in the self-management program. Although the implicit reward contingencies were not intentionally manipulated or controlled in previous studies, the hierarchical relationship between a student and a teacher created many opportunities for implicit incentives to shape student behavior. For example, because the teachers' control a variety of negative and positive consequences valued by their students

their approval can serve as a conditioned reinforce and their disapproval a conditioned punisher. Thus, control over the availability of conditioned reinforcement and punishment may enhance the value of the teacher's feedback, attention, and praise to a student.

Sources of implicit incentives were not as easy to identify in the current study. Teachers in the B&G Club were not held accountable for the performance of their students, they did not have extensive contact with parents, and there was no opportunity for additional incentives outside of their regular pay for improving their use of specific praise. The implicit incentives present in other self-management programs were presumed absent in this study because students at the B&G Club have little control over the sources of reinforcement for their teachers. However, this may not be true in every case. Some of the teachers may have been reinforced by contributing to improvements in the academic and social behavior of their students. For these teachers, higher student ratings may have been reinforcing.

Explicit rewards in previous self-management studies are delivered based on meeting key objectives. For example, as students improved their classroom behavior and became more accurate in their self-evaluations, they could earn preferred rewards such as preferred privileges (e.g., time to work with peers, a break from instruction, etc.), food, access to technology, or school supplies. These contingencies clearly identified the performance criteria associated with earning each incentive and gave the teacher the flexibility to adjust the criteria to encourage consistent improvement in behavior.

There were no explicit, programmed rewards delivered by the experimenter

during the Matching or Matching + Prompts phases of this study. The teachers collected and reviewed their data independently without any promised explicit incentives managed by the experimenter or the site supervisor. Incorporating an explicit reward contingency and active management of this contingency by the site supervisor are the two critical features of the Matching + Prompts + Rewards phase. Brendan met with his supervisor to determine a performance goal and select a reward for meeting the goal. His goal was achieving an average student rating greater than four over a four-day period and a perfect or next-door match each day. If he achieved his goal, the research team provided lunch for the B&G Club staff members at the South site. Brendan's specific praise rate during this period was slightly higher and more consistent than his performance during the previous intervention phases. These data suggest that explicit reward contingencies may enhance the effectiveness of the self-management procedure. However, these procedures come at a significant cost to the efficiency and feasibility of the procedures because they require external support and management.

Another limitation of this study is the lack of a systematic approach to the implementation of the additional intervention procedures. The implementation of these phases was accelerated because of a scheduling conflict between the local school district and the East site. This conflict was resolved by eliminating an entire week of the B&G Club summer program. As a result, the Match + Prompts phase was implemented with Kate almost immediately after Lilli started using the new procedures. A similar situation occurred with Ginny, who chose to end her employment with the summer program 2 weeks early to move to college, but in her case she was responding well to the Matching

phase procedures and her movement to a new phase was not justified. Regardless of these logistical challenges, the lack of a controlled implementation of these procedures limits the conclusions that can be drawn about the effects of these procedures on teachers' specific praise rates. There are other strategies including brief consultation, self-evaluation using the recorded audio tracks, and self-monitoring strategies that might have increased the potency and maintenance of these procedures (Briere et al. 2015; Simonsen et al., 2013; Keller et al., 2005). However, it is valuable to know that the low-cost, self-management procedures used in this study produced increased specific praise rates without more intensive procedures.

The lack of a direct measure of the impact of the intervention on student behavior is an important limitation. Although the increased use of specific praise has been linked to improvements in student behavior and student-teacher relationships (Mesa et al., 2005; Rathel et al., 2014), there were no data collected in this study to support the conclusion that increased praise rates improved the learning environment or performance of students in this study. One teacher remarked, "Wow! It's amazing how much smoother it goes when you can use specific praise! It was such a good day"! Aside from this anecdotal report, there is no evidence that the changes in teachers' specific praise rates corresponded with similar improvements in students' academic and social behavior. These data are also critical to determine the ultimate feasibility and efficiency of these interventions because if the procedures only change instructors' verbal behavior and do not produce corresponding changes in student performance then their feasibility is of little value. These data might also be helpful in determining why the effects of the

intervention were not maintained across all participants throughout the Matching phase. For example, if student behavior did not immediately improve with the increased use of specific praise then teachers may have found it difficult to maintain high rates with only the student feedback to encourage their use of praise.

The use of a treatment package designed around the self-management strategies to adolescents (Young et al., 1991) limited the conclusions that could be drawn regarding the effects of specific intervention components on teacher behavior. This was anticipated and simply means that future analytic research will be required to reveal the impact of individual components. The implementation of the treatment package is also a concern. Kate's data are difficult to interpret relative to those of her peers for two reasons. First, Kate was trained by her supervisor. The implementation fidelity checklist indicate that multiple features of the training were not completed during her training, including generating specific examples of specific and general praise. Kate's specific praise rate did not immediately increase after training, unlike the rates of the other participants on the first day of the Matching phase. Second, Kate was the only participant who shared a site and students with another participant. The degree to which prior potential treatment diffusion occurred due to learning about the intervention from Lilli or the FANZ students' experiences completing the procedures in Lilli's classroom and then sharing them with Kate is unknown.

Ginny's fidelity of implementation may have impacted her results as well. The phone assigned to Ginny for data collection contained only a small proportion of her visual performance feedback records. This may have been the result of Ginny deleting the

records, failing to view the data, or she may have toggled an option allowing the phone to replace rather than rename similar files. The latter seems the more likely possibility because her device contained no archived visual performance feedback files for the first seven days of the Matching phase and then a complete history for the next 5 days. Neither Ginny nor her supervisor reported having problems viewing the data. Ginny reported it was challenging to improve her performance, suggesting she was actively attending to the feedback.

Finally, the extent to which behavior changes generalized to other instructional environments outside the research sessions was not assessed. To be achieved, generalization should be programmed (Riley-Tillman & Eckert, 2001; Stokes & Baer, 1977) and no effort was made to provide this kind of programming within the implementation of the treatment package. In addition, maintenance of the treatment effects was not adequately addressed. The presence of decreasing trends in the data from Brendan and Lilli indicates that this may be an important area for future research. The presence of counter-therapeutic trends in Brendan's and Lilli's data during the Matching phase led to the development of additional intervention procedures to try to increase the potency of the intervention. As a result, maintenance and generalization were not examined in this study.

Future Research

In the future researchers should explore methods to further simplify and streamline the use of student feedback to enhance teacher self-management. The

technology used in this study enhanced data collection through the use of mobile audio recorders embedded in mobile phones; coding and data security through the use of cloud storage and management tools; and the collection and reporting of self-management data using electronic forms and databases. Despite these considerable advantages, participants and the research team alike experienced frustration with unreliable connections and unexpected glitches in the system. For example, the research team learned quickly that everyone previewing and coding tracks had to use the same web browser because each web browser displays the time stamp on the audio track differently resulting in biased calculations of observer agreement. In the future researchers should explore alternatives to these strategies to further streamline the collection and reporting of feedback for teachers and other participants.

In addition, in the future it would be helpful to understand more about the effects of improved instructional quality on student engagement, academic performance, social interactions within the classroom, teacher satisfaction, and other critical outcomes. These could also include measures of maintenance and generalization to explore how robust changes in teacher behavior are over time and in different instructional contexts. Some modifications that may be worth considering are fading away key components of the intervention to see which components produce the greatest impact on teacher's behavior. It may also be useful to consider broadening the focus of the self-management procedure beyond praise to include instructional strategies such as error correction, giving instructions, or implementing evidence-based practices.

Replicating this study in other school programs with new or veteran teachers

would help expand our understanding of the utility of this program beyond OST programs and the B&C Club. Future researchers could also explore the effects of using school or program staff to provide the teacher training and technical assistance during the study. Kate's experience suggests that program staff may need careful monitoring and support to effectively deliver these trainings and to support staff with various technology issues during implementation.

Finally, more research is needed to identify the most effective and efficient strategies for using student feedback. In this study, students reported the frequency of praise during the previous academic activity. In the future, researchers might explore the effectiveness of providing feedback over longer periods of time or on more than a single instructional strategy. In addition, it might be helpful to know more about how teachers interpret and use student feedback. Based on the results from this study, access to student feedback and self-evaluation data in the form of visual performance feedback influenced teachers' specific praise rates, but no measures were collected on teachers' use of these data to determine the most influential features of the independent variable. The teachers also indicated that they would like more detail concerning the student feedback they received.

Conclusion

The quality of OST programs is a concern for the parents and teachers of the 7 million students who participate in these programs (Carver & Iruka, 2006). Unfortunately, the expertise and resources necessary to ensure that every teacher provides

high quality instruction for every student, every day is not universally available. This study explored an approach to supporting teachers that was simple and cost-efficient in an OST program. The results of the study confirm that this modified self-management program using student feedback for matching was effective and valued by teachers. This study contributes to the existing literature by demonstrating the utility of student feedback as an integral component of self-management and by demonstrating that performance improvements can be made without extensive, ongoing coaching. In most studies, researchers relied heavily upon external observation and coaching to achieve similar gains, but this program required limited training and technical assistance to improve praise rates. In addition, this study provides evidence that self-management programs can be developed and used to support inexperienced teachers in an OST program. Additional research will be required to refine our understanding of the conditions, contexts, populations, and programs where self-management programs using student feedback can be effective.

REFERENCES

Ackland, R. (1991). A Review of the peer coaching literature. *Journal of Staff Development*, *12*(1), 22-27.

Afterschool Alliance. (2015). *Evaluations backgrounder: A summary of formal evaluations of afterschool programs impact on academics, behavior, safety and family life*. Washington, DC: After School Alliance. Retrieved from http://www.afterschoolalliance.org/

Alber, S. R., & Heward, W. L. (2000). Teaching students to recruit positive attention: A review and recommendations. *Journal of Behavioral Education*, *10*(4), 177-204.

Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, *13*(2), 153-166. Retrieved from http://doi.org/10.1023/A:1008168421283

Allinder, R. M., & Beckbest, M. A. (1995). Differential effects of two approaches to supporting teachers' use. *School Psychology Review*, *24*(2), 287-298.

Allinder, R. M., Bolling, R. M., Oats, R. G., & Gagnon, W. A. (2000). Effects of teacher self-monitoring on implementation of curriculum-based measurement and mathematics computation achievement of students with disabilities. *Remedial & Special Education*, *21*, 219-226. Retrieved from http://doi.org/10.1177/074193250002100403

Anderson, L. M., Evertson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first-grade reading groups. *The Elementary School Journal*, *79*, 193-223.

Andrews, L., & Kozma, A. (1990). Increasing teacher praise and altering its distribution to children of differing on-task levels. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, *22*(2), 110-120. Retrieved from http://doi.org/http://dx.doi.org.dist.lib.usu.edu/10.1037/h0078897

Apsler, R. (2009). After-school programs for adolescents: A review of evaluation research. *Adolescence*, *44*(173), 1-19.

Arbreton, A. J. A., Sheldon, J., & Herrera, C. (2005). *Beyond safe havens: A synthesis of 20 years of research on the Boys & Girls Clubs.* Philadelphia, PA: Public/Private Ventures.

Asher, R. (2012). *Voices from the field: Staffing Out-of-School Time Programs in the 21st Century*. Dover, DE: National Institute of Out-of-School Time.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1*(1), 91-97. Retrieved from http://doi.org/10.1901/jaba.1968.1-91

Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *20*, 313-327. Retrieved from http://doi.org/10.1901/jaba.1968.1-91

Baker, J. M., & Zigmond, N. (1990). Are regular education classes equipped to accommodate students with learning disabilities? *Exceptional Children*, *56*, 515-526.

Balch, R. T. (2012). *The validation of a student survey on teacher practice* (Doctoral Dissertation). Retrieved from https://etd.library.vanderbilt.edu

Barton, E. E., & Wolery, M. (2007). Evaluation of e-mail feedback on the verbal behaviors of preservice teachers. *Journal of Early Intervention*, *30*(1), 55-72. Retrieved from http://doi.org/10.1177/105381510703000105

Beaman, R., & Wheldall, K. (2000). Teachers' use of approval and disapproval in the classroom. *Educational Psychology*, *20*, 431-446. Retrieved from http://doi.org/10.1080/713663753

Becker, W. C., Madsen, C. H., Arnold, C. R., & Thomas, D. R. (1967). The contingent use of teacher attention and praise in reducing classroom behavior problems. *The Journal of Special Education*, *1*(3), 287-307. Retrieved from http://doi.org/10.1177/002246696700100307

Beckett, M., Borman, G., Capizzano, J., Parsley, D., Ross, S., Schirm, A., & Taylor, J. (2009). *Structuring out-of-school time to improve academic achievement: A practice guide* (No. NCEE #2009-012). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Belfiore, P. J., & Browder, D. M. (1992). The effects of self-monitoring on teacher's data-based decisions and on the progress of adults with severe mental retardation. *Education & Training in Mental Retardation*, *27*(1), 60-67.

Bell, L. M., & Aldridge, J. M. (2014). Investigating the use of student perception data for teacher reflection and classroom improvement. *Learning Environments Research*, *17*(3), 371-388. Retrieved from http://doi.org/10.1007/s10984-014-9164-z

Briere, D. E., Simonsen, B., Sugai, G., & Myers, D. (2015). Increasing new teachers' specific praise using a within-school consultation intervention. *Journal of Positive Behavior Interventions*, *17*(1), 50-60. Retrieved from http://doi.org/10.1177/1098300713497098

Brophy, J. (1981). Teacher praise: A functional analysis. *Review of Educational Research*, *51*(1), 5-32. Retrieved from http://doi.org/10.3102/00346543051001005

Browder, D. M., Liberty, K., Heller, M., & D'Huyvetters, K. K. (1986). Self-management by teachers: Improving instructional decision making. *Professional School Psychology*, *1*(3), 165-175. Retrieved from http://doi.org/10.1037/h0090506

Burgio, L. D., Engel, B. T., Hawkins, A., McCormick, K., Scheve, A., & Jones, L. T. (1990). A Staff management system for maintaining improvements in continence with elderly nursing home residents. *Journal of Applied Behavior Analysis*, *23*(1), 111-118. Retrieved from http://doi.org/10.1901/jaba.1990.23-111

Burnett, P. C. (2002). Teacher praise and feedback and students' perceptions of the classroom environment. *Educational Psychology*, *22*(1), 5-16. Retrieved from http://doi.org/10.1080/01443410120101215

Burnett, P. C., & Mandel, V. (2010). Praise and feedback in the primary classroom: Teachers' and students' perspectives. *Australian Journal of Educational & Developmental Psychology*, *10*, 145-154.

Cameron, J. (2002). *Rewards and intrinsic motivation: Resolving the controversy*. Westport, CT: Bergin & Garvey.

Cameron, J., & Pierce, W. D. (1996). The debate about rewards and intrinsic motivation: Protests and accusations do not alter the results. *Review of Educational Research*, *66*(1), 39-51. Retrieved from http://doi.org/10.3102/00346543066001039

Capizzi, A. M., Wehby, J. H., & Sandmel, K. N. (2010). Enhancing mentoring of teacher candidates through consultative feedback and self-evaluation of instructional delivery. *Teacher Education & Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, *33*(3), 191-212. Retrieved from http://doi.org/10.1177/0888406409360012

Carver, P. R., & Iruka, I. U. (2006). *After-school programs and activities* (No. NCES 2006-076). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Cavanaugh, B. (2013). Performance feedback and teachers' use of praise and opportunities to respond: A review of the literature. *Education & Treatment of Children*, *36*(1), 111-136. Retrieved from http://doi.org/10.1353/etc.2013.0001

Cherne, J. L. (2008). *Effects of praise on student behavior in the classroom* (Doctoral dissertation). Retrieved from https://www.lib.umn.edu/copyright/disstheses

Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014). *What makes great teaching? Review of the underpinning research*. London, England: The Sutton Trust.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98-101.

Cook, W. W., & Leeds, C. H. (1947). Measuring the teaching personality. *Educational and Psychological Measurement*, *7*, 399-410. Retrieved from http://doi.org/10.1177/001316444700700304

Cooper, J. A., Heron, T. E., & Heward, W. L. (2007). *Applied Behavior Analysis* (2nd ed.). Upper Saddle River, NJ: Pearson.

Cornelius, K. E., & Nagro, S. A. (2014). Evaluating the evidence base of performance feedback in preservice special education teacher training. *Teacher Education & Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, *37*(2), 133-146. Retrieved from http://doi.org/10.1177/0888406414521837

Cossairt, A., Hall, R. V., & Hopkins, B. L. (1973). The effects of experimenter's instructions, feedback, and praise on teacher praise and student attending behavior. *Journal of Applied Behavior Analysis*, *6*(1), 89-100. Retrieved from http://doi.org/10.1901/jaba.1973.6-89

Craft, M. A., Alber, S. R., & Heward, W. L. (1998). Teaching elementary students with developmental disabilities to recruit teacher attention in a general education classroom: Effects on teacher praise and academic productivity. *Journal of Applied Behavior Analysis*, *31*, 399-415. Retrieved from http://doi.org/10.1901/jaba.1998.31-399

Cross, A. B., Gottfredson, D. C., Wilson, D. M., Rorie, M., & Connell, N. (2010). Implementation quality and positive experiences in after-school programs. *American Journal of Community Psychology*, *45*, 370-380. Retrieved from http://doi.org/10.1007/s10464-010-9295-z

De Jong, R., & Westerhof, K. J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research*, *4*(1), 51-85. Retrieved from http://doi.org/http://dx.doi.org/10.1023/A:1011402608575

Deno, S., Maruyama, G., Espin, C., & Cohen, C. (1990). Educating students with mild disabilities in general education classrooms: Minnesota alternatives. *Exceptional Children*, *57*(2), 150-161.

Downer, J. T., Stuhlman, M., Schweig, J., Martínez, J. F., & Ruzek, E. (2015). Measuring effective teacher-student interactions from a student perspective: A multi-level analysis. *The Journal of Early Adolescence*, *35*, 722-758. Retrieved from http://doi.org/10.1177/0272431614564059

Duchaine, E. L., Jolivette, K., & Fredrick, L. D. (2011). The effect of teacher coaching with performance feedback on behavior-specific praise in inclusion classrooms. Education and Treatment of Children, 34(2), 209–227. Retrieved from http://doi.org/10.1353/etc.2011.0009

Dufrene, B. A., Lestremau, L., & Zoder-Martell, K. (2014). Direct behavioral consultation: Effects on teachers' praise and student disruptive behavior. Psychology in the Schools, 51(6), 567–580. Retrieved from http://doi.org/10.1002/pits.21768

Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology*, *45*, 294-309. Retrieved from http://doi.org/10.1007/s10464-010-9300-6

Dweck, C. S. (2002). Messages that motivate: How praise molds students' beliefs, motivation, and performance (in surprising ways). In J. Aronson (Ed.), *Improving academic achievement* (pp. 37-60). San Diego, CA: Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780120644551500063

Evertson, C. M., Anderson, C. W., Anderson, L. M., & Brophy, J. E. (1980). Relationships between classroom behaviors and student outcomes in junior high mathematics and English classes. *American Educational Research Journal*, *17*(1), 43-60. Retrieved from http://doi.org/10.3102/00028312017001043

Fallon, L. M., Collier-Meek, M. A., Maggin, D. M., Sanetti, L. M. H., & Johnson, A. H. (2015). Is performance feedback for educators an evidence-based practice? A systematic review and evaluation based on single-case research. *Exceptional Children*, *81*, 227-246. Retrieved from http://doi.org/10.1177/0014402914551738

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning & Instruction*, *29*, 1-9. Retrieved from http://doi.org/10.1016/j.learninstruc.2013.07.001

Ferguson, R. F. (2012). Can student surveys measure teaching quality? *Phi Delta Kappan*, *94*(3), 24-28.

Fisher, W. W., Ninness, H. A. C., Piazza, C. C., & Owen-DeSchryver, J. S. (1996). On the reinforcing effects of the content of verbal attention. *Journal of Applied Behavior Analysis*, *29*(2), 235-238. Retrieved from http://doi.org/10.1901/jaba.1996.29-235

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (No. 231 FMHI Publication). Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network. Retrieved from http://nirn.fmhi.usf.edu

Follman, J. (1992). Secondary school students' ratings of teacher effectiveness. *The High School Journal*, *75*(3), 168-178.

Follman, J. (1995). Elementary public school pupil rating of teacher effectiveness. *Child Study Journal*, *25*(1), 57.

Fraser, B. J., & Walberg, H. J. (1981). Psychosocial learning environment in science classrooms: A review of research. *Studies in Science Education*, *8*(1), 67-92. Retrieved from http://doi.org/10.1080/03057268108559887

Fullerton, E. K., Conroy, M. A., & Correa, V. I. (2009). Early childhood teachers' use of specific praise statements with young children at risk for behavioral disorders. *Behavioral Disorders*, 34(3), 118–135.

Gaertner, H. (2014). Effects of student feedback as a method of self-evaluating the quality of teaching. *Studies in Educational Evaluation*, 42, 91–99. Retrieved from http://doi.org/10.1016/j.stueduc.2014.04.003

Good, T. L., & Grouws, D. A. (1977). Teaching effects: Process-product relationships in fourth grade mathematics classrooms. *Journal of Teacher Education*, *28*(3), 49-54. Retrieved from http://doi.org/10.1177/002248717702800310

Gray, F., Graubard, P. S., & Rosenberg, H. (1974, March). Little brother is changing you. *Psychology Today*, 7(10), 42-46.

Granger, R. C. (2008). After-school programs and academics: Implications for policy, practice, and research. *Social Policy Report*, 22(2), 3–20.

Greene, K., Lee, B., Constance, N., & Hynes, K. (2013). Examining youth and program predictors of engagement in out-of-school time programs. *Journal of Youth & Adolescence*, *42*, 1557-1572. Retrieved from http://doi.org/10.1007/s10964-012-9814-3

Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, *52*, 1182-1186. Retrieved from http://doi.org/10.1037/0003-066X.52.11.1182

Grossman, J. B., Price, M. L., Fellerath, V., Jucovy, L. Z., Kotloff, L. J., Raley, R., & Walker, K. E. (2002). *Multiple choices after school: Findings from the Extended-Service Schools Initiative.* Philadelphia, PA: Public/Private Ventures. Retrieved from http://eric.ed.gov/?id=ED468056

Halpern, R. (2002). A different kind of child development institution: The history of after-school programs for low-income children. *Teachers College Record*, *104*(2), 178-211.

Hartje, J. A., Evans, W. P., Killian, E. S., & Brown, R. (2007). Youth worker characteristics and self-reported competency as predictors of intent to continue working with youth. *Child & Youth Care Forum*, *37*(1), 27-41. Retrieved from http://doi.org/10.1007/s10566-007-9048-9

Hasbrouck, J. E. (1997). Mediated peer coaching for training preservice teachers. *The Journal of Special Education*, *31*, 251-271. Retrieved from http://doi.org/10.1177/002246699703100206

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81-112. Retrieved from http://doi.org/10.3102/003465430298487

Hawkins, S. M., & Heflin, L. J. (2011). Increasing secondary teachers' behavior-specific praise using a video self-modeling and visual performance feedback intervention. *Journal of Positive Behavior Interventions*, 13(2), 97–108. Retrieved from http://doi.org/10.1177/1098300709358110

Hemmeter, M. L., Snyder, P., Kinder, K., & Artman, K. (2011). Impact of performance feedback delivered via electronic mail on preschool teachers' use of descriptive praise. *Early Childhood Research Quarterly*, 26(1), 96–109. Retrieved from http://doi.org/10.1016/j.ecresq.2010.05.004

Hirsch, B. J., Mekinda, M. A., & Stawicki, J. (2010). More than attendance: The importance of after-school program quality. *American Journal of Community Psychology*, *45*, 447-452. Retrieved from http://doi.org/10.1007/s10464-010-9310-4

Hoff, K. E., & Sawka-Miller, K. D. (2010). Self-management interventions. In G. G. Peacock, R. A. Ervin, E. J. Daly, III, & K. W. Merrell (Eds.), *Practical handbook of school psychology: Effective practices for the 21st Century* (pp. 337-352). New York, NY: Guilford.

Horner, R. H., Blitz, C., & Ross, S. W. (2014). *The importance of contextual fit when implementing evidence-based interventions*. Washington, DC: U.S. Department of Health and Human Services. Retrieved from https://aspe.hhs.gov/sites/default/files/pdf/77066/ib_Contextual.pdf

Horner, R. H., Carr, E., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*(2), 165-179. Retrieved from http://doi.org/10.1177/001440290507100203

James-Burdumy, S., Dynarski, M., & Deke, J. (2007). When elementary schools stay open late: Results from the national evaluation of the 21st Century Community Learning Centers program. *Educational Evaluation & Policy Analysis*, *29*(4), 296-318. Retrieved from http://doi.org/10.3102/0162373707309077

Jenkins, L. N., Floress, M. T., & Reinke, W. (2015). Rates and types of teacher praise: A review and future directions. *Psychology in the Schools*, *52*, 463-476. Retrieved from http://doi.org/10.1002/pits.21835

Joseph, L. M., Konrad, M., Cates, G., Vajcner, T., Eveleigh, E., & Fishley, K. M. (2012). A meta-analytic review of the cover-copy-compare and variations of this self-management procedure. *Psychology in the Schools*, *49*(2), 122-136. Retrieved from http://doi.org/10.1002/pits.20622

Joyce, B. R., & Showers, B. (1980). Improving inservice training: The message of research. *Educational Leadership*, *37*(5), 379-385.

Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Kachigan, S. K. (1986). *Statistical analysis: An interdisciplinary introduction to univariate and multivariate methods*. New York, NY: Radius.

Kalis, T. M., Vannest, K. J., & Parker, R. (2007). Praise counts: Using self-monitoring to increase effective teaching practices. *Preventing School Failure*, *51*(3), 20-27.

Keller, C. L., Brady, M. P., & Taylor, R. L. (2005). Using self-evaluation to improve student teacher interns' use of specific praise. *Education & Training in Developmental Disabilities*, *40*, 368-376.

Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., … Dunlop Velez, E. (2015). *The condition of education 2015* (No. NCES 2015-144). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Kilbourn, B. (1991). Self-monitoring in teaching. *American Educational Research Journal*, *28*(4), 721-736.

Kohn, A. (1999). *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes.* New York, NY: Houghton Mifflin Harcourt.

Kremer, K. P., Maynard, B. R., Polanin, J. R., Vaughn, M. G., & Sarteschi, C. M. (2014). Effects of after-school programs with at-risk youth on attendance and externalizing behaviors: A systematic review and meta-analysis. *Journal of Youth and Adolescence*, *44*, 616-636. Retrieved from http://doi.org/10.1007/s10964-014-0226-4

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education & Special Education*, *33*, 279-299. Retrieved from http://doi.org/10.1177/0888406410371643

Kunter, M., & Baumert, J. (2007). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9(3), 231–251. Retrieved from http://doi.org/10.1007/s10984-006-9015-7

Larson, R. W., & Walker, K. C. (2010). Dilemmas of practice: Challenges to program quality encountered by youth program leaders. *American Journal of Community Psychology*, *45*, 338-349. Retrieved from http://doi.org/10.1007/s10464-010-9307-z

Latham, G. (1997). *Behind the schoolhouse door: Eight skills every teacher should have*. Logan, UT: Utah State University. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED408735

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, *76*, 275-313. Retrieved from http://doi.org/10.3102/00346543076002275

Lauver, S., Little, P. M. D., & Weiss, H. B. (2004). *Moving beyond the barriers: Attracting and sustaining youth participation in Out-of-School Time programs*. Boston, MA: Harvard University. Retrieved from http://www.hfrp.org/publications-resources/publications-series/issues-and-opportunities-in-out-of-school-time-evaluation/moving-beyond-the-barriers-attracting-and-sustaining-youth-participation-in-out-of-school-time-programs

Lepper, M. R., Keavney, M., & Drake, M. (1996). Intrinsic motivation and extrinsic rewards: A commentary on Cameron and Pierce's meta-analysis. *Review of Educational Research*, *66*(1), 5-32. Retrieved from http://doi.org/10.3102/00346543066001005

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Little, P. M. D., Wimer, C., & Weiss, H. B. (2007). *After school programs in the 21st Century: Their potential and what it takes to achieve it*. Cambridge, MA: Harvard Family Research Project.

Long, J. D., & Williams, R. L. (1977). Self-management for teachers. *Education*, *97*(3), 242.

Lu, H.-L. (2010). Research on peer coaching in preservice teacher education: A review of literature. *Teaching & Teacher Education*, *26*, 748-753. Retrieved from http://doi.org/10.1016/j.tate.2009.10.015

Lylo, B. J., & Lee, D. L. (2013). Effects of delayed audio-based self-monitoring on teacher completion of learning trials. *Journal of Behavioral Education*, *22*(2), 120-138.

Mace, F. C., Cancelli, A. A., & Manos, M. J. (1983). Increasing teacher delivery of contingent praise and contingent materials using consultant feedback and praise. *School Psychology Review*, *12*, 340-346.

MacLean, W. E., Tapp, J. T., & Johnson, W. L. (1985). Alternate methods and software for calculating interobserver agreement for continuous observation data. *Journal of Psychopathology & Behavioral Assessment*, *7*(1), 65-73. Retrieved from http://doi.org/10.1007/BF00961847

Madsen, C. H., Becker, W. C., & Thomas, D. R. (1968). Rules, praise, and ignoring: elements of elementary classroom control. *Journal of Applied Behavior Analysis*, *1*(2), 139-150. Retrieved from http://doi.org/10.1901/jaba.1968.1-139

Maggin, D. M., Briesch, A. M., & Chafouleas, S. M. (2013). An application of the What Works Clearinghouse standards for evaluating single-subject research synthesis of the self-management literature base. *Remedial & Special Education*, *34*(1), 44-58. Retrieved from http://doi.org/10.1177/0741932511435176

Marchand-Martella, N. E., Lignugaris-Kraft, B., Pettigrew, T., Leishman, R., & Ross, S. W. (2010). *Direct Instruction Supervision System* (2nd ed.). Logan, UT: Utah State University.

Martens, B. K., Witt, J. C., Elliott, S. N., & Darveaux, D. X. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology: Research & Practice*, *16*(2), 191-198. Retrieved from http://doi.org/http://dx.doi.org.dist.lib.usu.edu/10.1037/0735-7028.16.2.191

Marzano, R. J. (2003). *What works in schools: Translating research into action*? Alexandria, VA: Association for Supervision and Curriculum Development.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation & Policy Analysis*, *21*(1), 29-45. Retrieved from http://doi.org/10.3102/01623737021001029

Maynard, B. R., Peters, K. E., Vaughn, M. G., & Sarteschi, C. M. (2013). Fidelity in after-school program intervention research A systematic review. *Research on Social Work Practice*, *23*, 613-623. Retrieved from http://doi.org/10.1177/1049731513491150

Measures of Effective Teaching. (2012a). *Asking students about teaching: Student perception surveys and their implementation*. Seattle, WA: Bill and Melinda Gates Foundation.

Measures of Effective Teaching. (2012b). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill and Melinda Gates Foundation.

Merrett, F., & Wheldall, K. (1987). Natural rates of teachers' approval and disapproval in British primary and middle school classrooms. *British Journal of Educational Psychology, 57,* 95-103.

Mesa, J., Lewis-Palmer, T., & Reinke, W. (2005). Providing teachers with performance feedback on praise to reduce student problem behavior. *Beyond Behavior*, *15*(1), 3-7.

Miller, K., & Snow, D. (2004). *Out of school time programs for at-risk students*. Aurora, CO: Mid-continent Research for Education and Learning. Retrieved from http://eric.ed.gov/?id=ED484550

Moffat, T. K. (2011). Increasing the teacher rate of behaviour specific praise and its effect on a child with aggressive behaviour problems. *Kairaranga*, *12*(1), 51-58.

Morgan, R. L., Menlove, R., Salzberg, C. L., & Hudson, P. (1994). Effects of peer coaching on the acquisition of direct instruction skills by low-performing preservice teachers. *The Journal of Special Education*, *28*(1), 59-76. Retrieved from http://doi.org/10.1177/002246699402800105

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality & Social Psychology*, *75*(1), 33-52. Retrieved from http://doi.org/10.1037/0022-3514.75.1.33

Myers, D. M., Simonsen, B., & Sugai, G. (2011). Increasing teachers' use of praise with a response-to-intervention approach. *Education & Treatment of Children*, *34*(1), 35-59.

Nafpaktitis, M., Mayer, G. R., & Butterworth, T. (1985). Natural rates of teacher approval and disapproval and their relation to student behavior in intermediate school classrooms. *Journal of Educational Psychology*, *77*, 362-367. Retrieved from http://doi.org/10.1037/0022-0663.77.3.362

Nee, J., Howe, P., Schmidt, C., & Cole, P. (2006). *Understanding the afterschool workforce: Opportunities and challenges for an emerging profession* Houston, TX: Cornerstones for Kids.

Oliver, R. M., Wehby, J. H., & Nelson, J. R. (2015). Helping teachers maintain classroom management practices using a self-monitoring checklist. *Teaching & Teacher Education*, *51*, 113-120. Retrieved from http://doi.org/10.1016/j.tate.2015.06.007

Parsonson, B. S., & Baer, D. M. (1978). The analysis and presentation of graphic data. In T. R. Kratochwill (Ed.), *Single subject research: Strategies for evaluating change* (1st ed., pp. 101-166). New York, NY: Academic.

Peterson, L. D., Young, K. R., Salzberg, C. L., West, R. P., & Hill, M. (2006). Using self-management procedures to improve classroom social skills in multiple general education settings. *Education & Treatment of Children*, *29*(1), 1-21.

Petscher, E. S., & Bailey, J. S. (2006). Effects of training, prompting, and self-monitoring on staff behavior in a classroom for students with disabilities. *Journal of Applied Behavior Analysis*, *39*(2), 215-226.

Phillips, R. H. (1984). Increasing positive self-referent statements to improve self-esteem in low-income elementary school children. *Journal of School Psychology*, *22*(2), 155-163. Retrieved from http://doi.org/10.1016/0022-4405(84)90035-9

Pierce, T., & Miller, S. P. (1994). Using peer coaching in preservice practica. *Teacher Education & Special Education*, *17*(4), 215-223. Retrieved from http://doi.org/10.1177/088840649401700402

Plavnick, J. B., Ferreri, S. J., & Maupin, A. N. (2010). The effects of self-monitoring on the procedural integrity of a behavioral intervention for young children with developmental disabilities. *Journal of Applied Behavior Analysis*, *43*, 315-320.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, *31*(7), 3-14. Retrieved from http://doi.org/10.3102/0013189X031007003

Praise. (2016). *Merriam-Webster.com*. Retrieved from http://www.meriamwebster.com/dictionary/praise

Rathel, J. M., Drasgow, E., Brown, W. H., & Marshall, K. J. (2014). Increasing induction-level teachers' positive-to-negative communication ratio and use of behavior-specific praise through e-mailed performance feedback and its effect on students' task engagement. *Journal of Positive Behavior Interventions*, *16*, 219-233. Retrieved from http://doi.org/10.1177/1098300713492856

Rathel, J. M., Drasgow, E., & Christle, C. C. (2008). Effects of supervisor performance feedback on increasing preservice teachers' positive communication behaviors with students with emotional and behavioral disorders. *Journal of Emotional & Behavioral Disorders*, *16*(2), 67-77. Retrieved from http://doi.org/10.1177/1063426607312537

Reinke, W. M., Lewis-Palmer, T., & Martin, E. (2007). The effect of visual performance feedback on teacher use of behavior-specific praise. *Behavior Modification*, *31*(3), 247-263. Retrieved from http://doi.org/10.1177/0145445506288967

Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The Classroom Check-Up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review*, *37*(3), 315-332.

Riley-Tillman, T. C., & Eckert, T. L. (2001). Generalization programming and school-based consultation: An examination of consultees' generalization of consultation-related skills. *Journal of Educational & Psychological Consultation*, *12*(3), 217-241. Retrieved from http://doi.org/10.1207/S1532768XJEPC1203_03

Roth, J. L., Malone, L. M., & Brooks-Gunn, J. (2010). Does the amount of participation in afterschool programs relate to developmental outcomes? A review of the literature. *American Journal of Community Psychology*, *45*, 310-324. Retrieved from http://doi.org/10.1007/s10464-010-9303-3

Ryan, R. M., & Deci, E. L. (1996). When paradigms clash: Comments on Cameron and Pierce's claim that rewards do not undermine intrinsic motivation. *Review of Educational Research*, *66*(1), 33-38. Retrieved from http://doi.org/10.3102/00346543066001033

Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11*(1), 57-67. Retrieved from http://doi.org/10.1023/A:1007999204543

Scheeler, M. C., Ruhl, K. L., & McAfee, J. K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education, 27*, 396-407. Retrieved from http://doi.org/10.1177/088840640402700407

Schwartz, I. S., & Baer, D. M. (1991). Social validity assessments: Is current practice state of the art? *Journal of Applied Behavior Analysis*, *24*, 189-204. Retrieved from http://doi.org/10.1901/jaba.1991.24-189

Scott-Little, C., Hamann, M. S., & Jurs, S. G. (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, *23*, 387-419. Retrieved from http://doi.org/10.1177/109821400202300403

Serbin, L. A., Tonick, I. J., & Sternglanz, S. H. (1977). Shaping cooperative cross-sex play. *Child Development*, *48*, 924-929. Retrieved from http://doi.org/10.1111/1467-8624.ep10402886

Shapiro, E. S., & Cole, C. L. (1994). *Behavior change in the classroom: Self-management interventions*. New York, NY: Guilford.

Sheffield, K., & Waller, R. J. (2010). A review of single-case studies utilizing self-monitoring interventions to reduce problem classroom behaviors. *Beyond Behavior*, *19*(2), 7-13.

Shores, R. E., Gunter, P. L., & Jack, S. L. (1993). Classroom management strategies: Are they setting events for coercion? *Behavioral Disorders*, *18*(2), 92-102.

Showers, B., Joyce, B. R., & Bennett, B. (1987). Synthesis of research on staff development: A framework for future study and a state-of-the-art analysis. *Educational Leadership*, *45*(3), 77-87.

Silvestri, S. M. (2004). *Effects of self-scoring on teachers' rates of positive and negative statements during classroom instruction* (Doctoral dissertation). Retrieved from https://etd.ohiolink.edu/

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, *31*, 351-380. Retrieved from http://doi.org/10.1353/etc.0.0007

Simonsen, B., MacSuga, A. S., Fallon, L. M., & Sugai, G. (2013). The effects of self-monitoring on teachers' use of specific praise. *Journal of Positive Behavior Interventions*, *15*(1), 5-15. Retrieved from http://doi.org/10.1177/1098300712440453

Simonsen, B., MacSuga-Gage, A. S., Briere, D. E., Freeman, J., Myers, D., Scott, T. M., & Sugai, G. (2014). Multitiered support framework for teachers' classroom-management practices overview and case study of building the triangle for teachers. *Journal of Positive Behavior Interventions*, *16*(3), 179-190. Retrieved from http://doi.org/10.1177/1098300713484062

Simonsen, B., Myers, D., & DeLuca, C. (2010). Teaching teachers to use prompts, opportunities to respond, and specific praise. *Teacher Education & Special Education*, *33*(4), 300-318. Retrieved from http://doi.org/10.1177/0888406409359905

Sloat, K. C. M., Tharp, R. G., & Gallimore, R. (1977). The incremental effectiveness of classroom-based teacher-training techniques. *Behavior Therapy*, *8*, 810-818. Retrieved from http://doi.org/10.1016/S0005-7894(77)80152-4

Smith, D. J., Young, K. R., West, R. P., Morgan, D. P., & Rhode, G. (1988). Reducing the disruptive behavior of junior high school students: A classroom self-management procedure. *Behavioral Disorders*, *13*, 231-239.

Sprick, R. S., Knight, J., Reinke, W., & McKale, T. (2008). *Coaching: Approaches and perspectives*. Thousand Oaks, CA: Corwin.

State, T. & Harrison, J. (2016, March). Social acceptability: Measurement and reporting in intervention research for students with EBD. Paper presented at the annual conference of the Association of Positive Behavior Support, San Francisco, CA.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, *10*, 349-367. Retrieved from http://doi.org/10.1901/jaba.1977.10-349

Stokes, T. F., Fowler, S. A., & Baer, D. M. (1978). Training preschool children to recruit natural communities of reinforcement. *Journal of Applied Behavior Analysis*, *11*(2), 285-303. Retrieved from http://doi.org/10.1901/jaba.1978.11-285

Stormont, M. A., Smith, S. C., & Lewis, T. J. (2007). Teacher implementation of precorrection and praise statements in Head Start classrooms as a component of a program-wide system of positive behavior support. Journal of Behavioral Education, 16(3), 280–290. http://doi.org/10.1007/s10864-007-9040-3

Strain, P. S., Lambert, D. L., Kerr, M. M., Stagg, V., & Lenkner, D. A. (1983). Naturalistic assessment of children's compliance to teachers' requests and consequences for compliance. *Journal of Applied Behavior Analysis*, *16*, 243-249. Retrieved from http://doi.org/10.1901/jaba.1983.16-243

Sutherland, K. S. (2000). Promoting positive interactions between teachers and students with emotional/behavioral disorders. *Preventing School Failure*, *44*(3), 110-115.

Sutherland, K. S., Alder, N., & Gunter, P. L. (2003). The effect of varying rates of opportunities to respond to academic requests on the classroom behavior of students with EBD. *Journal of Emotional & Behavioral Disorders*, *11*(4), 239-248. Retrieved from http://doi.org/10.1177/10634266030110040501

Sutherland, K. S., Copeland, S. R., & Wehby, J. H. (2001). Catch them while you can: Monitoring and increasing the use of effective praise. Beyond Behavior, 11, 46–49.

Sutherland, K. S., & Wehby, J. H. (2001). The effect of self-evaluation on teaching behavior in classrooms for students with emotional and behavioral disorders. *The Journal of Special Education*, *35*(3), 161-171. Retrieved from http://doi.org/10.1177/ 002246690103500306

Sutherland, K. S., Wehby, J. H., & Copeland, S. R. (2000). Effect of varying rates of behavior-specific praise on the on-task behavior of students with EBD. *Journal of Emotional & Behavioral Disorders*, *8*(1), 2-8. Retrieved from http://doi.org/10.1177/106342660000800101

Taheri, S. A., & Welsh, B. C. (2015). After-school programs for delinquency prevention A systematic review and meta-analysis. *Youth Violence & Juvenile Justice*. Advance online publication. Retrieved from http://doi.org/10.1177/1541204014567542

Taylor, J. J. (20133). *Inclusion of youth with disabilities in afterschool programs*. Oregon State University, Corvallis, OR. Retrieved from http://ir.library.oregonstate.edu/xmlui/handle/1957/42773

Tschecthelin, M. A., Hipskind, M. J. F., & Remmers, H. H. (1940). Measuring the attitudes of elementary-school children toward their teachers. *Journal of Educational Psychology*, *31*(3), 195-203. Retrieved from http://doi.org/10.1037/h0055853

Tuckman, B. W., & Oliver, W. F. (1968). Effectiveness of feedback to teachers as a function of source. *Journal of Educational Psychology*, *59*, 297-301. Retrieved from http://doi.org/10.1037/h0026022

U.S. Department of Education. (2015, May 15). *21st Century Community Learning Centers: Funding status*. Retrieved from http://www2.ed.gov/programs/21stcclc/funding.html

Van Houten, R., & Sullivan, K. (1975). Effects of an audio cueing system on the rate of teacher praise. *Journal of Applied Behavior Analysis*, *8*(2), 197-201. Retrieved from http://doi.org/10.1901/jaba.1975.8-197

Vollmer, T. R., & Hackenberg, T. D. (2001). Reinforcement contingencies and social reinforcement: Some reciprocal relations between basic and applied research. *Journal of Applied Behavior Analysis*, *34*, 241-253. Retrieved from http://doi.org/10.1901/jaba.2001.34-241

West, R. P. (2005). Self-instruction. In G. Sugai & R. H. Horner (Eds.), *Encyclopedia of behavior modification and cognitive behavior therapy* (Vol. 3, pp. 1511-1513). Thousand Oaks, CA: Sage.

West, R. P., Smith, T. G., & Taylor, M. J. (2004). Principles for effective schools. *The Utah Special Educator*, *25*, 10-11.

Wheatley, R. K. (2015, May). *The effectiveness of an intervention designed to increase the positive to negative ratio of instructor interactions during afterschool programming* (Master's Thesis). Retrieved from http://digitalcommons.usu.edu/etd/

Wheldall, K., Houghton, S., & Merrett, F. (1989). Natural rates of teacher approval and disapproval in British secondary school classrooms. *British Journal of Educational Psychology, 59,* 38-48.

White, M. A. (1975). Natural rates of teacher approval and disapproval in the classroom. *Journal of Applied Behavior Analysis*, *8*, 367-372. Retrieved from http://doi.org/10.1901/jaba.1975.8-367

Wilkerson, D. J., Manatt, R. P., Rogers, M. A., & Maughan, R. (2000). Validation of student, principal, and self-ratings in 360° Feedback® for teacher evaluation. *Journal of Personnel Evaluation in Education*, *14*(2), 179-192. Retrieved from http://doi.org/10.1023/A:1008158904681

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, *11*(2), 203-214. Retrieved from http://doi.org/10.1901/jaba.1978.11-203

Woodland, M. H. (2008). Whatcha doin' after school? A review of the literature on the influence of after-school programs on young black males. *Urban Education*, *43*, 537-560. Retrieved from http://doi.org/10.1177/0042085907311808

Wyatt, W. J., & Hawkins, R. P. (1987). Rates of teachers' verbal approval and disapproval relationship to grade level, classroom activity, student behavior, and teacher characteristics. *Behavior Modification*, 11(1), 27–51. Retrieved from http://doi.org/10.1177/01454455870111003

Young, K. R., West, R. P., Smith, D. J., & Morgan, D. P. (1991). *Teaching Self-Management Strategies to Adolescents*. Longmont, CO: Sopris West.

Zhao, J., & Gallant, D. J. (2012). Student evaluation of instruction in higher education: Exploring issues of validity and reliability. *Assessment & Evaluation in Higher Education*, *37*, 227-235. Retrieved from http://doi.org/10.1080/02602938.2010.523819

Zimmerman, E. H., & Zimmerman, J. (1962). The alteration of behavior in a special classroom situation. *Journal of the Experimental Analysis of Behavior*, *5*(1), 59-60. Retrieved from http://doi.org/10.1901/jeab.1962.5-59

APPENDICES

Appendix A

Informed Consent Documents

Informed Consent Letter for Teachers

INFORMED CONSENT

Effects Of A Self-Management Procedure Using Student Feedback On Staff Members' Use Of Effective Instructional Practices In An Out-Of-School Time Program

Dr. Richard P. West in the Department of Special Education and Rehabilitation at Utah State University is conducting a research study to find out more about the effects of a self-management procedure on teachers' use of effective instructional practices in out-of-school-time programs. You have been asked to take part because you are a teacher at the Boys and Girls Club of Northern Utah. There will be approximately four total participants in this research. This study is being conducted in partial fulfillment of the requirements for the degree of doctor of philosophy for Cade Charlton.

If you agree to be in this research study, you will be asked to participate in daily data collection and a self-management procedure. It is anticipated that the study will last approximately 10 weeks.

To participate in data collection, you will be asked to record your interactions with students using a small, digital audio recorder. These recordings will be collected during a selected session lasting no more than 30 minutes each day. You will be asked to upload the audio recordings to a secure cloud storage folder, which will be accessible only to the investigator, Cade Charlton, and the assigned data collector. Independent, off-site data collectors assigned to review these recordings will document your instructional interactions with students. The recorded audio will be destroyed immediately after the records have been completed and verified for accuracy.

To participate in the self-management procedure, you will be asked to attend a brief training, collect performance feedback from your students, and complete self-appraisals as part of a systematic research process. If necessary, you may be asked to meet briefly with the program coordinator to review the research project and your participation

Risks Participation in this study may involve some added risks or discomforts. These include additional time and effort required to complete assigned research activities and the potential loss of confidentiality. However, the research team will make every effort to mitigate these risks and to reduce the burden of participation. First, your name will not be associated with the recorded audio you collect. You will be assigned a participant number and the number and name will be recorded on a separate hard-copy document stored within a locked cabinet in the locked office of the co-researcher. In addition, the recordings will be destroyed after all data collection has been completed and all research data have been verified for accuracy. The transcribed, de-identified data collected from

the recordings will be kept for five years.

Benefits There are several potential direct benefits to you for participating. First, the purpose of this study is to enhance the quality of instruction your students receive. We hypothesize that the quality of your relationships and interactions with students will increase through participation in this study making you more effective and potentially enhancing your job satisfaction. Second, you will have opportunities to earn an opportunity to win a small monetary incentive for participating. In addition, your participation may produce important, indirect benefits to you. For example, your participation may lead to the identification of effective procedures that enhance the quality of the club experience for the students and teachers who may use these procedures.

Participation in research is entirely voluntary. You may refuse to participate or withdraw at any time without consequence or loss of benefits. You may also be withdrawn from this study without your consent by the investigator. The most common reason for dismissal is failure to follow the research procedures as outlined by the investigator.

Research records will be kept confidential, consistent with federal and state regulations. To protect your privacy, personal, identifiable information will be removed from study documents and replaced with a study identifier known only to the investigator and Cade Charlton. Information regarding you and your assigned study identifier will be stored in a locked file cabinet in a locked room at Utah State University. Research documents and records will be kept for a period of five years.

Cade Charlton has explained this research study to you and answered your questions. If you have other questions or research-related problems, you may reach Dr. Richard P. West at (435) 797-3243 or via email at rich.west@usu.edu.

The Institutional Review Board for the protection of human participants at Utah State University has approved this research study. If you have any questions or concerns about your rights or a research-related injury and would like to contact someone other than the research team, you may contact the IRB Administrator at (435) 797-0567 or email irb@usu.edu to obtain information or to offer input. You have been given two copies of this Informed Consent. Please sign both copies and keep one copy for your files.

“I certify that the research study has been explained to the individual, by me or my research staff, and that the individual understands the nature and purpose, the possible risks and benefits associated with taking part in this research study. Any questions that have been raised have been answered.”

Letter of Information for Students

INFORMED CONSENT

Effects Of A Self-Management Procedure Using Student Feedback On Staff Members' Use Of Effective Instructional Practices In An Out-Of-School Time Program

Introduction Dr. Richard P. West and Cade Charlton from the Department of Special Education and Rehabilitation at Utah State University are conducting a research study on the effects of a self-management program for staff members in the Boys and Girls Club summer program. Cade Charlton is a graduate researcher and will coordinate this study along with Tara Peterson, the Boys and Girls Club Director of Programming. Four staff members from the Logan, Garland, and Brigham City programs will be recruited to participate in this study. Since the focus of the study is on the behavior of staff members, no data will be collected on the activities of your child(ren). The study will last approximately 10 weeks.

Procedures Children will be automatically included in the study, but you our your child(ren) may opt out if you wish. If you allow your child(ren) to participate, she or he will be asked to complete a daily survey about the participating staff member (i.e., their teacher). They will be asked to answer two yes/no questions:

1. Did your teacher praise you today?
2. Did he/she say which behavior he/she was praising?

Your child's responses to the questions above will be collected anonymously and reported in aggregate to the teacher. This survey will be collected every day for a short period of the study and then used periodically (e.g., every other day or weekly) thereafter.

In addition, audio recordings of a 30-minute session will be collected throughout the study. These recordings will only be used to document the teacher's use of specific instructional strategies. Although your child(ren)'s voice(s) may be incidentally recorded, no data will be collected on the children's behavior.

Risks Participation in this research study may involve some added risks or discomforts. These include a potential loss of confidentiality and the possibility that your child(ren) will not enjoy completing the survey. If for any reason your child does not want to complete the survey, he/she will not be required to complete it.

Benefits The purpose of this study is to help your child's teacher increase the quality of his/her instruction. Your child(ren)'s participation(s) may lead to an enhanced learning experience at the summer program and improved teacher-student relationships.

Voluntary nature of participation and right to withdraw without consequence Participation in research is entirely voluntary. Also, your child may be asked not to participate if they fail to follow the research protocols as explained to them. You may ask

to have your child(ren) removed from the study at any time by contacting Tara Peterson at (435) 723-6224 or email tpeterson@bcbrclub.org or Cade Charlton (435) 797-7471 or email cade.charlton@usu.edu.

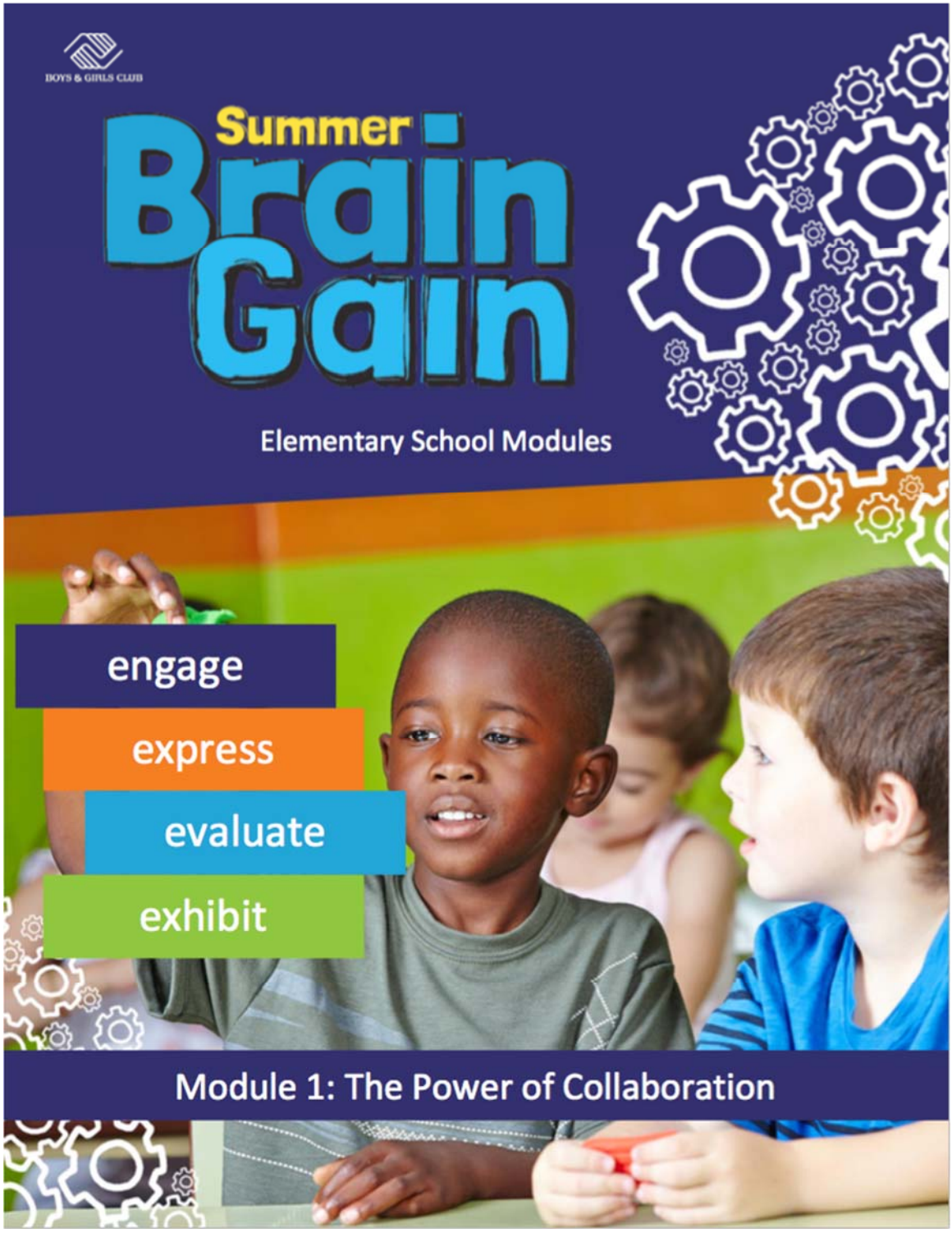
Confidentiality Research records will be kept confidential, consistent with federal and state regulations. Only the Dr. Richard West and Cade Charlton will have access to the data, which will be kept in a locked file cabinet or on a password protected computer in a locked room. To protect the privacy of your child(ren), audio recordings will be destroyed after the study is completed and the accuracy of data have been verified. All remaining de-identified research records will be kept for five years and then destroyed.

IRB Approval Statement The Institutional Review Board for the protection of human participants at Utah State University has approved this research study. If you have any questions or concerns about your rights or a research-related injury and would like to contact someone other than the research team, you may contact the IRB Administrator at (435) 797-0567 or email irb@usu.edu to obtain information or to offer input.

Appendix B

Summer Brain Gain Sample Lesson

Summer Brain Gain Sample Lesson



About This Module

Overview	This module will introduce members to the major components of the solar system including Earth. Through explorations of Earth's characteristics, the Sun, the workings of spacecraft and patterns of orbits, members will better understand how our solar system works. Activities will cover how we leave Earth and what differences we would encounter outside of our planet. Members will end the week with a better grasp of what is beyond Earth and what our relationship is to what's out there.
Driving Questions	How are the different parts of the solar system connected?
	What theories have yet to be confirmed about solar system?
Products of the Week	Model Earth (Members will complete a creative model of the Earth, displaying the major parts of Earth's surface and layers.)
	Film Canister Rocket Ship (Film Canister Rocket ships are small model rockets. Members will launch them to get an idea of what occurs during a space rocket launch. Members will brainstorm ideas about how rockets propel spacecraft through Earth's atmosphere into outer space.)
	Interactive Solar System Exhibit (Members will construct a model of the solar system that is life sized and interactive.)
Community Sharing Event	Solar System Exhibit At the end of the week, on Friday, members will hold a solar system exhibit, in which participants can view and interact with a life-sized model of the solar system. Members will participate as experts, explaining how the parts of the solar system function and are related. Club staff should be asked in advance to participate. The venue set up expo style mostly standing room with a few areas designated for presentations. Projectors and speaking podiums (areas) should be set up and available.

NASA Module 1: Our Solar System

Introduction	Did you know that where you live goes beyond, your city, state, and country? Your world is a part of a larger system called the solar system. The solar system consists of one star, the sun, 8 planets, including earth, and other celestial bodies. Earth is the 3rd planet from the sun; the only planet upon which we know life exists. Years of study have taught us many things about the solar system, including the fact that "Pluto" is not a planet, but rather a celestial snowball. As the study of the solar system extends into the galaxy and other parts of the universe, how can we continue to update what we know about the world beyond our world?
Special Notes	• This module is written to accommodate members of all elementary ages. It is important that the facilitator adjust activities if necessary, to better suit older or younger ages. Some suggestions for activity changes are included in this curriculum. Keep an eye out for any areas where activities can be adapted for the group. • The Solar System exhibit is the opportunity to share the work that members have done with the community. Feel free to add or swap out activities on that day at your discretion. Take into account the materials needed to pull off demonstrations more than once. Also take into account the extra materials necessary that may be needed to allow community members to participate in activities.

NASA Module 1: Our Solar System

Monday	**Activity 1:** Geocentrism vs. Heliocentrism
Location: Learning Center or Tech Center	**Estimated Time**: 30min
Description: In order for members to better understand life outside of Earth, they will understand that the sun is the center of the solar system by modeling our position and how we revolve around it.	
Objective: Members will model the position of the Sun, Earth, and Earth's moon in the solar system. Members will recognize that the solar system is heliocentric.	

What You Need		Handout: Moon Sign (per facilitator)
		Handout: Planet Sign (per facilitator)
		Handout: Sun sign (per facilitator)
How to Prepare		Review **Steps to Follow** (below).
		Set up the work area and pull out all materials.

Steps to Follow	1	**Say** "Geo" and **Allow** members to respond. **Say** "Geo" means earth. **Have** students repeat after you,"Geocentrism." **Allow** members to respond.
	2	**Say** Geocentrism is an idea that Earth was the center of the solar system and that objects in space traveled around Earth.
	3	**Say** "Helio." **Allow** members to respond. **Say** "Helio" means sun. **Have** students repeat after you, "Heliocentrism." **Allow** members to respond.
	4	**Say** Heliocentrism is the theory that the Sun is the center of the solar system and all objects are positioned around it. If this theory were to be true, our planet, Earth, would be one of the objects travelling around the sun. What do you think? Which theory is the true theory?
	5	**Organize** the members into a circle.
	6	**Choose** a volunteer to be the sun and have them hold the sun sign. While the sun remains standing still, **choose** another member and have them stand with their back to "the sun". This second member

		is a planet.
	7	**Tell** the Planet to take 5 big steps away from the sun and begin spinning around in place.
	8	**Say** "planet" while you are spinning around in place, slowly, begin to move around the sun in a circle. (*You may need to demonstrate)
	9	After the planet has made one rotation around the sun, **tell** the members to stop.
	10	**Have** them explain what that was like. The planet will probably be dizzy.
	11	**Explain** to members that now everyone will get a chance to participate.
	12	Organize the members into pairs and assign each pair a role (Planet or Sun).
	13	**Give** the members 1 minute to replicate the demonstration they just saw with their partner. When the minute is up, **organize** the members back into a large circle. Remind the members of the definition of geocentrism and heliocentrism.
	14	**Ask** did we just demonstrate heliocentrism or geocentrism? (heliocentrism)
	15	**Say** our solar system follows a heliocentric model. The Sun is the center and all of the planets in space revolve around it. Some planets like Earth, spin on their own axis as they revolve around the sun. That is what you guys were modeling.
	16	**Ask** How would you describe the shape of the path you travelled around the sun?
	17	**Say** Each planet has an elliptical orbit. This shape looks more like and oval than a circle. **Draw** an example of and ellipse for the members to observe. An ellipse looks like this: 

	18	**Have** your original volunteers resume their positions.
	19	**Tell** the "planet" to take eight steps away from the sun this time. This is just to create enough space for traveling, feel free to adjust the number of steps as needed.
	20	**Choose** an additional volunteer. This is the moon.
	21	**Say** "Planet" when I say "go" I want you to go back to spinning in place and around the "sun". "Moon", when I say go, walk around the planet as it spins and revolves around the sun.
	22	**Discuss** this demonstration, comparing it to the first one.

Appendix C

Audio Coding Form

Audio Coding Form

#	(min.sec)	Praise Statement	Type
EX	2.01	Great Job! Sitting quietly	S
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			

Coder:
Recording #:
Recording Length:
Date Coded:

# of specific praise (S) 0
# of general praise (G) 0

S = SPECIFIC PRAISE - a positive verbal response emitted by a teacher following a desired behavior emitted by a student or group of students that describes the desired behavior in observable terms.

G = General Praise - global or broad phrases that reflect a positive response to a desired behavior spoken by a teacher to a student or group of students (e.g., “Good job”, “Nice”, etc.).

Appendix D

Self-Management Forms and Visual Performance Feedback

Student Feedback Form

Remember, when someone praises you he/she says something nice about your behavior.

* Required

Did John praise your behavior today? *

- ○ No
- ○ Yes, one time
- ○ Yes, more than one time

Did John tell you which behavior he was praising?

- ○ Yes
- ○ No

Submit

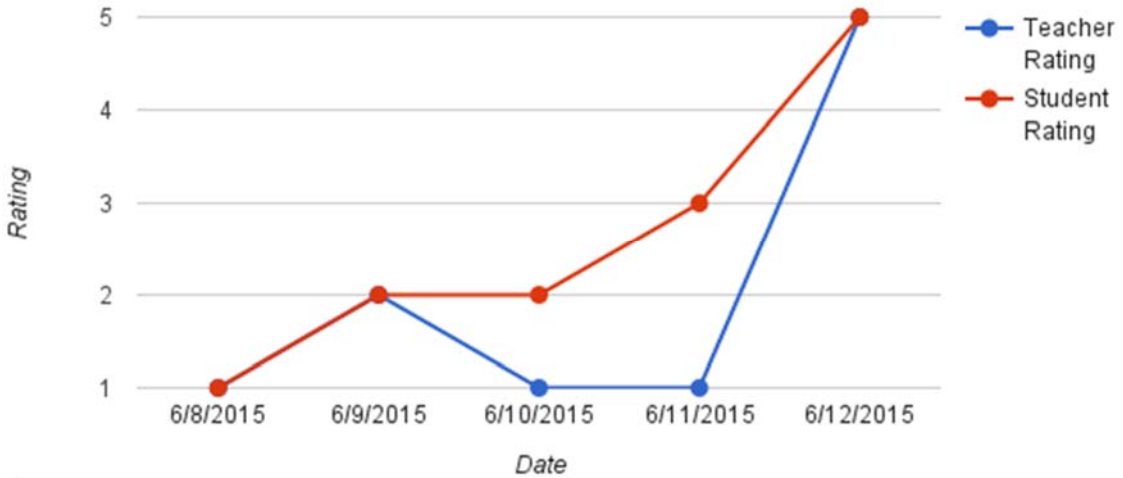
Teacher Rating Form

Please, indicate your rating for today

- 5 -- Outstanding -- I specifically praised 80-100% of my students
- 4 -- Very Good -- I specifically praised 60-80% of my students
- 3 -- Good -- I specifically praised 40-60% of my students
- 2 -- Ok -- I specifically praised 20-40% of my students
- 1 -- Poor -- I specifically praised 0-20% of my students

Submit

Visual Performance Feedback



6/8/2015	6/9/2015	6/10/2015	6/11/2015	6/12/2015
Perfect Match	Perfect Match	Next Door	No Match	Perfect Match

Figure D1. Comparison of teacher and student ratings.

Appendix E

Implementation Fidelity Checklists

Implementation Fidelity Checklists

Teacher Training Checklist

Trainer _____

Participant # _____

Date _____

Completed by _____

- ☐ Identify three benefits of using praise (e.g., reduced problem behavior, increased task engagement, enhanced student-teacher relationships, etc.)
- ☐ Define general praise statements and give examples/non-examples
- ☐ Define specific praise statements and give examples/non-examples
- ☐ Generate 6-10 examples of contextually appropriate specific praise statements with the teachers
- ☐ Describe and practice collecting student feedback using Google Forms
- ☐ Describe and practice completing the Teacher Rating Form using Google Forms
- ☐ Review examples of the performance feedback system
- ☐ Identify specific strategies to use when viewing the performance feedback
- ☐ Check for understanding

Student Training Checklist

Participant # ____

Site _____

Date _____

Completed by _____

- ☐ The trainer stated why these data are being collecting (e.g., "I'm asking you to do this to help me be a better teacher and club member.")
- ☐ The trainer described how these data will be collected each day
- ☐ The trainer defined praise
- ☐ The trainer provided an example and non-example of praise
- ☐ Students gave examples of praise
- ☐ The trainer described how students will complete the Student Feedback Form using the tablet
- ☐ The trainer explained that these data are anonymous

Appendix F

Modified Intervention Rating Profile-15

Modified Intervention Rating Profile-15

	Strongly Disagree	*Disagree*	*Slightly Disagree*	*Slightly Agree*	*Agree*	*Strongly Agree*
1. This intervention was effective in changing my behavior.	1	2	3	4	5	6
2. I would suggest the use of this intervention to other teachers.	1	2	3	4	5	6
3. I would be willing to use this intervention in the classroom setting.	1	2	3	4	5	6
4. This intervention would *not* result in negative side effects for the child.	1	2	3	4	5	6
5. This intervention would be appropriate for a variety of teachers.	1	2	3	4	5	6
6. This intervention improved students' behavior.	1	2	3	4	5	6
7. The time and effort required to participate in this intervention is reasonable.	1	2	3	4	5	6
8. I liked the procedures used in this intervention.	1	2	3	4	5	6
9. Overall, this intervention was good for our students.	1	2	3	4	5	6

What did you like about the training you received on praise?

What, if anything, would you change about the training?

What did you like about collecting self-appraisals or student feedback?

What, if anything, would you change about the collection of self-appraisals or student feedback?

What did you like about the performance feedback (graphs) you received?

What, if anything, would you change about the graphs?

CURRICULUM VITAE

CADE TOMAS CHARLTON

Counseling Psychology & Special Education
Brigham Young University
MCKB 340-A
Provo, UT 84602
(801) 422-1238
cade_charlton@byu.edu

EDUCATION

2016	Doctorate of Philosophy, Emma Eccles Jones College of Education, Utah State University, Disability Disciplines
2008	Master of Science, Jon Huntsman School of Business, Utah State University, Business Administration
2007	Bachelor of Science, Utah State University Major: Psychology Honors: Magna Cum Laude
2007	Bachelor of Science, Utah State University Major: Economics Honors: Magna Cum Laude

HONORS AND AWARDS

2008	Jon Huntsman School of Business Commitment to Excellence Award
2007	Utah State University, Department of Psychology Robins Award Nominee
2007	Utah State University, Department of Psychology Outstanding Student
2005	Utah State University A-pin Scholar (3 semesters 4.0 GPA maintained)

RESEARCH EXPERIENCE

2010-2013	Project Manager, Project LEARN Principal Investigator: Richard P. West, PhD *Managed a national project focusing on the identification of school trust lands in 12 western states. Activities included: evaluating the current management and investment practices of state land commissioners*

managing school trust lands; promoting best practices for land management and public policy; and developing stakeholder networks to support the preservation of school trust lands.

2008-2009 School Coordinator, Prevention Plus
Principal Investigators: Pamela Hallam, PhD and Richard P. West, PhD
Assisted in the selection of participant schools, training of school faculty, and evaluation of progress in the prevention of juvenile justice. Activities included: conducting school wide academic and behavioral screening, providing data-based decision making training, and completing project reports.

2008 Team Accountant, MBA Field Studies Consulting
Principal Investigator: Edwin Stafford, PhD
Interviewed fitness equipment consumers about their preferences for fitness equipment. Analyzed results using grounded theory for ICON Health and Fitness.

2007 Research Assistant, Public School Teacher Retention Study
Principal Investigator: Matthew J. Taylor, PhD
Contacted school administrators across the U.S. to determine the prevalence and causes of teacher termination. Assisted in the analysis and summary of data.

2005-2006 Undergraduate Laboratory Instructor: Utah State University
Psychology 1410: Learning and Behavior,
Instructor: Carl D. Cheney PhD
Operated and maintained operant lab for Psychology 1400.

2005 Research Assistant, Animal Safety and Training Research
Principal Investigator: Susan Friedman, PhD
Conducted a systematic literature review of training procedures to support appropriate feeding behaviors in zoo animals.

SCHOLARSHIP

Refereed Articles

Snyder, K., Pyle, D., **Charlton, C. T.**, Sabey, C., Lund, E. M., & Ross, S. W. (2015). A systematic review of the empirical support for Check-In Check-Out. *Journal of Positive Behavior Interventions,* Advance online publication. doi:10.1177/1098300715595957

Campbell, K., **Charlton, C. T.,** Maynes, J., & West, R. P. (2014). Utah. In D. Thompson & F. Crampton (Eds) State of the States 2013. *Journal of Education Finance*, 39(3), 283-286.

Taylor, M. J., Hallam, P. **Charlton, C. T.**, Wall, D. G. (2013). The Formative Assessment of Collaborative Teams. *NASSP Bulletin*, 98, 26-52 doi:10.1177/0192636513514109.

Wheatley, R. K., West, R. P., **Charlton, C. T.**, Sanders, R. B., Smith, T. G., & Taylor, M. J. (2009). Improving behavior through differential reinforcement: A praise note system for elementary school students. *Education and Treatment of Children, 32*, 551-571.

Trade journal articles

West, R. P., **Charlton, C. T.,** & Taylor, M. J. (In press). Students are the real experts on school reform and improvement. UASSP Impact Journal, xx(x), xx-xx.

Manuscripts under review or in preparation

Charlton, C. T., Ross, S. W., & West, R. P. (under review). Effects of a check-in/check-out system on the classroom engagement of at-risk high school students. Journal of Positive Behavior Interventions, xx(x), xx-xx.

Sabey, C., **Charlton, C. T.**, Pyle, D., Ross, S. W. (under review). A review of classwide or universal social, emotional, behavioral programs for students in kindergarten. Review of Educational Research, xx(x), xx-xx.

Ross, S.W., Lund, E. M., Sabey, C., & **Charlton, C. T.** (revise-resubmit). A behavioral definition of bullying. *Behavior and Social Issues.*

Taylor, M. J., West, R. P., **Charlton, C. T.,** & Smith T. G. (revise-resubmit). The link between school learning environment and academic achievement. Journal of Learning Environments Research, xx(x), xx-xx.

Charlton, C. T., West, R. P., Ross, S. W., (in preparation). Effects of a self-management procedure using student feedback on the use of behavior specific praise. Journal of Positive Behavior Interventions, xx(x), xx-xx.

Charlton, C. T., Sabey, C., Dawson, M., Pyle, D., & Ross, S. W. (in preparation). Critical incidents in the state-wide implementation of a multi-tiered system of support. Journal of Positive Behavior Interventions, xx(x), xx-xx.

Maynes, J., **Charlton, C. T.**, Bird, M., Campbell, K., & West, R. P. (in preparation). America's school trust lands: An untapped resource for school funding.

Sabey, C., Ross, S. W., **Charlton, C. T.,** & Pyle, D. (in preparation). Integrating social skills within Check-in Check-out: CICO+. *Education and Treatment of Children*, xx(x), xx-xx.

FUNDED PROJECTS

2014-2018 District-wide Evaluation of the Conditions for Learning and School Climate for the School Leadership Pipeline Grant in Granite School District. Scope of work includes data collection for 90+ schools, consultation with school principals, and work with district administration to examine the effects of the grant on school climate. (Total Funds: $201,000).
Role: Project Director & Co-Principal Investigator

2013-2014 Utah Personnel Development Center graduate research award to fund an exploration of the effects of Check-in/Check-out on the academic behavior of secondary school students (Total funds: $1,000).
Role: Principal Investigator

2012-2013 School Support Team member at Granger High School (ARRA SIG), Millville Elementary School, and Greenwood Elementary School. Provided assessment, evaluation, and professional development to priority schools during multi-year school improvement projects. (Total funds: appr. $280,000)
Role: School Improvement Consultant

2010 Funds for the expansion of the Prevention Plus project and the creation of the Land-grant Education and Research Network (Project LEARN) included in H.R.1105 - Omnibus Appropriations Act of 2009 (Total funds: $920,000).
Role (Prevention Plus): Grant Writer & Project Coordinator
Role (Project LEARN): Team Leader.

2009 AmeriCorps* VISTA School Improvement Partnership expansion funded by the American Reinvestment and Recovery Act (Total funds: $62,000).
Role: Co-Principal Investigator

2008-2009 Office of Juvenile Justice and Delinquency Prevention awarded the McKay School of Education at Brigham Young University funds to collaborate with the Center for the School of the Future and five project schools in northern Utah on innovative delinquency prevention efforts (Total funds: $268,000).
Role: Project Coordinator.

2006-2008 AmeriCorps* VISTA operational grant for the School Improvement Partnership (Total Funds through 2012: $1,276,000).
Role: Co-Principal Investigator

TEACHING EXPERIENCE

2016 Instructor: Brigham Young University
Counseling Psychology & Special Education 460: Collaboration

2015-2016 Instructor: Brigham Young University
Counseling Psychology & Special Education 402: Educating Students with Disabilities in Secondary Settings

2015 Instructor: Brigham Young University
Counseling Psychology & Special Education 300: Exceptional Students in Primary Settings: Principles of Collaboration

2014 Instructor: Utah State University
Psychology 1730: Strategies for Academic Success
Aggies Elevated

2014 Guest Lecturer and Teaching Assistant: Utah State University
Special Education 6770: Coaching and Systems Change

2012-2014 Student Teaching Supervisor: Utah State University
Special Education 5410: Practicum: Direct Instruction Reading & Language Arts for Students with Mild/Moderate Disabilities
Special Education 6030: Student Teaching in Special Education

2012 Instructor: Utah State University
Special Education 6280: Instructional Leadership for At-Risk Students.

2009 Teaching Assistant: Utah State University
Education 6570: Introduction to Educational and Psychological Research.
Instructor: Matthew J. Taylor, PhD

2009 Teaching Assistant: Utah State University
Special Education 6280: Instructional Leadership for At-Risk Students.
Instructor: Richard P. West, PhD

2006-2007 Undergraduate Teaching Assistant: Utah State University
Psychology 3460: Physiological Psychology.
Instructor: Carl D. Cheney, PhD

2004-2005 Undergraduate Teaching Assistant: Utah State University
Psychology 1400: Learning and Behavior.
Instructors: Carl D. Cheney, PhD Susan G. Friedman, PhD, and Ana Neves

PROFESSIONAL POSITIONS

2015-to date Visiting Instructor, Counseling Psychology & Special Education,
Brigham Young University, Provo, UT

2014-2015 Vice President of Client Engagement and Support,
Tetra Analytix, LLC, Logan, UT

2008-2013 Program Coordinator, The Center for the School of the Future
Utah State University, Logan, UT

2006-2008 AmeriCorps* VISTA Member,
Corporation for National and Community Service, Logan, UT

2006 Skills Development Specialist
Bear River Mental Health, Logan, UT

CONFERENCE PRESENTATIONS

Charlton, C. T., West, R. P., Hopkins, K. (2015). Creating the conditions for learning. Presentation at the Utah Multi-Tiered System of Supports Conference, Layton, UT.

Charlton, C. T., Sabey, C., Charlton, S. R. (2015). Examining the Effects of Varying the Availability of Positive Feedback on Performance of a Multiplication Task. Poster at the Association for Behavior Analysis International, San Antonio, TX.

Sabey, C., **Charlton, C. T.,** Charlton, S. R. (2015). The Effects of a 3:1 Positive to Negative Ratio on Performance During a Pattern Recall Task. Poster at the Association for Behavior Analysis International, San Antonio, TX.

Charlton, C. T., West, R. P., & Young, R. K. (2014). *Administrative Intervention*. Presentation at the Utah Multi-Tiered System of Supports Conference, Layton, UT.

Charlton, C. T. (2013). *Beyond storming: Building collaboration teams*. Utah Conference of School Community Councils, Invited Presentation, Draper, UT.

Snyder, K., Ross, S. W., Sabey, C., **Charlton, C. T.**, Pyle, D., Lund, E. M., & Slocum, T. A. (2013). Check-in/Check-out and Check, Connect, and Expect: A systematic review of common secondary interventions. Symposium at the Association for Behavior Analysis International, Minneapolis, MN.

Teemant, B., **Charlton, C. T.**, Wheatley-Wardle, R. K., West, R. P., Wessman, R., & Taylor, M. J. (2013). Improving attendance and school climate while reducing at-risk behaviors in a troubled urban high school implementing the Administrative Intervention model. Symposium at the Association for Behavior Analysis International, Minneapolis, MN.

Wessman, R., **Charlton, C. T.**, West, R. P., Teemant, B., Wheatley-Wardle, R. K., & Taylor, M. J. (2013). The effects of a school-based, targeted dropout prevention program on attendance and academic performance of at-risk students in an urban high school. Symposium at the Association for Behavior Analysis International, Minneapolis, MN.

Oaks, R., Wheatley-Wardle, R. K., **Charlton, C. T.**, West, R. P., Teemant, B., & Taylor, M. J. (2013). An analysis of a schoolwide praise note system implementation in an urban high school: Methods, penetration, and teacher participation. Symposium at the Association for Behavior Analysis International, Minneapolis, MN.

Charlton, C. T., Bird, M., Maynes, J., Campbell, K., & West, R. P. (2013, May). America's School Trust Lands: An Untapped Resource for School Funding. Paper session at the National Education Finance Conference, Indianapolis, IN.

Ross, S. W., Sabey, C., **Charlton, C. T.**, & Pyle, D. (2012). *Evaluation of a Novel Approach to Social Skills*. Paper presentation at the Utah Conference on Effective Practices For Teachers and Human Service Professionals: Interventions Across the Lifespan, Logan, UT.

Charlton, C. T., Taylor, M. J., & West, R. P. (2011, Sep). *FACT: Developing and sustaining collaborative learning communities through self-management tools*, Paper presentation at the International Conference on Children with Behavioral Disorders, New Orleans, LA.

Wheatley, R. K., **Charlton, C. T.,** & West, R. P. (2011, May). *Using ClockLight to increase student on-task behavior in an after school club setting: Effectiveness, sustainability, and teacher satisfaction,* Poster presentation at the International Conference of the Association for Behavior Analysis International, Denver, CO.

Taylor, M. J., West, R. P., **Charlton, C. T.**, & Wheatley, R. K. (2009). *Elements of effective instruction and academic performance: The achievement gap explained*. *Poster*, Logan, UT: The Center for the School of the Future.

Wheatley, R. K., West, R. P., & **Charlton, C. T.** (2009, March). *Increasing Appropriate Lunchroom Behavior: A Praise Note System for Elementary Students*, Paper presentation at the Association of Positive Behavior Support Conference, Jacksonville, FL.

Taylor, M. J. & **Charlton, C. T.** (2008, September). *Teacher Behaviors that Predict Average Daily School Attendance*, Paper presentation at the Teacher Educators for Children with Behavioral Disorders Conference, Tempe, AZ.

West, R.P., Taylor, M.J. Wheatley, R.K., and **Charlton, C.T.** (2008, February) *Four Critical Elements of Instruction and their relationship to overall school quality*. Address, California Association of Behavior Analysis: Garden Grove, CA.

Wardle, R.K. & **Charlton, C.T.** (2007, March) *ALERT: Student report as a more sensitive measure of school learning environment.* [Poster]. Association of Positive Behavior Support: Boston, MA.

Larson, L.L., Wardle, R.K., & **Charlton, C.T.** (2007, March) *Increasing attendance and punctuality and decreasing office discipline referrals: The role of school-wide behavior support*. [Poster]. Association of Positive Behavior Support: Boston, Ma.

Charlton, C.T., & Charlton, S. R. (2006, May). *The Role of Commodity Type and Reciprocity in the Iterated Prisoner's Dilemma*. [International Poster Session]. Association for Behavior Analysis, International: Atlanta, GA.

Branch, M. N., **Charlton, C**., Fisher, W., Madden, G. J., & Mawhinney, T. C. (2006, May). *Professional Development Series: Perspectives on the Future of Behavior Analysis*. [Panel Discussion] Association for Behavior Analysis, International: Atlanta, GA.

Wheatley, R.K & **Charlton, C.T.** (May, 2006). *Clocklight: Affecting group behavior through immediate feedback*. Association for Behavior Analysis, International: Atlanta, GA.

Bailey, J.S., Johnston, J.M., Mellichamp, F., Shook, G.L., **Charlton, C.T.** (2005, May) *Professional development series: On being a board certified behavior analyst.* Assocaition for Behavior Analysis, International: Chicago, IL.

Wheatley, Rikki K & **Charlton, C.T.** (May, 2006). *Increasing appropriate lunchroom behavior: A praise Note System for Elementary Students*. Association for Behavior Analysis, International: Atlanta, GA.

West, R.P., Smith, T.G., Wheatley, R.K., and **Charlton, C.T.** (2006, March) *Clocklights: Affecting group behavior using immediate feedback*. Poster Presentation, Association for Positive Behavior Support: Reno, NV.

West, R.P., Smith, T.G., Wheatley, R.K., and **Charlton, C.T.** (2006, March) *Increasing appropriate lunchroom behavior: A praise note system for elementary students*. Poster Presentation, Association for Positive Behavior Support: Reno, NV.

Wheatley, R.K., **Charlton, C.T.**, Willett, B., Torres, M.M., Fullmer IV, W.P., Heninger, M. & Wheatley, M. (July, 2006). *Principles for effective teaching*, Boys and Girls Club of Cache Valley Employee Training.

HONORS SOCIETIES

2008-to date	The International Honor Society Beta Gamma Sigma
2007-2008	The Honor Society of Phi Kappa Phi
2007-2008	The National Society of Collegiate Scholars
2001	National Honors Society

PROFESSIONAL AFFILIATIONS

2015-to-date	USOE School Turnaround Experts-Tetra Analytix
2010-to date	Utah State Office of Education School Support Team
2004-to date	Association for Behavior Analysis International
2006-to date	Organizational Behavior Management Network
2006-to date	Association of Positive Behavior Support