Classwide Peer Tutoring in Physical Education: Assessing Its Effects With Kindergartners With Autism

Phillip Ward and Shiri Ayvazo
The Ohio State University

Researchers, textbooks authors, and educational policy makers recommend peer tutoring as an inclusion strategy for students with autism. However, there is little, if any, research supporting these recommendations in physical education. We assessed the effects of classwide peer tutoring (CWPT) in teaching catching skills to two typically developing peers and two children diagnosed with autism in kindergarten. A single subject withdrawal design assessed the effects of CWPT on total catches and correct catches. Results show that CWPT improved total and correct catches for the two students with autism. The results for the typically developing peers were mixed. These findings, while requiring further research, provide initial evidence to support CWPT as an inclusion strategy for children with autism in physical education.

Autism and related disorders, also called autism spectrum disorders or pervasive developmental disorders, are characterized by atypical social interaction, communication, and behavioral development (DSM-IV-TR; American Psychiatric Association, 2000). Although prevalence estimates for autism spectrum disorders vary, the frequency is generally reported as 1-2 individuals per 1000 in the US population (Bryson & Smith, 1998). Inclusion for children with autism is critical if they are to acquire developmentally appropriate social, communicative, and behavioral skills. In their review of effective teaching strategies for children with autism in inclusive settings, Harrower and Dunlop (2001, p. 763) note that students with autism who are fully included “(a) display higher levels of engagement and social interaction, (b) give and receive higher levels of social support, (c) have larger friendship networks, and (d) have developmentally more advanced individualized education plan goals than their counterparts in segregated placements.”

Despite these findings, there is still much debate on the topic of inclusion in the literature. Harrower and Dunlop (2001) note that the question ought not to be framed as a dichotomy between segregated versus inclusive placements, but rather should be a question of determining effective supports for students with autism in inclusive settings. However, because autism encompasses a broad range of disor-
ders and a large variance of social, communicative, and developmental behaviors, determining effective supports is not a “one size fits all” response.

There is increasing empirical evidence supporting inclusion in the adapted physical education literature (Porretta & Sherrill, 2005). However, there is a dearth of research focusing on autism. For example, Porretta and Sherrill (2005) report that only one study addressing autism was published in *Adapted Physical Activity Quarterly* from 1994-2004. One strategy that has been promoted as a best practice for inclusion is peer tutoring (Ward & Lee, 2005). Peer tutoring has been extensively investigated in physical education settings with typically developing children where it has been shown to improve the percentage of correct performance of motor skills (Crouch, Ward, & Patrick, 1997; Ward, Smith, Makasci, & Crouch, 1998) and to increase student scores on motor skill tests (Ernst & Byra, 1998; Goldberger & Gerney, 1986; Goldberger, Gerney, & Chamberlain, 1982; Virgilio, 1985).

Peer tutoring has also been successful in improving the performances of students with disabilities. For example, it has helped to increase the moderate to vigorous physical activity of students who are deaf (Lieberman, Dunn, van der Mars, & McCubbin, 2000), to increase developmentally disabled students’ correct performance of motor skills (Houston-Wilson, Dunn, van der Mars, & McCubbin, 1997), and to increase the academic learning time of children with moderate to severe developmental disabilities (DePaepe, 1985; Webster, 1987).

Classwide peer tutoring (CWPT) is a particular version of peer tutoring. In CWPT, the entire class is involved in reciprocal roles of tutor and tutee. The effectiveness of CWPT has been demonstrated in both regular and special education and in elementary, secondary, and college settings (Greenwood, Maheady, & Carta, 1991). In CWPT, a class is divided into small groups of 4-6 students. Within each team, students pair themselves or may be paired by the teacher. Instructional tasks are presented on task cards in text or with pictures. Following teacher demonstration of the tasks, students practice and are given a fixed amount of time to conduct a partner assessment. As the tutee performs each trial, the tutor provides feedback and records the number of correct performances. After all members of the team have completed their assessments, a member of each team posts the number of correct performances performed by the group under the team name on a poster board. At the start of subsequent lessons, the teacher establishes a goal for each group to meet. Reinforcers are often tied to meeting the daily goals set by the teacher.

The CWPT strategy utilizes an interdependent group contingency (Cooper, Heron, & Heward, 1987), holding students individually and as a group accountable for performance. Only one study has assessed this strategy in physical education. Johnson and Ward (2001) assessed a 3rd-grade striking unit and reported that CWPT was effective in increasing correct and total catches, for both low and high skilled students and for both boys and girls. The purpose of this study was to extend the research by assessing the effects of CWPT in an inclusive setting with children with autism and typically developing peers. In particular, we wanted to assess the extent to which students with autism could participate in the same activities as their typically developing peers. We chose catching because it is a critical skill for kindergartners (Thomas, Lee, & Thomas, 2000) and because it requires, for most tasks, the cooperation of students in pairs.
Method

Setting
This study was conducted in a K-8 charter school in the Midwest, specializing in inclusion placements for children with autism. The teacher had an undergraduate degree in physical education and two years direct experience working with children who had emotional and behavioral disorders. Each classroom served 5-7 children with autism and 8-10 typically developing peers. Although all students were in classrooms that used peer tutoring, it had not been previously used in their physical education lessons in formal ways. This study was conducted with a kindergarten class of 16 children. Six children in the class were diagnosed with autism. Data were collected on four participants.

Description of Participants
Ben, 8 years old, had been diagnosed with behavioral and developmental patterns consistent with autism by a licensed psychologist. Though verbally capable, he had difficulty expressing his wants and needs. He did not respond independently unless motivated or prompted to do so. Peter, 8 years old, had received an independent psychiatric diagnosis of autism described as pervasive developmental disorder by a licensed psychologist. His full-scale IQ test score was 54, based on the Wechsler Intelligence Scale for Children-Revised III (1992). Most verbal communication was prompted. He did not respond independently unless motivated or prompted. In addition to Ben and Peter, two typically developing kindergarten peers, Holly and Sarah, were purposely selected by the teacher in response to the investigator’s request for students who in the opinion of the teacher would be good peer tutors for Ben and Peter. Sarah served as Ben’s peer tutor, while Holly served as Peter’s peer tutor.

Physical education lessons were held twice a week for 30 min. This study occurred during a 26-lesson motor skill unit focusing on catching and striking. The lessons were conducted in an irregular shaped indoor area, approximately 400 sq ft (122 sq m) in size.

Data Collection and Procedure
Two dependent measures are reported in this study. First, the number of catches participants made during each session was used to provide a measure of the level of engagement and the work completed by students. Our focus was on catches rather than time on learning, because time is a proxy rather than direct measure of student learning. Second, the number of catches that were correct was used to provide a measure of the degree to which students were able to perform skills that their peers were performing (i.e., all students were performing the same skills). The criteria for correct performance included two critical elements stated by the teacher as well as the successful accomplishment of the task. An example of a task and its critical elements is “successfully catch a self tossed ball below head height, with both hands” (a complete list of critical elements is available from the authors upon request). Thus, observers coding each skill were required to make decisions about the critical element and the success of the performance.
The entire unit, including skills, the critical elements, and the development sequences appropriate for kindergarteners was derived from the Thomas et al. (2000) and the Graham, Holt-Hale, and Parker (1998) texts. The unit was then reviewed by an elementary physical education and motor development expert to determine (a) if each task was equivalent to the previous one in level of difficulty and (b) if the tasks were developmentally appropriate.

Each lesson was arranged in the following manner. Students entered the gymnasium and met with the teacher. Teacher aids were present observing but did not interfere with the participating students during the lesson unless the child “wandered off.” After a brief introduction, students engaged in moderate to vigorous physical activity for 10-min. The lesson then began with task demonstrations followed by practice. Each task taught by the teacher was followed by a two-min practice episode. During each lesson, students performed four activities, two catching tasks and two striking tasks. Generally, the first activity for each skill had been introduced the previous lesson, while the second activity was typically new.

Observer Training. All 26 lessons were videotaped and coded. Observers received training prior to the start of data collection that included a copy of definitions for all tasks, explanations about each task, and its critical elements. Observers then practiced coding from videotapes and received feedback. Finally, observers were assessed on their coding accuracy using a 30-min tape and were required to meet a criterion level of 95% accuracy or better.

Experimental Design and Conditions

An A-B-A-C single-subject withdrawal design was used to evaluate the effects of CWTP on students’ motor performance (Cooper et al., 1987).

Whole Group Direct-Instruction (WG). This served as the baseline. The teacher provided explanations and demonstrations after which students performed the task while she monitored the activity and provided feedback. Hot spots were used to cue students where to stand and music was used to cue the students and the teacher as to the duration of practice times (i.e., different music for each 2-min phase of the practice time).

Classwide Peer Tutoring-1. The CWPT-1 condition was arranged in the following manner. Because of her knowledge of the students, the classroom teacher divided the class into groups of four students. Students with autism were paired with typically developing peers in each group. In these pairings, the tutoring was unidirectional. In dyads where typically developing children were together, the tutoring was reciprocal. Following a demonstration of the task, students performed each task during a 2-min period that was timed by music. When the music was playing, students were to perform the task and when the music was not playing, they were to stop and listen for instructions. During practice, the peer tutor asked his or her partner to perform the task and waited a short time for a response (approximately 4 seconds). If no response occurred, the tutor prompted or assisted the partner to perform the task. If a correct response occurred, the tutor praised the performance. If an incorrect response occurred, the tutor prompted or assisted the partner in
performing the correct task. Assistance was typically provided by placing the ball in the hand of the child and prompting performance with a statement such as, “Watch me—this is how you do it!”

As part of the CWPT independent variable, we added a reinforcement component. The physical education teacher monitored the activity during practice time and placed a sticker next to a student’s name on a wall chart organized by groups as she identified correct performance. The sticker functioned as a generalized reinforcer contingent upon correct performance of the task (Cooper et al., 1987). At the initial stage of the intervention, students were encouraged to earn two stickers to be eligible to pull a tangible reinforcer (e.g., pencil, eraser, key-chain) from a “mystery bag” at the end of each class. After the second day of the intervention, the criterion was increased to three stickers.

**Classwide Peer Tutoring-2.** This phase was different from the CWPT-1 in only one respect. During the second minute, the tutors focused on their own performance with the goal of modeling both the behavior and the on-task engagement. Each minute of practice time was cued with a different type of music.

**Training of Tutors.** All students received a single 30-min training prior to the beginning of intervention CWPT-1. When the intervention was introduced for the second time (i.e., CWPT-2), students received 10-min “refresher” training session. The initial training session consisted of a short discussion led by the teacher of the value of cooperation. The students practiced, in turn, demonstrating, providing feedback and encouragement using mock tasks. During this time the teacher provided feedback as to the appropriate decisions being made.

**Treatment Integrity**

Fifty percent of the intervention sessions were assessed for treatment integrity. We assessed whether peers responded to each error or period of off-task behavior by observing if the tutor provided a verbal or physical prompt or used a statement such as “Watch me—this is how you do it.” However, since appropriate catches often occurred continuously, we simply coded if encouragement occurred during the period of correct responses. The data show that Holly responded to off-task or incorrect responses by Peter 87% of the time. She provided encouragement during 84% of the intervals when Peter was successful. Sarah responded to off-task or incorrect responses by Ben 80% of the time. She provided encouragement during 70% of the intervals when Ben was successful.

**Interobserver Agreement**

Interobserver agreement (IOA) was assessed on a randomly chosen sample of 37% of the observations. IOA was calculated for the total and correct number of catches for each participant by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100% (Cooper et al., 1987). The mean IOA was 92% for Holly (range, 87%-100%), 95% for Ben (range, 85%-100%), 93% for Sarah (range, 85%-100%), and 99% for Peter (range, 92%-100%).
Results

Figure 1 presents the data for the number of total and correct catches performed by the participants. The data for day 4 were excluded since the task was found to be unequal (i.e., easier) than the other tasks the students were required to perform.

Holly. During the first baseline (WG-1), the data for total catches were stable with a mean of 10 catches (range, 4-14 in 2 minutes) as were the data for correct catches with a mean of 6 (range, 1-11). During phase B (CWPT-1), the data for total catches increased to a mean of 18 (range, 12-25). A slight upward trend was demonstrated for correct catches with a mean of 8 (range, 2-16). Total and correct catches decreased during the second baseline (WG-2). The mean of total catches

![Graph showing data for Holly and Ben](image-url)

*Figure 1* — Figure 1 presents the data for the number of total and correct catches performed by Holly, Ben, Sarah, and Peter, respectively.
was 9 (range, 3-13) and the mean of correct catches was .3 (range, 0-1). The number of total and correct catches increased during phase CWPT-2. The mean of total catches was 18 (range, 8-28) and the mean of correct catches was 10 (range, 0-22). No data were collected on day 23 because Holly was absent.

For Holly 38% of the total trial data points overlapped between WG-1 and CWPT-1, and 17% overlapped between WG-2 and CWPT-2. Total catches increased from WG-1 and CWPT-1. Baseline was recovered during WG-2, followed by an increased level during CWPT-2. For correct catches 75% of the data points overlapped between WG-1 and CWPT-1, and 17% overlapped between WG-2 and CWPT-2. Correct catches were generally stable during WG-1 and became more instable with a slight upward trend during CWPT-2. Few correct catches were performed during WG-2, while an increase in correct catches was demonstrated during CWPT-2.
Ben. During WG-1, the data for total catches were generally stable with a mean of 7 catches (range, 4-11). Correct catches were stable with a mean of .4 (range, 0-2). No data were collected on day 3 since Ben was absent, or reported for day 8 because he received additional support on that day. During CWPT-2, both total and correct catches increased though they were highly variable. The mean of total catches was 21 (range, 5-38) and the mean of correct catches was 13 (range, 0-35). Baseline was recovered during WG-2 with a mean of 5 (range, 0-9) for total catches and a mean of 1 (range, 0-2) for correct catches. During CWPT-2, total and correct catches increased in level. The mean of total catches was 30 (range, 18-40) and the mean of correct catches was 20 (range, 12-29). No data were collected on day 20 because Ben was absent.

For Ben, 25% of the data points for total catches overlapped between WG-1 and CWPT-1, and there were no overlapping data points (0%) between WG-2 and CWPT-2. For correct catches, 38% of the data points overlapped between WG-1 and CWPT-1, and there were no overlapping data points (0%) between WG-2 and CWPT-2. Total and correct catches were generally stable during WG-1, but were variable during CWPT-1. During WG-1, total and correct catches returned to baseline levels. Total and correct catches increased in level during CWPT-2 yet were variable.

Sarah. During WG-1 a descending trend was demonstrated for total catches with a mean of 15 catches (range, 5-26) and for correct catches with a mean of 5 (range, 0-15). During CWPT-1, a descending variable trend was demonstrated for total catches with a mean of 17 (range, 11-34). A variable trend was also demonstrated for correct catches with a mean of 7 (range, 0-18). Baseline was not recovered during WG-2. The mean of total catches was 13 (range, 7-19) and the mean of correct catches was 7 (range, 2-13). A variable trend was demonstrated for total and correct catches during CWPT-2. The mean of total catches was 19 (range, 14-30) and the mean of correct catches was 9 (range, 2-13). No data were collected on days 6, 19, and 23 because Sarah was absent.

The majority of data points (75%) for total catches overlapped between WG-1 and CWPT-1, and 67% overlapped between WG-2 and CWPT-2. The variable trend for total and correct catches was maintained throughout the phases of the study. The majority of data points for correct catches overlapped between WG-1 and CWPT-1 (88%) and between WG-2 and CWPT-2 (83%).

Peter. During WG-1, the data for total catches were overall stable with a mean of 5 (range, 0-18) and there were no correct catches. During CWPT-1 a descending trend followed by a sharp ascending trend was demonstrated for total catches, with a mean of 18 (range, 4-43). A stable trend was demonstrated for correct catches with a mean of 3 (range, 0-12). Total catches decreased during WG-2 with a mean of 3 (range, 0-8), while no correct catches were performed during this phase. A higher level of total catches, with a descending trend, was demonstrated during CWPT-2. The mean was 13 (range, 5-22). Data were for the most part stable for correct catches during CWPT-2 with a mean of 4 (range, 1-17).

For total catches, 50% of the data points overlapped between WG-1 and CWPT-1, and 29% overlapped between WG-2 and CWPT-2. Total catches increased from WG-1 and CWPT-1. An instable recovery to baseline occurred on WG-2, and an increased level occurred again on CWPT-2, with a descending trend. For correct catches, 38% of the data points overlapped between WG-1 and CWPT-1, and 14%
overlapped between WG-2 and CWPT-2. Correct catches were not performed during WG-1 or WG-2, but were performed during CWPT-1 and CWPT-2.

**Discussion**

The purpose of this study was to assess the effects of CWPT as an inclusion strategy for children with autism in physical education. If one accepts that engagement (i.e., total catches) is one indicator of successful inclusion, then one conclusion of this study is that merely placing Ben and Peter in the class did not result in successful inclusion during either baseline condition. During the first introduction of CWPT, the data for both students showed an immediate improvement; however, the data path was variable for both. In both cases, however, performance during CWPT-1 was substantively better than baseline. Original baseline levels of performance were replicated during the reintroduction of the baseline indicating the stimulus control properties of CWPT. During the CWPT-2, performance for both students immediately and substantively improved. However, there was a descending trend for both toward the latter part of this phase. These data, although variable, do provide evidence that for Ben and Peter, the CWPT intervention was more effective than whole group instruction in meeting the goal of including these students into physical education. The results for correct performance, albeit not as substantive as those for total catches, nonetheless, show that Ben and Peter performed more correct catches during CWPT than during regular whole group direct instruction.

Among the limitations of this study are that the initial baseline performance for Peter and in the return to the baseline performance for Sarah and Ben, were improving before the intervention. Because this was a group setting and the CWPT intervention was delivered to the entire class, we made decisions on the basis of the general trend of all four participants rather than each individual participant. This makes experimental control problematic for Sarah and to a lesser extent with Peter because the differences in performance between baseline and peer tutoring are not highly discriminable.

At least two factors may account for idiosyncratic variance in performance across students. First, it may have been too fast a pace within and across classes for these students to learn these new skills. Perhaps allowing more time to practice within and across lessons would have increased opportunities to more correctly perform these skills. Second, the “one size fits all” approach (i.e., all students performed the same skills at the same time) might not have been the best choice when students have markedly different baseline performances. An alternative to this might be to use differentiated tasks. One way this might be achieved is to use a circuit approach with each task to be performed at each station having three levels of difficulty. All students might start at the lowest level and then progress toward the more advance level when the teacher indicates to them to do so. In this way, the character of the task for students would stay the same but it would incorporate three levels of difficulty at each station. To date, no studies have investigated this modification of CWPT. It clearly warrants inquiry.

These findings need to be viewed in the context of research on students with autism in general education. The literature suggests two reasons for the inclusion of children with autism in regular classrooms. First it is argued that because of the unique social challenges that children with autism have, inclusion is critical
if they are to be placed in situations where they can observe, imitate, and receive discriminated feedback and prompts (Harrower & Dunlop, 2001). Second, segregating children with autism from regular classrooms is a denial of basic civil rights of education (Ramey & Ramey, 1998). These rationales have shaped research in classrooms toward social outcomes, rather than academic outcomes. At present there is limited literature addressing academic outcomes for children with autism and similar to the findings of this study, mixed results have been reported (Harrower & Dunlap, 2001). In this study we chose to focus on “academic” outcomes of physical education (i.e., motor learning) rather than social outcomes.

The data for the two typically developing students do not show as substantive improvement as has been reported in other studies (e.g., Johnson & Ward, 2001). Holly’s performance shows improvement during CWPT for both total catches and correct catches, while Sarah shows little improvement for total catches and marginal improvement for correct catches during the CWPT. It is difficult to interpret this finding. In the majority of studies using PT with typically developing students, the tutoring arrangement has been reciprocal (Ward & Lee, 2005). In contrast, in the majority of the studies using students with disabilities, the studies have been unidirectional. That is the performance for typically developing peers has not been reported. So it is unknown what the “cost” in terms of learning is to students who peer tutor students with disabilities. Although our findings require further study, they do give rise to the question, “What is the cost to the tutor?” At the very least, changes in procedures or cycling through different peers may be a strategy to share the cost among students in the class.

There are two issues regarding the independent variable that require elaboration. First, in this discussion we have referred to the independent variable as CWPT. However, in the methods section, we describe the difference between the CWPT-1 and the CWPT-2. In general, students did better in the CWPT-2 than during CWPT-1. But because neither the CWPT phases were replicated, we cannot draw conclusions as to the relative effects of each. Our anecdotal observations suggest that they were not too different. Second, the package in this study included a generalized reinforcer for tasks completed. We cannot determine the differential effects of the generalized reinforcer because we did not include a component analysis in this study. Our anecdotal observations suggest that the generalized reinforcer may have been a component for the typically developing students, but we think it was not for the students with autism as evidenced by their minimal interest in the “grab bag.”

Due to our focus, we did not measure communicative and social variables, such as tracking verbal initiations of peers and children with autism, as well as cooperative and compliance behaviors. However, given the importance of these variables for children with autism, they ought to be measured in future studies. Certainly the issues of time in terms of task accomplishment and differentiated instruction discussed earlier should be investigated. These are easily alterable variables that based on our experiences in this study warrant investigation.

Horner et al. (2005) suggest that single subject research can contribute to evidence based practice if studies operationally define the practice, specify the contexts in which the practices are to be used, demonstrate fidelity of
treatment, and demonstrate both internal and external validity. This study meets the first four of these five criteria. Demonstration of external validity will wait on future replications of these findings, which, although not conclusive, provide initial support for the use of CWPT in inclusive settings and for the direction for future research.

References


