# The Effect of Therapy Balls on the Classroom Behavior and Learning of Children with Dyslexia

Leilani B. Goodmon
Raven Leverett
Amanda Royer
Gracia Hillard
Tracey Tedder
Lori Rakes
Florida Southern College

Author Note

Correspondence concerning this article should be addressed to: Dr. Leilani B. Goodmon, Department of Psychology, 111 Lake Hollingsworth Drive, Lakeland, FL 33801-5698, lgoodmonriley@flsouthern.edu.

Author's note: The Office of the Provost at Florida Southern College provided partial funding for this research. The authors thank the following: the associate head of The Roberts Academy, Adella Spires, M.Ed., and her assistant Lisa South; the teachers, Shari Richard, M.Ed, and Sonja Skaggs, M.A.; and the following research assistants for their help with data collection: Lauren Dill, Siobhan Buckley, Brandon Hardy, Stephanie Thacker, and Rachel Rodgers.

*Abstract*

*The purpose of this study was to determine if the behavioral and learning benefit found from the use of therapy balls as classroom seats in children with attention deficit and hyperactivity disorder (ADHD) (e.g., Schilling, Washington, Billingsley, & Deitz, 2003) generalizes to children with dyslexia who suffer from similar attention problems as children with ADHD (e.g., Germano, Gagliano, & Curatolo, 2010). We found a positive impact of therapy balls on behavior from the independent observers' and teachers' perspectives in two classes. Although there was no significant improvement in reading comprehension or in student self-reports of behavior, most students indicated an improvement in attention and motivation and a preference for their continued use. Results are discussed in terms the importance of evaluating the individual needs of the students when determining whether the therapy balls should be used as a sensory modulation technique in the classroom.*

Dyslexia is a specific learning disorder characterized by a pattern of difficulties with word recognition and spelling despite adequate intelligence and educational instruction (American Psychiatric Association, 2013). In addition to reading impairments, children with dyslexia also suffer from other attention deficits that might impair the learning process, including slower visual and auditory attention (Facoetti, Lorusso, Cattaneo, Galli, & Molteni, 2005), problems with fast attention shifts (for review see Jaskowski & Rusiak, 2005), slower mental rotation (Kershner, 1979), and greater cognitive impulsivity (for review see Donfrancesco, Mugnaini, & Dell'Uomo, 2005).

Others report that children with learning disabilities in reading tend to behave more impulsively in the classroom (Routh, 1979) and are more distractible and hyperactive (Kavale & Forness,

1996). Children with dyslexia are often also diagnosed with ADHD (Willcutt & Pennington, 2000a), with reported rates of comorbidity ranging from 16.9% (Germano et al., 2010) to 20% (Karande et al., 2007). Others report that children or adolescents with dyslexia are more withdrawn, anxious, and depressed, and have more somatic complaints, social problems, and attention problems (e.g., Dahle, Knivsberg, & Andreassen, 2011). There are also reports of more aggressive and rule breaking behaviors (Willcutt & Pennington, 2000b). Given these cognitive and behavioral issues, it is important to explore classroom techniques that might improve the attention, engagement, behavior, and learning of children with dyslexia.

Some physical and occupational therapy researchers suggest that children with attention problems suffer from sensory modulation deficits (e.g., Mulligan, 2001) and move excessively in order to increase stimulation (Zentall, 2007). In support of Optimal Stimulation Theory (for review see Kercood, Grskovic, Lee, & Emmert, 2007), physical activity (e.g., yoga) appears to have a calming effect on children with ADHD (Jensen & Kenny, 2004). Within the context of the classroom, others report the effectiveness of various sensory processing techniques (e.g., Watling, Deitz, Kanny, & McLaughlin, 1999) (also called sensory modulation strategies) designed to provide learning disabled children with the sensory input needed to maintain effective arousal states, appropriate classroom behavior, and academic performance.

One classroom intervention that has shown some promise in improving attention, behavior, and some learning outcomes in children with attention or behavior problems (e.g., ADHD, Specific Learning Disabilities in reading or math, or Autism Spectrum Disorder (ASD)) is the use of therapy balls in place of classroom seats. In addition to the emission of excess energy, the reported benefits of therapy balls in the classroom include improved focus (Fedewa & Erwin, 2011; Schilling et al., 2003; Schilling & Schwartz, 2004), engagement (Fedewa & Erwin, 2011; Schilling & Schwartz, 2004), in-seat behavior (Fedewa & Erwin, 2011; Schilling et al., 2003; Schilling & Schwartz, 2004), legible word productivity (Schilling et al., 2003), and listening comprehension (Kercood & Banda, 2012). The therapy balls are believed to have these benefits because they require the appropriate implementation of physical activity in the classroom allowing the children to emit excess energy in order to attain and sustain an optimal level of arousal needed for learning (Mulligan, 2001).

Although several researchers have reported positive benefits of therapy balls, it is important to note that in addition to a small sample sizes, the majority of these studies were exclusively qualitative in nature, as the researchers did not conduct any significance tests (with the exception of the hyperactivity scores in the Fedewa and Erwin (2011) study). Therefore, it is unclear whether the reported improvements in attention, learning, behavior, and social validity were of any practical significance. Also because the studies did not include more than one sample, there is no way to discern if the improvements would generalize to other children with similar attention problems or to other classrooms with different student dynamics and teaching styles. In only two of the studies (Schilling & Schwartz, 2004; Schilling et al., 2003) did the samples entirely consist of children who had a formal diagnosis of an attention problem in either the form of ADHD or ASD. In the majority of the studies (with the exception of Fedewa and Erwin, 2011) the teachers' social validity reports and preferences were based on teacher comments instead of quantitative, scaled measures.

Finally, to date there is no specific research on the use of therapy balls on children with dyslexia. There are several reasons to believe that children with dyslexia will exhibit similar improvements from the use of therapy balls as children with attention problems in the previously described studies. In addition to the high rate of comorbidity between dyslexia and ADHD, like those with ADHD, participants with dyslexia yield similar time on task behavior (Dhar, Been, Minderaa, & Althaus, 2010), have similar cognitive deficits and slower processing speeds (Willcutt et al., 2010), and fail to develop age appropriate reading skills (Robertson & Joanisse, 2010).

Given the aforementioned gap in the literature, the purpose of the current study was to determine the influence of therapy balls on measures of desirable (e.g., *staying in seat*) and undesirable classroom behaviors (e.g., *talking off topic to classmate*), and reading comprehension in two samples of fifth grade students with dyslexia (with and without a diagnosis of ADHD). Although Schilling and colleagues (2003) found an improvement in legible word productivity and Kercood and Banda (2012) found an improvement in listening comprehension, to date no researchers have examined reading comprehension as a learning outcome.

In addition to examining a sample of children with dyslexia, we also improved the scientific rigor of the aforementioned studies by increasing the sample size and objectively measuring behavior from three different perspectives (i.e., independent observers', students', and teachers' quantitative ratings). Also, in order to determine if the results generalized to other students, we included two samples of students from separate classrooms in a time series switching replication design so that the original control class utilized the therapy balls later in the study. With the exception of historical change patterns that match the time sequence of the treatment introductions, this design rules out the majority of threats to internal validity (see Shadish, Cook, & Campbell, 2002).

## *Method*

### *Participants*

Twenty-four fifth grade students (12 from each class) from a school for children with dyslexia participated in this study. All of the students were formally diagnosed with either dyslexia only or with dyslexia and ADHD. The dyslexia and ADHD comorbidity rate was 41.67% in one class and 50% in the other. The sample was 79.2% male and 20.8% female and ranged in age from 9-11 years ($M$ = 10). Both classes followed the same curriculum and schedule, and had other similar demographic characteristics.

### *Apparatus*

The therapy balls were systematically sized to each child using a standard height chart (adjusted by weight) so that each child could sit on the center of the ball with hips and knees at a 90-degree angle and feet flat on the floor.

*Design and Procedure*

We employed a time series (pretest–posttest control group) with switching replications design (shown below, where *T* is Time Wave and *X* is Treatment).

| | | | | | |
|---|---|---|---|---|---|
| Class A | $T_1$ | X | $T_2$ | | $T_3$ |
| Class B | $T_1$ | | $T_2$ | X | $T_3$ |

The three time waves of data collection ($T_1$, $T_2$, $T_3$) within the classrooms lasted 15 days each. In each time wave, we collected reading comprehension scores and behavioral data from the independent observers', teachers', and students' perspectives (see measures). After baseline measures were collected ($T_1$), Class A was randomly assigned to sit on the therapy balls between $T_1$ and $T_2$, while Class B sat on normal classroom chairs and served as the control. We then switched the conditions between $T_2$ and $T_3$ so that Class B sat on the balls while Class A served as the control. This second phase ($T_2$-$T_3$) allowed us to determine if there were any long-term benefits after the students returned to normal classroom chairs and to determine if the benefits observed in Class A ($T_1$-$T_2$) could be replicated in Class B ($T_2$-$T_3$).

Prior to the introduction of therapy balls, a certified fitness instructor showed the students how to properly sit on the balls (i.e., both feet flat on the floor with backs in a straight and upright posture position). In addition, students in the experimental conditions completed a post-treatment questionnaire designed to assess their seating enjoyment and preference.

*Measures*

*Independent Observations of Undesirable Behavior*

Prior to the study, the teachers and experimenters developed a list of 14 undesirable behaviors considered detrimental to the learning process such as "*not participating in choral activities*" and "*not sitting in the WOW position*"[-1] (see list of behaviors in Table 1). Five undergraduate research assistants from Florida Southern College served as independent observers in the study. As part of their training, the observers sat in each classroom for 45 minutes across three days in order to acclimate the students to the their presence, reduce demand effects (e.g., Steele-Johnson, 2000), and familiarize themselves with the environments and the 14 behaviors of interest. During the study, the observers conducted naturalistic observations in each classroom by recording the frequency with which the students engaged in any of the 14 undesirable behaviors. They observed behavior five times (each 35 minutes long) in both classrooms across each of the three time waves (for a total of 15 observation sessions per class). Two research assistants observed in each classroom in order to establish inter-rater reliability and they observed in both classrooms at the same time of day in order to control for time differences. Inter-rater reliability was high in both Class A (.79) and Class B (.74), indicating that the observers were consistent with each other in their observations within the classrooms.

*Teacher Reports of Desirable Behavior*

In each time wave, the teachers completed a 15-item questionnaire designed to measure the percentage of time that the teachers observed each student engaging in desirable classroom behaviors (e.g., *looking at the teacher when appropriate*) (see list of behaviors in Table 3). Using a 5-point Likert scale, from 1 (*never or 0% of the time*) to 5 (*always or 100% of the time)*, the teachers were asked, "How often on average in the last time wave has the child exhibited the following behaviors?"

*Student Self-Report of Desirable Behavior*

In each time wave, the students completed a 15-item self-assessment of behavior (see Table 5). This questionnaire was identical in content (using the same Likert scale) as the teacher observation questionnaire except that it was designed to measure the percentage of time that each student believed that they engaged in desirable classroom behaviors (e.g., "How often on average in the last time wave did I look at the teacher when appropriate?"). We also included two questions designed to assess each student's internal attention level (e.g., "My mind wandered during class") and motivation level (e.g., "I felt motivated to complete my work to the best of my ability"). These two questions were averaged into a composite attention and motivation score.

*Student Seating Enjoyment, Focus, and Preference*

Following each treatment phase, the students in the experimental group completed a 3-item questionnaire designed to assess whether they enjoyed / liked using the therapy balls, whether it increased their focus, and whether they found the use of balls by their classmates distracting (reverse scored). The questionnaire utilized a 5-point Likert scale, from 1 (*not at all*) to 5 (*extremely*). An additional yes/no item assessed seating preference (e.g., "Would you like to continue using the therapy balls instead of classroom chairs?").

*Reading Comprehension*

In each time wave, students in both classes completed three elementary grade level (first through fourth grade) reading comprehensions tests as part of their normal curriculum. These tests consisted of 12 to 15 sequencing, multiple choice, and true/false questions from the source, "*Reading Comprehension in Varied Subject Matter*" (Ervin, 1997). The tests assessed ability to comprehend the information (e.g., identify the main idea) from one to two single spaced passages on various subjects (e.g., making the world flat). Students were given adequate time to read the passage and were then tested directly after.

## *Results*

We transformed all of the data collected across all of the time wave observation periods ($T_1$- $T_2$ and $T_2$- $T_3$) into difference scores by subtracting the previous observation period data from the subsequent observation period data (e.g., $T_2$ minus $T_1$). Therefore, positive scores indicate an increase in that measure (e.g., behavior, attention, motivation, etc.) from the previous

observation phase (from either the independent observers', teachers' or students' perspective), and negative scores indicate a decrease in that measure.

There were no significant three-way interactions between diagnosis (dyslexia only, dyslexia and ADHD), time wave ($T_1$, $T_2$, $T_3$), and class (Class A, Class B) on any of the measures. In other words, there were no significant differences between children with dyslexia only and children with a dual diagnosis of dyslexia and ADHD on any of the measures we collected. Therefore, we present the results of the data analysis without diagnosis as a variable.

*Independent Observations of Undesirable Behavior*

The mean differences in observations of behavior, results of the planned comparisons, and effect sizes (*Cohen's d*) across all 14 behavioral indices from the $T_1$ to $T_2$ time waves are shown in Table 1 and time waves $T_2$ to $T_3$ are shown in Table 2. We computed a composite measure of undesirable behavior by summing the total number of undesirable behaviors that the research assistants observed across 14 individual behavioral dimensions. Data from all five observations sessions per time period in each class were included in a 2 x 3 repeated measures factorial ANOVA with time wave of observations of composite undesirable behaviors ($T_1$, $T_2$, $T_3$) as the within-subjects factor and Class (A, B) as the between subjects factor. There was a significant interaction between time wave and Class, $F$ (2, 16) = 12.81, $p$ = .000, $\eta_p^2$ = .62. Further analysis of this two-way interaction (see below) revealed that the assistants observed a significant decrease is undesirable behaviors overall (from baseline to treatment) in both classes (see last row of Tables 1 and 2).

*Planned Comparisons: Independent Observations of Undesirable Behavior, T1 to T2*

As shown in the composite score row at the bottom of Table 1, planned paired sample t-tests revealed that from $T_1$ to $T_2$, the research assistants observed significantly fewer undesirable classroom behaviors in the treatment class (Class A) overall (*Mean difference* = -12.00). A Cohen's *d* effect size of 2.83 indicated high practical significance of this reduction. Specifically, there was a significant reduction in the following four undesirable behaviors (*ps* < .05): (1) *looking away from teacher*, (2) *looking away from material*, (3) *fidgeting*, and (4) *displaying a negative attitude*. Cohen's *d* effect size values ranged from .64 to 4.95 for these four undesirable behaviors, suggesting moderate to high practical significance of these reductions. There was also a moderate reduction in *talking off topic to classmates* in the treatment condition that approached significance ($p$ = .080, $d$ = .71). Of the remaining nine non-significant comparisons (*ps* > .05), there was a moderate to large reduction (*ds* > .50) from $T_1$ to $T_2$ in five of the undesirable behaviors.

In contrast, the assistants did not observe significantly fewer undesirable behaviors in the control class (Class B, $T_1$ to $T_2$) overall (*Mean difference* = -2.17, $p$ = .860, $d$ = .11) (see bottom row of Table 1). Specifically, there was only a significant reduction in *looking away from teacher* and *responding inappropriately (ps* < .05*)*. Cohen's *d* effect size values were 3.25 and 1.61 respectively, suggesting high practical significance of these two reductions. None of the other comparisons approached significance. Despite a reduction in two of the undesirable behaviors in

the control, there was a significant increase in *complaining* (with a large effect size of 2.56), and all other comparisons were not significant (*ps* > .05). Of the remaining 11 non-significant comparisons, there was a moderate to large increase (*ds* > .50) from $T_1$ to $T_2$ in four of the undesirable behaviors in the control condition.

*Planned Comparisons: Independent Observations of Undesirable Behaviors T2 to T3*

As shown in the composite score at the bottom of Table 2, the research assistants also observed significantly fewer undesirable behaviors overall from $T_2$ to $T_3$ in the treatment class (Class B) ($M$ = -34.90, $p$ = .037), and this reduction was large in size ($d$ = 1.83). Specifically, there was a significant reduction in the following five undesirable behaviors in the treatment condition (*ps* < .05): (1) *responding inappropriately,* (2) *failing to raise hand,* (3) *getting out of seat,* (4) *fidgeting*, and (5) *complaining*. Cohen's *d* effect size values (.94 to 4.24) indicated large reductions in these five undesirable behaviors. There was also a large reduction in *talking off topic to classmates* in the treatment condition that approached significance ($p$ = .068, $d$ = 1.41). All other comparisons in the treatment condition were not significant (*ps* > .05). However, of the remaining 10 non-significant comparisons, there was a moderate to large reduction (*ds* > .50) from $T_2$ to $T_3$ in two of the undesirable behaviors in the treatment condition.

In contrast, they observed significantly more undesirable behaviors overall in the control class (Class A) ($M$ = 14.20, $p$ = .029), and this increase was large in size ($d$ = 1.80) (see bottom row of Table 2). Specifically, there was a significant increase in the undesirable behavior of *not sitting in the WOW position* (with a large effect size of 2.30). There was also an increase in *fidgeting* and *talking off topic to classmates* that approached significance ($p$ = .080 and .053, respectively) and Cohen's *d* effect size values of .94 and 1.68 (respectively) suggested large increases in these two undesirable behaviors. Of the 11 remaining non-significant comparisons (*ps* > .05), there was 0% change in four of the behaviors and there was a moderate to large increase (*ds* > .50) from $T_2$ to $T_3$ in three of the undesirable behaviors in the control.

*Teacher Reports of Desirable Behavior*

We computed a composite measure of teachers' reports of the percentage of time during the observation period that individual students exhibited desirable behaviors by averaging the teachers' rating of the 15 individual behavioral dimensions shown in Table 3. The average teacher ratings, results of the planned comparisons, and effect sizes (*Cohen's d*) across all 15 behavioral indices are shown in Table 3 ($T_1$ to $T_2$ times waves) and Table 4 ($T_2$ to $T_3$ time waves). A 2 x 3 repeated measures factorial ANOVA with time wave of teachers' composite reports of desirable behaviors ($T_1$, $T_2$, $T_3$) as the within-subjects factor and Class (A, B) as the between subjects factor, revealed a significant interaction between time wave and Class, $F$ (2, 44) = 6.45, $p$ = .003, $\eta_p^2$ = .23 (see last row of Table 3 and 4). Further analysis of this two-way interaction (see below) revealed that teachers reported a significant increase in desirable behaviors overall (from baseline to treatment) for both classes (see last row of Table 3 and 4).

*Planned Comparisons: Teacher Reports of Desirable Behaviors, T1 to T2*

As shown in the composite score in the bottom row of Table 3, planned paired sample t-tests revealed that there was a moderate, significant increase of 9.00% ($d$ = .68) in the teacher's composite reports of the percentage of time that students exhibited desirable classroom behaviors for the treatment class (Class A) between $T_1$ and $T_2$. Specifically, there was a significant increase in the following eight desirable behaviors ($ps$ < .05): (1) *sitting in the WOW position*, (2) *raising hand*, (3) *participating in choral activities*, (4) *staying in seat*, (5) *participating in conversations on topic*, (6) *talking on topic to classmate*, (7) *keeping hands to oneself*, and (8) *not complaining*. For seven of these eight behaviors, the Cohen's $d$ effect size values ranged from .62 to 1.20, suggesting moderate to high practical significance. There was also an increase in the following four behaviors that approached significance ($ps$ < .09): (1) *looking at teacher*, (2) *looking at materials*, (3) *completing assignments on time*, and (4) *responding appropriately*. Two of these increases were of moderate effect size ($ds$ > .50). Of the remaining four non-significant comparisons ($ps$ > .05), there was a moderate increase from $T_1$ to $T_2$ in *completing assignments on time* in the treatment condition ($d$ = .51).

In contrast, there was no significant difference in teacher reports of desirable behaviors between $T_1$ to $T_2$ in the control class (Class B) overall ($M$ = -.71%, $p$ = .613, $d$ = .06) (see bottom row, Table 3). Specifically, there was a moderate, significant increase in *looking at the teacher* ($p$ = .039, $d$ = .69) and a moderate, non-significant increase in *looking at materials* ($p$ = .096, $d$ = .65). However, the behavior of *talking on topic to classmate* decreased significantly ($p$ = .039, $d$ = .46) and there was a moderate decrease in *sitting in the WOW position* that approached significance ($p$ = .104, $d$ = .69). Furthermore, all 11 of the remaining comparisons in the control condition from $T_1$ to $T_2$ were not significant and had small effect sizes ($ds$ < .35).

*Planned Comparisons: Teacher Reports of Desirable Behaviors, T2 to T3*

As shown in the bottom row of Table 4, there was a significant increase of 4.25% in teacher's composite reports of percentage of time that student exhibited desirable behaviors between $T_2$ and $T_3$ for the treatment class (Class B), but this effect was small in size ($d$ = .31). Specifically, there was a significant increase in the following three desirable behaviors: (1) *displaying a positive attitude*, (2) *talking on topic to classmate*, and (3) *keeping hands to oneself.* The Cohen's $d$ effect size values ranged from .40 to .43, suggesting small practical significance. There was also a small increase in *not complaining* in the treatment condition that approached significance ($p$ = .082, $d$ = .24). All of the other comparisons in the treatment class from $T_2$ to $T_3$ were not significant ($ps$ > .05), with small effect sizes ($ds$ < .35). However, of the remaining 11 non-significant comparisons, there was a moderate increase from $T_2$ to $T_3$ in *sitting in the WOW position* in the treatment condition ($d$ = .58).

In contrast, there was no significant difference in teacher reports of desirable behaviors between $T_2$ and $T_3$ in the control class (Class A) overall ($M$ = .04%, $p$ = .987, $d$ = .00). Specifically, although there was a moderate, significant increase in *not fidgeting* ($p$ = .017, $d$ = .74), there were either decreases or no changes in the remaining behaviors. For example, there was a moderate decrease in *participating in conversations on topic* that approached significance ($p$ =

.082, $d = .55$) and although not significant, there was also a moderate decrease in *talking on topic to classmates* ($p = .137$, $d = .51$). All 11 of the remaining comparisons in the control class were not significant ($ps > .05$), and had small effect sizes ($ds < .42$).

*Student Self-Report of Composite Desirable Behaviors*

We computed a composite measure of student self-reports of the percentage of time during the observation period that they exhibited desirable behaviors by averaging across the 15 individual behavioral dimensions shown in Table 5. The average student ratings, results of the planned comparisons, and effect sizes (*Cohen's d*) across all 15 behavioral indices are shown in Table 5 ($T_1$ to $T_2$ time waves) and Table 6 ($T_2$ to $T_3$ time waves). We conducted a 2 x 3 repeated measures factorial ANOVA with time wave of students self-reports of composite desirable behaviors ($T_1$, $T_2$, $T_3$) as the within-subjects factor and Class (A, B) as the between subjects factor. There was interaction between time wave and Class that approached significance, $F(2, 44) = 2.37$, $p = .11$, $\eta_p^2 = .097$. Further analysis (see below) revealed that there was only a significant increase in students' self-report of desirable behavior in the treatment condition (Class A) between $T_1$ and $T_2$.

*Student Self-Report of Desirable Behaviors, T1 to T2*

As shown in the bottom row of Table 5, planned paired sample t-tests revealed that in the treatment class (Class A) there was a significant increase of 7.92% in the students' self-reports of the percentage of time that they exhibited desirable classroom behaviors between $T_1$ and $T_2$. The Cohen's *d* effect size value of .93 suggested high practical significance of this increase in student self report. Specifically, there was a significant increase in the following four desirable behaviors: (1) *talking on topic to classmate,* (2) *keeping hands to oneself*, (3) *not complaining*, and (4) *asking related questions.* Cohen's *d* effect size values ranged from 1.41 to 2.33, suggesting large increases in these four behaviors. There was a large, significant reduction in *raising hands* ($p = .004$, $d = 1.41$) and a large reduction in *participating in choral activities* that approached significance ($p = .089$, $d = .87$). However, all nine of the remaining comparisons in the treatment condition were not significant ($ps > .05$), and eight of them had small effect sizes ($ds < .47$).

In contrast, there was no significant difference in student self-reports of desirable behaviors between $T_1$ and $T_2$ in the control class (Class B) overall ($M = 5.27\%$, $p = .142$, $d = .56$). There were large, significant increases in (1) *completing assignments on time*, (2) *talking on topic to classmate*, (3) *keeping hands to oneself*, and (4) *not complaining* ($ds = .77$ to 1.44). However, there was a large, significant decrement in *not fidgeting* ($p = .017$, $d = 1.46$), and a large, decrement in *displaying a positive attitude* that approached significance ($p = .056$, $d = .92$). All nine other comparisons in the control condition between $T_1$ and $T_2$ were not significant ($ps > .05$), and six had small effect sizes ($ds < .42$).

*Student Self-Report of Desirable Behaviors, T2 to T3*

As shown in the bottom row of Table 6, there was a small, non-significant increase of 5.15% (*p*

= .159, $d$ = .46) between $T_2$ and $T_3$ in students' composite self-reports of desirable behaviors in the treatment class (Class B) overall. Although there were moderate, significant reductions in *responding appropriately* and *not fidgeting* ($p$ = .034, $d$ = .69 and $p$ = .044, $d$ = .66, respectively), all of the other 13 comparisons in the treatment condition between $T_2$ and $T_3$ were not significant ($ps$ > .05) and 10 had small effect sizes ($ds$ < .48).

In contrast, there was a moderate decrease of -4.73% ($d$ = .56) in self-reports of desirable behaviors in the control class (Class A) between $T_2$ and $T_3$ that approached significance ($p$ = .09). Specifically, there was a significant decrement in *sitting in the WOW position* and *displaying a positive attitude towards learning*. The Cohen's $d$ effect size values of .88 and 1.16 (respectively) suggested high practical significance of these reductions. Although, there was a moderate increase in *raising hand when responding* that approached significance ($p$ = .05, $d$ = .66), all of the other 13 comparisons in the control condition between $T_2$ and $T_3$ were not significant ($ps$ > .05) and 11 had small effect sizes ($ds$ < .49).

*Student Self-Report of Attention and Motivation*

We computed a composite measure of student self-report of the percentage of time during the observation period that they were attentive and motivated by averaging across the two questions that assessed attention and motivation. A 2 x 3 repeated measures factorial ANOVA with time wave of student self-report of composite attention and motivation ($T_1$, $T_2$, $T_3$) as the within-subjects factor and Class (A, B) as the between subjects factor, revealed a significant interaction between time wave and Class, $F$ (2, 44) = 7.59, $p$ = .001, $\eta_p^2$ = .26.

Planned comparisons revealed that between $T_1$ and $T_2$ there was a significant increase of 13.54% in student self-reports of attention and motivation in the treatment condition (Class A) from $T_1$ ($M$ = 53.65%, $SD$ = 7.75) to $T_2$ ($M$ = 67.19%, $SD$ = 11.03), $t$(11) = -2.60, $p$ = .025. A Cohen's $d$ effect size value of 1.42 suggested a high practical increase in attention and motivation in the treatment condition. In contrast, there was a significant, large decrement in attention and motivation (-13.02%) between $T_1$ ($M$ = 65.12%, $SD$ = 11.45) and $T_2$ ($M$ = 52.08%, $SD$ = 16.50) in the control condition (Class B), $t$(11) = 2.23, $p$ = .047, $d$ = .92.

There was less of an impact of treatment on students' self-reports of attention and motivation between $T_2$ and $T_3$. In the control condition (Class A, $T_2$ to $T_3$), there was no change in attention and motivation (0%) between $T_2$ and $T_3$ ($M$ = 67.19%, $SD$ = 11.96), $t$(11) = .00, $p$ = 1.00, $d$ = .00. In contrast, there was an increase of 6.25% in student self-reports of attention and motivation between $T_2$ and $T_3$ ($M$ = 65.12%, $SD$ = 11.45) in the treatment class (Class B), $t$(11) = -1.27, $p$ = .230. Although this increase in attention and motivation was not significant, a *Cohen's d* value of .92 suggested that this increase in the treatment condition was of large practical significance.

*Student Seating Enjoyment and Focus*

On a 5 point Likert scale where 1 was the "*least enjoyable / focused – attentive*" and 5 was the "*most enjoyable / focused-attentive*," students reported an average enjoyment level of 3.96 ($SD$ =

1.12) and an average focus - attentive level of 3.88 (*SD* = 1.30) while sitting on the therapy balls. On a 5 point Likert scale, where 1 was "*strongly disliked the therapy balls*" and 5 was "*strongly liked the therapy balls*," students reported an average likeability rating level of 3.71(*SD* = .85). Both classes reported high positive reactions to the use of the therapy balls because a series of between subject ANOVAs revealed that neither enjoyment ratings, $F < 1$, nor focus - attentive ratings, $F(1, 22) = 3.30$, $p = .080$, $\eta_p^2 = .13$, nor likeability ratings differed significantly as a function of Class, $F < 1$.

*Student Seating Preference Frequencies*

According to a chi-square analysis, significantly more students reported a desire to continue using the therapy balls ($n = 19$) than students who reported a desire to return to their regular classroom chairs ($n = 5$), $\chi^2 = 8.17$, $p = .004$. However, unlike the ratings of enjoyment, preference frequencies did vary as a function of Class: significantly more Class A students indicated a desire to continue using the balls ($n = 10$) than students who wanted to return to their normal chairs ($n = 2$), $\chi^2 = 5.33$, $p = .021$, but in Class B this preference for the therapy balls only approached significance, (9 vs. 3), $\chi^2 = 3.00$, $p = .083$.

*Reading Comparison*

A 2 x 3 repeated measures factorial ANOVA with time wave of reading comprehension scores ($T_1$, $T_2$, $T_3$) as the within-subjects factor and Class (A, B) as the between subjects factor revealed no significant interaction between time wave and Class, $F < 1$. Paired sample t-tests also revealed no significant improvement in reading comprehension scores between $T_1$ ($M = 90.02$, $SD = 7.93$) and $T_2$ ($M = 86.08$, $SD = 8.39$) in the treatment class (Class A), $t(11) = 1.92$, $p = .08$, $d = .48$, or between $T_1$ ($M = 86.61$, $SD = 9.28$) and $T_2$ ($M = 83.96$, $SD = 14.83$) in the control class (Class B), $t(11) = .68$, $p = .510$, $d = .21$. There was also no significant improvement in reading comprehension scores between $T_2$ and $T_3$ ($M = 82.08$, $SD = 14.47$) in the treatment class (Class B), $t(11) = .41$, $p = .690$, $d = .13$, or between $T_2$ and $T_3$ ($M = 83.55$, $SD = 8.32$) in the control class (Class A), $t(11) = 1.10$, $p = .291$, $d = .30$.

## Discussion

With the exception of the students' self-ratings of behavior, the behavioral results and the students' attention and motivation reports and seating preferences support the use of therapy balls in the classroom. The results of both the independent observations and the teacher reports suggest that the therapy balls significantly improved behavior in both classrooms. Importantly, by replicating the overall results in Class B, we provide evidence that the benefits were not class, student, or teacher specific. These results are consistent with previous research showing a behavioral benefit from the use of therapy balls in the classroom on children with attention issues (e.g., Schilling et al., 2003). The behavioral benefits in our study did not depend on the presence of ADHD, for students with dyslexia only, yielded the same improvements as those who also had an ADHD diagnosis. Thus, our results provide evidence that the behavioral benefit found in other studies on children with ADHD or ASD also generalizes to children with dyslexia only as well as those with a dual diagnosis of dyslexia and ADHD.

Although both classes benefited from the use of therapy balls, each class exhibited a unique

pattern of improvement. However, the differential benefits we found as a function of class are most likely due to different levels of specific behaviors during baseline. A review of the research assistants' data, for example, revealed that a class failed to exhibit a decrease in a specific undesirable behavior because they either exhibited that problem behavior at a very low rate or not at all during baseline (i.e., a floor effect). For example, Class A exhibited *complaining about school work* at a very low rate during baseline ($M$ = .33, $SD$ = .41), while Class B exhibited this behavior at a higher rate ($M$ = 1.50, $SD$ = .50). This meant that although both Class A and B exhibited reductions in *complaining about school work* post-treatment (*mean difference* = -.13 and -1.50, respectively), only Class B exhibited a statistically significant reduction because Class A's baseline was near the floor. This floor effect left relatively little quantitative "room" for a reduction in *complaining* behavior post-treatment in Class A. In short, there was only a significant reduction in the undesirable behaviors that appeared to be problematic for the class (A or B) at baseline.

Similarly, our failure to find a significant improvement in some of the teacher's report of a specific behavior was the result of a ceiling effect or a high reported rate of that behavior during baseline, that left little statistical room for improvement of that behavior during treatment. In support of this explanation, there was only a significant increase in the desirable behaviors that the teachers reported at a lower rate during baseline. Overall, the results of the independent observations and teacher reports, suggest that therapy balls can be used to modify any number of different behaviors within the context of a classroom.

Students' subjective reports of behavior paralleled those of the teachers and the research assistants, but to a much lesser degree and with not as much consistency. However there are reasons to question the validity of the student data. For example, the teachers had to read some of the items on the student self-report questionnaire to the students because they had difficulties reading and understanding some of them. Because the students' reading comprehension levels ranged from first to fourth grade, it was difficult to construct a questionnaire that not only mirrored the behaviors assessed on the teachers' report, but that also matched each student's reading level. It is therefore unclear whether our null effect with regard to students' self report was in fact credible (i.e., the students did not perceive an improvement in their behavior) or whether the results reflect a lack of understanding of the questions. In addition to misunderstanding some items, the students had relatively little experience answering Likert scaled items. Future research should include a manipulation check to ensure that the students truly understand the questions and student training on how to answer Likert scaled items. In addition to improving the validity of the student questionnaires, we propose adding parental screenings of behavior (Reddington & Wheeldon, 2002).

Although students' subjective reports of behavior failed to completely mirror the observations of the research assistants or the teachers, the students reported high levels of enjoyment and focus while seated on the balls. The majority of students also clearly preferred the balls to normal classroom seats. Furthermore, the students in both classes also indicated improvements in attention and motivation. Thus the majority of the present findings support the use of therapy balls in the classroom.

While the behavioral, attention, and motivation results of the experiment could be due to maturation (i.e., students growing older and behaving better as the experiment goes on), we can assume that both classes matured at the same rate. If maturation was responsible for improvements, then we would not have found improved behavior during treatment compared to control in both classes. Another argument against maturation as an explanation is the fact that the benefits of the balls were short-lived. After the balls were removed from Class A, the research assistants observed significantly more undesirable behaviors and the teachers reported decreases or no change in 14 desirable behaviors. If maturation was responsible for improvements, there would have been behavioral improvement in absence of the balls in the later time waves. Therefore, it is unlikely that the results are simply due to maturation.

We failed to find an impact of therapy balls on reading comprehension. This null learning result is inconsistent with the previous research showing a learning benefit in the form of legible word productivity (Schilling et al., 2003) and listening comprehension (Kercood & Banda, 2012). Therefore the learning benefit found in other studies may not generalize to reading comprehension skills. The high average score on the reading comprehension tests overall ($M$ = 85.38, $SD$ = 10.54) may account for our failure to find any significant improvement. Future studies might increase the number of reading comprehension measures (e.g., fluency or reading speed) (Cutting, Materek, Cole, Levine & Mahone, 2009) in order to more accurately determine the effects of therapy balls on reading comprehension as a learning outcome.

Poor motor coordination paired with abnormal eye movements might also explain why reading comprehension scores did not improve. Magnocellular theory lends support for the idea that the area in the cerebellum that controls reading deficits also controls difficulties in motor coordination (Benassi, Simonelli, Giovagnoli & Bolzani, 2010), motion coherence (Cornelissen, Richardson, Mason & Fowler, 1995), and mental rotation (Kershner, 1979). Researchers have also shown that dyslexics exhibit impairments if balancing is paired with another task (Kaltner & Jansen, 2014; Brookes, Tinkler, Nicolson, & Fawcett, 2010). Dyslexics also tend to have problems with fast attention shifts, symptoms of unilateral neglect syndrome and abnormal eye movements, which have all been linked to the parietal lobe (Jaskowski & Rusiak, 2005; Facoetti et al., 2005). Therefore, poor motor coordination, inhibited motion coherence, and abnormal eye movements could explain why reading comprehension did not improve while on the ball because the students may have found it difficult to read and balance simultaneously.

Overall our results support the use of therapy balls in the classroom for improving the attention, motivation, and behavior of children with dyslexia. Our results are noteworthy given that we observed benefits after only five days of treatment. However, it is important to note that not every child enjoyed or preferred sitting on the ball. Thus, educators need to evaluate the individual needs of their students when determining whether the balls should be used a sensory modulation technique.

## References

*American Psychiatric Association (2003) Diagnostics and Statistical Manual of Mental Disorders* (5th Ed.). American Psychiatric Publishing, Washington DC.

Benassi, M., Simonelli, L., Giovagnoli, S., & Bolzani, R. (2010). Coherence motion perception in developmental dyslexia: A meta-analysis of behavioral studies. *Dyslexia*, *16*(4), 341-357.

Brookes, R. L., Tinkler, S., Nicolson, R. I., & Fawcett, A. J. (2010). Striking the right balance: Motor difficulties in children and adults with dyslexia. *Dyslexia*, *16*(4), 358-373.

Cornelissen, P., Richardson, A., Mason, A., & Fowler, S. (1995). Contrast sensitivity and coherent motion detection measured at photopic luminance levels in dyslexics and controls. *Vision Research*, *35*(10), 1483-1494.

Cutting, L. E., Materek, A., Cole, C. S., Levine, T. M., & Mahone, E. (2009). Effects of fluency, oral language, and executive function on reading comprehension performance. *Annals of Dyslexia*, *59*(1), 34-54.

Dahle, A., Knivsberg, A., & Andreassen, A. (2011). Coexisting problem behaviour in severe dyslexia. *Journal of Research in Special Educational Needs, 11*(3), 162-170.

Dhar, M., Been, P. H., Minderaa, R. B., & Althaus, M. (2010). Information processing differences and similarities in adults with dyslexia and adults with attention deficit hyperactivity disorder during a continuous performance test: A study of cortical potentials. *Neuropsychologia*, *48*(10), 3045-3056.

Donfrancesco, R., Mugnaini, D., & Dell'uomo, A. (2005). Cognitive impulsivity in specific learning disabilities. *European Child & Adolescent Psychiatry*, *14*(5), 270-275.

Ervin, J. (1997). Reading comprehension in varied subject matter. Cambridge, MA: Educators Pub. Service.

Facoetti, A., Lorusso, M., Cattaneo, C., Galli, R., & Molteni, M. (2005). Visual and auditory attentional capture are both sluggish in children with developmental dyslexia. *Acta Neurobiologiae Experimentalis*, *65*(1), 61-72.

Fedewa, A. L., & Erwin, H. E. (2011). Stability balls and students with attention and hyperactivity concerns: Implications for on-task and in-seat behavior. *American Journal of Occupational Therapy*, *65*(4), 393-399.

Germanò, E., Gagliano, A., & Curatolo, P. (2010). Comorbidity of ADHD and dyslexia. *Developmental Neuropsychology*, *35*(5), 475-493.

Helland, T. (2007). Dyslexia at a behavioural and a cognitive level. *Dyslexia: An International Journal of Research and Practice*, *13*(1), 25-41.

Jaskowski, P., & Rusiak, P. (2005). Posterior parietal cortex and developmental dyslexia. *Acta Neurobiologiae Experimentalis*, *65*(1), 79-94.

Jensen, P. S., & Kenny, D. T. (2004). The effects of yoga on the attention and behavior of boys with Attention-Deficit/hyperactivity Disorder (ADHD). *Journal of Attention Disorders, 7*(4), 205-216.

Kaltner, S., & Jansen, P. (2014). Mental rotation and motor performance in children with developmental dyslexia. *Research in Developmental Disabilities*, *35*(3), 741-754.

Karande, S., Satam, N., Kulkarni, M., Sholapurwala, R., Chitre, A., & Shah, N. (2007). Clinical and psychoeducational profile of children with specific learning disability and co-occurring attention deficit hyperactivity disorder. *Indian Journal of Medical Sciences, 61*(12), 639-647.

Kavale, K., & Forness, S. (1996). Social skill deficits and learning disabilities: a meta-analysis. *Journal of Learning Disabilities, 29*(3), 226-237.

Kercood, S., & Banda, D. R. (2012). The effects of added physical activity on performance during a listening comprehension task for students with and without attention problems. *International Journal of Applied Educational Studies*, *13*(1), 19-32.

Kercood, S., Grskovic, J. A., Lee, D. L., & Emmert, S. (2007). The effects of fine motor movement and tactile stimulation on the math problem solving of students with attention problems. *Journal of Behavioral Education, 16*(4), 303-310.

Kershner, J. R. (1979). Rotation of mental images and asymmetries in word recognition in disabled readers. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, *33*(1), 39-50.

Mulligan, S. (2001). Classroom strategies used by teachers of students with attention deficit hyperactivity disorder. *Physical & Occupational Therapy in Pediatrics*, *20*(4), 25-44.

Reddington, J., & Wheeldon, A. (2002). Involving parents in baseline assessment: Employing developmental psychopathology in the early identification process. *Dyslexia*, *8*(2), 119-122.

Robertson, E. K., & Joanisse, M. F. (2010). Spoken sentence comprehension in children with dyslexia and language impairment: The roles of syntax and working memory. *Applied Psycholinguistics, 31*(1), 141-165.

Routh, D. K. (1979) Activity, attention, and aggression in learning disabled children. *Journal of Clinical Child Psychology, 8*(3), 183-187.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.

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

Schilling, D., Washington, K., Billingsley, F. F., & Deitz, J. (2003). Classroom seating for children with attention deficit hyperactivity disorder: Therapy balls versus chairs. *American Journal of Occupational Therapy*, *57*(5), 534-541.

Steele-Johnson, D. (2000). Goal orientation and task demand effects on motivation, affect, and performance. *The Journal of Applied Psychology*, *85*(5), 724–738.

Watling, R., Deitz, J., Kanny, E., & McLaughlin, J. (1999). Current practice of occupational therapy for children with autism. *American Journal of Occupational Therapy, 53,* 498-505.

Willcutt, E. G., Betjemann, R. S., McGrath, L. M., Chhabildas, N. A., Olson, R. K., DeFries, J. C., & Pennington, B. F. (2010). Etiology and neuropsychology of comorbitiy between RD and ADHD: The case for multiple-deficit models. *Cortex*, *46*(10), 1345-1361.

Willcutt, E.G. & Pennington, B.F (2000a). Comorbidity of reading disability and attention deficit/hyperactivity disorder: differences by gender and subtype. *Journal of Learning Disabilities, 33*(2), 179-191.

Willcutt, E.G., Pennington, B.F & DeFries, J.C., (2000b). Etiology of inattention and hyperactivity/impulsivity in a community sample of twins with learning difficulties. *Journal of Abnormal Child Psychology, 28*(2),149-159.

Zentall, S. S. (2007). Math performance of students with ADHD: Cognitive and behavioral contributors and interventions. *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 219-243). Baltimore, MD: Paul H Brookes Publishing.

*Footnote*

[-1.] WOW stands for "watch our writing" and includes the five following components: 1. Feet placed flat on the floor, 2. Back positioned straight but with slight arch at the top leaning toward paper, 3. Proper placement of writing hand and the "bossy" hand to guide the paper upright as one moves down the paper so as to keep the writing arm on the table instead of off the table, 4. Paper placed in correct direction, and, 5. Gripping pencil.

Table 1.
*Independent observations of the differences in average number of undesirable behaviors from $T_1$ to $T_2$.*

| Undesirable Behavior | Class A | | | | Class B | | | |
|---|---|---|---|---|---|---|---|---|
| | *Mean* Difference (*SD*) | *t* (4) | *p* | *d* | *Mean* Difference (*SD*) | *t* (4) | *p* | *d* |
| Looking away from teacher when inappropriate | **-1.87(0.61)** | **6.89**** | **.002** | **2.25** | -8.33 (1.96) | 9.52* | .001 | 3.25 |
| Looking away from materials when inappropriate | **-2.13 (0.73)** | **6.53**** | **.003** | **4.95** | 1.33 (2.34) | -1.27 | .273 | 0.68 |
| Responding inappropriately | **-1.20 (1.30)** | **2.06** | **.109** | **1.54** | -1.92 (.72) | 5.94* | .004 | 1.61 |
| Not sitting in WOW position | **-1.40 (2.61)** | **1.20** | **.296** | **0.96** | 0.17 (4.71) | -0.08 | .941 | 0.05 |
| Failing to raise hand when responding | **-0.53 (0.93)** | **1.28** | **.269** | **0.55** | 1.92 (4.93) | -0.87 | .434 | 0.58 |
| Silent during choral activities | **-0.67 (0.82)** | **1.83** | **.142** | **1.16** | 1.08 (1.48) | -1.64 | .177 | 0.78 |
| Getting out of seat when inappropriate | **0.20 (0.45)** | **-1.00** | **.374** | **0.63** | -0.08 (2.20) | 0.08 | .937 | 0.06 |
| Not participating in conversations on topic | **-0.13 (0.61)** | **0.49** | **.648** | **0.30** | 0.25 (2.86) | -0.20 | .855 | 0.09 |
| Fidgeting | **-2.20 (1.10)** | **4.49*** | **.011** | **0.64** | 0.83 (1.26) | -1.48 | .213 | 0.24 |
| Talking off topic to classmate | **-0.47 (0.45)** | **2.33** | **.080** | **0.71** | 0.75 (11.79) | -0.14 | .894 | 0.10 |
| Putting hands on classmate | **0.00 (0.00)** | **0.00** | **1.00** | **0.00** | -0.50 (1.50) | 0.75 | .497 | 0.63 |
| Failing to complete assignments on time | **-0.33 (0.41)** | **1.83** | **.142** | **1.14** | 0.50 (.50) | -2.24 | .089 | 1.41 |
| Complaining about work | **-0.13 (0.18)** | **1.63** | **.178** | **0.30** | 1.17 (.50) | -5.22* | .006 | 2.56 |
| Displaying negative attitude towards learning | **-1.13 (0.73)** | **3.47*** | **.026** | **2.63** | 0.67 (1.47) | -1.01 | .368 | 0.49 |
| Composite negative behavior score | **-12.00 (6.78)** | **3.96*** | **.017** | **2.83** | -2.17 (25.22) | 0.19 | .860 | 0.11 |

Note. Bold indicates treatment phase. Class A received the therapy balls treatment during the $T_1$ to $T_2$ period. Difference scores were calculated by subtracting the $T_1$ observation period data from the $T_2$ observation data ($T_2 - T_1$). Therefore, positive scores indicate an increase in observations of undesirable behavior from $T_1$ to $T_2$, and negative scores indicate a decrease. Composite negative behavior scores were calculated by summing all of the 14 undesirable behaviors that were observed. *indicates a significant difference at the .05 level and **indicates a significant difference at the .01 level.

Table 2.
*Independent observations of the differences in average number of undesirable behaviors from $T_2$ to $T_3$.*

| | Class A | | | | Class B | | | |
|---|---|---|---|---|---|---|---|---|
| Undesirable Behavior | *Mean* Difference (*SD*) | *t* (4) | *p* | *d* | *Mean* Difference (*SD*) | *t* (4) | *p* | *d* |
| Looking away from teacher when inappropriate | 0.00 (1.00) | 0.00 | 1.00 | 0.00 | **-0.20 (2.28)** | **0.20** | **.854** | **0.15** |
| Looking away from materials when inappropriate | 1.20 (1.30) | -2.06 | .109 | 1.38 | **-0.80 (1.30)** | **1.37** | **.242** | **0.37** |
| Responding inappropriately | 0.40 (.55) | -1.63 | .178 | 0.41 | **-1.15 (.78)** | **3.29*** | **.030** | **0.94** |
| Not sitting in WOW position | 4.40 (2.41) | -4.09* | .015 | 2.30 | **-1.50 (3.39)** | **0.99** | **.379** | **0.41** |
| Failing to raise hand when responding | 0.00 (1.58) | 0.00 | 1.00 | 0.00 | **-8.25 (4.66)** | **3.96*** | **.017** | **2.50** |
| Silent during choral activities | 0.20 (.45) | -1.00 | .374 | 0.63 | **-0.35 (2.21)** | **0.36** | **.741** | **0.17** |
| Getting out of seat when inappropriate | -0.20 (.45) | 1.00 | .374 | 0.63 | **-2.25 (1.79)** | **2.82*** | **.048** | **1.78** |
| Not participating in conversations on topic | -0.20 (.46) | 1.00 | .381 | 0.63 | **-1.85 (4.21)** | **0.98** | **.382** | **0.66** |
| Fidgeting | 1.40 (1.34) | -2.33 | .080 | 0.94 | **-5.30 (4.09)** | **2.90*** | **.044** | **1.85** |
| Talking off topic to classmate | 6.60 (5.41) | -2.73 | .053 | 1.68 | **-9.75 (8.79)** | **2.48** | **.068** | **1.41** |
| Putting hands on classmate | 0.20 (.45) | -1.00 | .374 | 0.63 | **0.50 (1.50)** | **1.29** | **.266** | **0.81** |
| Failing to complete assignments on time | 0.00 (.00) | 0.00 | 1.00 | 0.00 | **-0.10 (1.24)** | **0.18** | **.866** | **0.14** |
| Complaining about work | 0.20 (.45) | -1.00 | .374 | 0.40 | **-1.50 (.50)** | **6.71*** | **.003** | **4.24** |
| Displaying negatives attitude towards learning | 0.00 (.71) | 0.00 | 1.00 | 0.00 | **-1.40 (2.88)** | **1.09** | **.338** | **0.86** |
| Composite negative behavior score | 14.20 (9.58) | -3.32* | .029 | 1.80 | **-34.90 (25.39)** | **3.07*** | **.037** | **1.83** |

Note. Bold indicates treatment phase. Class B received the therapy balls treatment during the $T_2$ to $T_3$ period. Difference scores were calculated by subtracting the $T_2$ observation period data from the $T_3$ observation data ($T_3 - T_2$). Therefore, positive scores indicate an increase in observations of undesirable behavior from $T_2$ to $T_3$, and negative scores indicate a decrease. Composite negative behavior scores were calculated by summing all of 14 observed undesirable behaviors. *indicates a significant difference at the .05 level and **indicates a significant difference at the .01 level.

Table 3.
*Teacher reports of the average percent of time that students exhibited desirable behaviors from $T_1$ to $T_2$.*

| Desirable Behavior | Class A | | | | | | Class B | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $T_1$ *Mean* (*SD*) | $T_2$ *Mean* (*SD*) | *Mean* Difference (*SD*) | *t* (11) | *p* | *d* | $T_1$ *Mean* (*SD*) | $T_2$ *Mean* (*SD*) | *Mean* Difference (*SD*) | *t* (11) | *p* | *d* |
| Looking at teacher when appropriate | **54.17** | **60.42** | **6.25** | -1.92 | 0.082 | 0.29 | 56.25 | 64.58 | 8.33 | -2.35* | 0.039 | 0.69 |
| | **(20.87)** | **(22.51)** | **(11.31)** | | | | (11.31) | (12.87) | (12.31) | | | |
| Looking at materials when appropriate | **54.17** | **60.42** | **6.25** | -1.92 | 0.082 | 0.29 | 56.25 | 66.67 | 10.42 | -1.82 | 0.096 | 0.65 |
| | **(20.87)** | **(22.51)** | **(11.31)** | | | | (15.54) | (16.28) | (19.82) | | | |
| Responding appropriately | **52.08** | **62.50** | **10.42** | -2.16 | 0.054 | 0.52 | 64.58 | 64.58 | 0.00 | 0.00 | 1.000 | 0.00 |
| | **(19.82)** | **(19.94)** | **(16.71)** | | | | (12.87) | (12.87) | (10.66) | | | |
| Sitting in WOW position | **52.08** | **66.67** | **14.58** | -3.92** | 0.002 | 0.99 | 56.25 | 47.92 | -8.33 | 1.77 | 0.104 | 0.69 |
| | **(12.87)** | **(16.28)** | **(12.87)** | | | | (15.54) | (7.22) | (16.28) | | | |
| Raising hand when responding | **47.92** | **64.58** | **16.67** | -3.55** | 0.005 | 1.00 | 58.33 | 60.42 | 2.08 | -0.43 | 0.674 | 0.14 |
| | **(16.71)** | **(16.71)** | **(16.28)** | | | | (16.28) | (12.87) | (16.71) | | | |
| Participating in choral activities | **56.25** | **64.58** | **8.33** | -2.35* | 0.039 | 0.47 | 62.50 | 62.50 | 0.00 | 0.00 | 1.000 | 0.00 |
| | **(18.84)** | **(16.71)** | **(12.31)** | | | | (16.86) | (13.06) | (15.06) | | | |
| Staying in seat when appropriate | **64.58** | **75.00** | **10.42** | -2.80* | 0.017 | 1.14 | 68.75 | 72.92 | 4.17 | -1.48 | 0.166 | 0.34 |
| | **(12.87)** | **(0.00)** | **(12.87)** | | | | (11.31) | (12.87) | (9.73) | | | |
| Participating in conversations on topic | **52.08** | **70.83** | **18.75** | -4.18* | 0.020 | 1.20 | 62.50 | 60.42 | -2.08 | 0.43 | 0.674 | 0.14 |
| | **(19.82)** | **(9.73)** | **(15.54)** | | | | (16.86) | (12.87) | (16.71) | | | |
| Not fidgeting | **47.92** | **50.00** | **2.08** | -0.29 | 0.777 | 0.11 | 50.00 | 43.75 | -6.25 | 1.39 | 0.191 | 0.31 |
| | **(22.51)** | **(15.08)** | **(24.91)** | | | | (15.08) | (24.13) | (15.54) | | | |
| Talking on topic to classmate | **47.92** | **66.67** | **18.75** | -2.69* | 0.021 | 0.95 | 54.17 | 45.83 | -8.33 | 2.35* | 0.039 | 0.46 |
| | **(19.82)** | **(22.19)** | **(24.13)** | | | | (14.43) | (20.87) | (12.31) | | | |
| Keeping hands to oneself | **91.68** | **77.08** | **14.58** | -3.02* | 0.012 | 0.62 | 75.00 | 77.92 | 2.08 | 0.43 | 0.674 | 0.12 |
| | **(22.19)** | **(24.91)** | **(16.71)** | | | | (28.20) | (19.82) | (16.71) | | | |
| Completing assignments on time | **56.25** | **66.67** | **10.42** | -2.16 | 0.054 | 0.51 | 62.50 | 64.58 | 2.08 | -0.36 | 0.723 | 0.11 |
| | **(26.38)** | **(12.31)** | **(16.71)** | | | | (13.06) | (22.51) | (19.82) | | | |
| Not complaining about work | **52.08** | **68.75** | **16.67** | -3.55** | 0.005 | 0.80 | 56.25 | 52.08 | -4.17 | 1.48 | 0.166 | 0.17 |
| | **(19.82)** | **(21.65)** | **(16.28)** | | | | (24.21) | (24.91) | (9.73) | | | |
| Positive attitude towards learning | **60.42** | **66.67** | **6.25** | -1.39 | 0.191 | 0.43 | 52.08 | 47.92 | -4.17 | 1.48 | 0.166 | 0.23 |
| | **(12.87)** | **(16.28)** | **(15.54)** | | | | (16.71) | (19.82) | (9.73) | | | |
| Asking related questions | **66.67** | **70.83** | **4.17** | -0.62 | 0.551 | 0.22 | 58.33 | 56.25 | -.2.08 | 0.56 | 0.586 | 0.09 |
| | **(12.63)** | **(23.44)** | **(23.44)** | | | | (24.62) | (24.13) | (12.87) | | | |
| Composite desirable behavior score (Teacher observation) | **57.11** | **66.11** | **9.00** | -3.56** | 0.004 | 0.68 | 59.61 | 58.90 | -0.71 | 0.52 | 0.613 | 0.06 |
| | **(12.63)** | **(13.69)** | **(8.75)** | | | | (11.91) | (12.65) | (22.08) | | | |

Note. Bold indicates treatment phase. Class A received the therapy balls treatment during the $T_1$ to $T_2$ period. Difference scores were calculated by subtracting the $T_1$ observation period data from the $T_2$ observation data ($T_2 - T_1$). Therefore, positive scores indicate an increase in teachers' observations of desirable behavior from $T_1$ to $T_2$, and negative scores indicate a decrease. *indicates a significant difference at the .05 level and **indicates a significant difference at the .01 level.

Table 4.
*Teacher reports of the average percent of time that students exhibited desirable behaviors from $T_2$ to $T_3$.*

| Desirable Behavior | Class A | | | | | | Class B | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $T_2$ *Mean* (*SD*) | $T_3$ *Mean* (*SD*) | *Mean* Difference (*SD*) | *t* (11) | *p* | *d* | $T_2$ *Mean* (*SD*) | $T_3$ *Mean* (*SD*) | *Mean* Difference (*SD*) | *t* (11) | *p* | *d* |
| Looking at teacher when appropriate | 60.42 (22.51) | 64.58 (12.87) | 4.17 (14.43) | -1.00 | 0.339 | 0.23 | **64.58 (12.87)** | **64.59 (12.87)** | **0.00 (0.00)** | 0.00 | 1.000 | 0.00 |
| Looking at materials when appropriate | 60.42 (22.51) | 64.58 (12.87) | 4.17 (14.43) | -1.00 | 0.339 | 0.23 | **66.67 (16.28)** | **62.50 (13.06)** | **-4.17 (14.43)** | 1.00 | 0.339 | 0.28 |
| Responding appropriately | 62.50 (19.94) | 64.58 (12.87) | 2.08 (12.87) | -0.56 | 0.586 | 0.12 | **64.58 (12.87)** | **64.58 (16.71)** | **0.00 (.60)** | 0.00 | 1.000 | 0.00 |
| Sitting in WOW position | 66.67 (16.28) | 70.83 (14.43) | 4.17 (9.73) | -1.48 | 0.166 | 0.27 | **47.92 (7.22)** | **56.25 (18.84)** | **8.33 (19.46)** | -1.48 | 0.166 | 0.58 |
| Raising hand when responding | 64.58 (16.71) | 66.67 (12.31) | 2.08 (16.71) | -0.43 | 0.674 | 0.14 | **60.42 (12.87)** | **64.58 (16.71)** | **4.17 (20.87)** | -0.69 | 0.504 | 0.28 |
| Participating in choral activities | 64.58 (16.71) | 62.50 (13.06) | -2.08 (12.87) | 0.56 | 0.586 | 0.14 | **62.50 (13.06)** | **62.50 (16.86)** | **0.00 (10.66)** | 0.00 | 1.000 | 0.00 |
| Staying in seat when appropriate | 75.00 (0.00) | 72.92 (7.22) | -2.08 (7.22) | 1.00 | 0.339 | 0.41 | **72.92 (12.87)** | **72.92 (12.87)** | **0.00 (0.00)** | 0.00 | 1.000 | 0.00 |
| Participating in conversations on topic | 70.83 (9.73) | 64.58 (12.87) | -6.25 (11.31) | 1.92 | 0.082 | 0.55 | **60.42 (12.87)** | **62.50 (13.06)** | **2.08 (12.87)** | -0.56 | 0.586 | 0.16 |
| Not fidgeting | 50.00 (15.08) | 60.42 (12.87) | 10.42 (12.87) | -2.80* | 0.017 | 0.74 | **43.75 (24.13)** | **52.08 (24.91)** | **8.33 (16.28)** | -1.77 | 0.104 | 0.34 |
| Talking on topic to classmate | 66.67 (22.19) | 56.25 (18.84) | -10.42 (22.51) | 1.60 | 0.137 | 0.51 | **45.83 (20.87)** | **54.17 (20.87)** | **8.33 (12.31)** | -2.35* | 0.039 | 0.40 |
| Keeping hands to oneself | 77.08 (24.91) | 75.00 (23.81) | -2.08 (16.71) | 0.43 | 0.674 | 0.09 | **72.92 (19.82)** | **81.25 (21.65)** | **8.33 (12.31)** | -2.35* | 0.039 | 0.40 |
| Completing assignments on time | 66.67 (12.31) | 64.58 (12.87) | -2.08 (7.22) | 1.00 | 0.339 | 0.17 | **64.58 (22.51)** | **68.75 (15.54)** | **4.17 (14.43)** | -1.00 | 0.339 | 0.22 |
| Not complaining about work | 68.75 (21.66) | 68.75 (15.54) | 0.00 (21.32) | 0.00 | 1.00 | 0.00 | **52.08 (24.91)** | **58.33 (26.83)** | **6.25 (11.31)** | -1.92 | 0.082 | 0.24 |
| Positive attitude towards learning | 66.67 (16.28) | 62.50 (13.06) | -4.17 (14.43) | 1.00 | 0.339 | 0.28 | **47.92 (19.82)** | **56.25 (18.84)** | **8.33 (19.46)** | -2.35* | 0.039 | 0.43 |
| Asking related questions | 70.83 (23.44) | 72.92 (12.87) | 2.08 (16.71) | -0.43 | 0.674 | 0.11 | **56.25 (24.13)** | **62.50 (29.19)** | **6.25 (15.54)** | -1.39 | 0.191 | 0.23 |
| Composite desirable behavior score (Teacher observation) | 66.11 (13.69) | 66.15 (8.71) | 0.04 (8.44) | -0.02 | 0.987 | 0.00 | **58.90 (11.65)** | **62.90 (13.41)** | **4.25 (3.39)** | -4.09 | 0.002 | 0.31 |

Note. Bold indicates treatment phase. Class A received the therapy balls treatment during the $T_2$ to $T_3$ period. Difference scores were calculated by subtracting the $T_2$ observation period data from the $T_3$ observation data ($T_3 - T_2$). Therefore, positive scores indicate an increase in teachers' observations of desirable behavior from $T_2$ to $T_3$, and negative scores indicate a decrease. *indicates a significant difference at the .05 level and **indicates a significant difference at the .01 level.

Table 5.
*Student self-reports of the average percent of time that they exhibited desirable behaviors from $T_1$ to $T_2$.*

| Desirable Behavior | Class A | | | | | | Class B | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $T_1$ *Mean* (*SD*) | $T_2$ *Mean* (*SD*) | *Mean* Difference (*SD*) | *t* (11) | *p* | *d* | $T_1$ *Mean* (*SD*) | $T_2$ *Mean* (*SD*) | *Mean* Difference (*SD*) | *t* (11) | *p* | *d* |
| Looking at teacher when appropriate | **83.33** | **79.17** | **-4.17** | 1.00 | 0.339 | 0.27 | 70.83 | 77.08 | 6.25 | -1.15 | 0.275 | 0.40 |
| | **(16.28)** | **(14.43)** | **(14.43)** | | | | (14.43) | (16.71) | (18.84) | | | |
| Looking at materials when appropriate | **83.33** | **72.92** | **-10.42** | 0.89 | 0.392 | 0.39 | 75.00 | 64.58 | -10.42 | 1.00 | 0.339 | 0.41 |
| | **(24.62)** | **(29.11)** | **(40.53)** | | | | (21.32) | (29.11) | (36.08) | | | |
| Responding appropriately | **79.17** | **91.67** | **12.50** | -1.32 | 0.214 | 0.55 | 79.18 | 62.50 | -16.68 | 1.54 | 0.151 | 0.63 |
| | **(29.84)** | **(12.31)** | **(32.86)** | | | | (20.87) | (31.08) | (37.44) | | | |
| Sitting in WOW position | **54.17** | **60.42** | **6.25** | -0.90 | 0.389 | 0.40 | 54.17 | 62.50 | 8.33 | -1.00 | 0.339 | 0.30 |
| | **(17.79)** | **(12.87)** | **(24.13)** | | | | (29.84) | (25.00) | (28.87) | | | |
| Raising hand when responding | **85.42** | **56.25** | **-29.17** | 3.63** | 0.004 | 1.41 | 60.42 | 64.58 | 4.17 | -0.43 | 0.674 | 0.17 |
| | **(16.71)** | **(24.13)** | **(27.87)** | | | | (16.71) | (31.00) | (33.43) | | | |
| Participating in choral activities | **93.75** | **79.18** | **-14.58** | 1.87 | 0.089 | 0.87 | 83.33 | 83.33 | 0.00 | 0.00 | 1.00 | 0.00 |
| | **(15.54)** | **(17.94)** | **(27.09)** | | | | (22.19) | (22.19) | (33.71) | | | |
| Staying in seat when appropriate | **85.42** | **77.08** | **-8.33** | 1.17 | 0.266 | 0.45 | 75.00 | 75.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| | **(12.87)** | **(22.51)** | **(24.62)** | | | | (26.11) | (23.84) | (15.08) | | | |
| Participating in conversations on topic | **77.08** | **70.83** | **-6.25** | 0.67 | 0.515 | 0.31 | 56.25 | 68.75 | 12.50 | -1.48 | 0.166 | 0.55 |
| | **(16.71)** | **(23.44)** | **(32.20)** | | | | (28.45) | (15.54) | (29.19) | | | |
| Not fidgeting | **41.67** | **47.92** | **6.25** | -0.38 | 0.714 | 0.20 | 79.17 | 31.25 | -47.92 | 2.82* | 0.017 | 1.46 |
| | **(32.57)** | **(29.11)** | **(57.53)** | | | | (31.68) | (33.92) | (58.83) | | | |
| Talking on topic to classmate | **31.25** | **58.33** | **27.08** | -3.22** | 0.008 | 1.41 | 39.58 | 66.67 | 27.08 | -2.72* | 0.020 | 1.29 |
| | **(21.65)** | **(16.28)** | **(29.11)** | | | | (22.51) | (19.46) | (34.47) | | | |
| Keeping hands to oneself | **14.58** | **64.58** | **50.00** | -5.14** | 0.000 | 1.72 | 20.83 | 64.58 | 43.75 | -3.02* | 0.012 | 1.44 |
| | **(29.11)** | **(29.11)** | **(33.71)** | | | | (23.44) | (36.08) | (50.14) | | | |
| Completing assignments on time | **64.58** | **70.83** | **6.25** | -0.64 | 0.536 | 0.26 | 54.17 | 75.00 | 20.83 | -3.08* | 0.010 | 0.77 |
| | **(24.91)** | **(23.44)** | **(33.92)** | | | | (27.87) | (26.11) | (23.44) | | | |
| Not complaining about work | **29.17** | **75.00** | **45.83** | -5.70** | 0.000 | 2.33 | 20.83 | 54.17 | 33.33 | -2.60* | 0.025 | 1.29 |
| | **(23.44)** | **(15.08)** | **(27.87)** | | | | (20.87) | (29.84) | (44.38) | | | |
| Positive attitude towards learning | **79.17** | **70.83** | **-8.33** | 1.08 | 0.305 | 0.46 | 79.17 | 60.42 | -18.75 | 2.14 | 0.056 | 0.92 |
| | **(14.43)** | **(20.87)** | **(26.83)** | | | | (14.43) | (24.91) | (30.39) | | | |
| Asking related questions | **29.17** | **75.00** | **45.83** | -4.16** | 0.002 | 1.94 | 43.75 | 60.42 | 16.67 | -1.38 | 0.194 | 0.60 |
| | **(20.87)** | **(26.11)** | **(38.19)** | | | | (28.45) | (27.09) | (41.74) | | | |
| Composite self-report desirable behavior score | **62.06** | **69.98** | **7.92** | -3.15** | 0.009 | 0.93 | 59.46 | 64.73 | 5.27 | -1.58 | 0.142 | 0.56 |
| | **(7.74)** | **(9.22)** | **(8.72)** | | | | (8.10) | (10.63) | (11.55) | | | |

Note. Bold indicates treatment phase. Class A received the therapy balls treatment during the $T_1$ to $T_2$ period. Difference scores were calculated by subtracting the $T_1$ observation period data from the $T_2$ observation data ($T_2 - T_1$). Therefore, positive scores indicate an increase in students' self-reports of desirable behavior from $T_1$ to $T_2$, and negative scores indicate a decrease. *indicates a significant difference at the .05 level and **indicates a significant difference at the .01 level.

Table 6.
*Student self-reports of the average percent of time that they exhibited desirable behaviors from $T_2$ to $T_3$.*

| Desirable Behavior | Class A | | | | | | Class B | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $T_2$ *Mean* (*SD*) | $T_3$ *Mean* (*SD*) | *Mean* Difference (*SD*) | *t* (11) | *p* | *d* | $T_2$ *Mean* (*SD*) | $T_3$ *Mean* (*SD*) | *Mean* Difference (*SD*) | *t* (11) | *p* | *d* |
| Looking at teacher when appropriate | 79.17 (14.43) | 77.08 (16.71) | -2.08 (12.87) | 0.56 | 0.586 | 0.13 | **77.08 (16.71)** | **75.00 (15.08)** | **-2.08 (12.87)** | 0.56 | 0.586 | 0.13 |
| Looking at materials when appropriate | 72.92 (29.11) | 62.50 (25.00) | -10.42 (29.11) | 1.24 | 0.241 | 0.38 | **64.58 (29.11)** | **68.75 (15.52)** | **4.17 (27.87)** | -0.52 | 0.615 | 0.18 |
| Responding appropriately | 91.67 (12.31) | 85.42 (16.71) | -6.25 (18.84) | 1.15 | 0.275 | 0.43 | **62.50 (31.08)** | **83.33 (28.87)** | **20.83 (29.84)** | -2.42* | 0.034 | 0.69 |
| Sitting in WOW position | 60.42 (12.87) | 50.00 (10.66) | -10.42 (12.87) | 2.80* | 0.017 | 0.88 | **62.50 (25.00)** | **62.50 (32.86)** | **.00 (42.64)** | 0.00 | 1.000 | 0.00 |
| Raising hand when responding | 56.25 (24.13) | 68.75 (11.31) | 12.50 (19.94) | -2.17 | 0.053 | 0.66 | **64.58 (31.00)** | **66.67 (28.87)** | **2.08 (19.82)** | -0.36 | 0.723 | 0.07 |
| Participating in choral activities | 79.17 (17.94) | 79.17 (17.94) | 0.00 (28.20) | 0.00 | 1.00 | 0.00 | **83.33 (22.19)** | **70.83 (29.84)** | **-12.50 (40.59)** | 1.07 | 0.309 | 0.48 |
| Staying in seat when appropriate | 77.08 (22.51) | 75.00 (26.11) | -2.08 (19.82) | 0.36 | 0.723 | 0.09 | **75.00 (23.84)** | **87.50 (19.94)** | **12.50 (50.09)** | -1.59 | 0.139 | 0.57 |
| Participating in conversations on topic | 70.83 (23.44) | 66.67 (19.46) | -4.17 (23.44) | 0.62 | 0.551 | 0.19 | **68.75 (15.54)** | **77.08 (16.71)** | **8.33 (24.62)** | -1.17 | 0.266 | 0.52 |
| Not fidgeting | 47.92 (29.11) | 60.42 (22.51) | 12.50 (29.19) | -1.48 | 0.166 | 0.48 | **31.25 (33.92)** | **52.08 (29.11)** | **20.83 (31.67)** | -2.28* | 0.044 | 0.66 |
| Talking on topic to classmate | 58.33 (16.28) | 58.33 (22.19) | 0.00 (23.84) | 0.00 | 1.00 | .00 | **66.67 (19.46)** | **60.42 (19.82)** | **-6.25 (18.84)** | 1.15 | 0.275 | 0.52 |
| Keeping hands to oneself | 64.58 (29.11) | 54.17 (35.09) | -10.42 (31.00) | 1.16 | 0.269 | 0.32 | **64.58 (36.08)** | **75.00 (21.32)** | **10.42 (36.08)** | -1.00 | 0.339 | 0.32 |
| Completing assignments on time | 70.83 (23.44) | 68.75 (26.38) | -2.08 (12.87) | 0.56 | 0.586 | 0.08 | **75.00 (26.11)** | **77.08 (27.09)** | **2.08 (32.78)** | -0.220 | 0.830 | 0.35 |
| Not complaining about work | 75.00 (15.08) | 64.58 (19.82) | -10.42 (24.91) | 1.45 | 0.175 | 0.59 | **54.17 (29.84)** | **64.58 (24.91)** | **10.42 (29.11)** | -1.24 | 0.241 | 0.09 |
| Positive attitude towards learning | 70.83 (20.87) | 41.67 (28.87) | -29.17 (45.02) | 2.24* | 0.046 | 1.16 | **60.42 (24.91)** | **60.42 (29.11)** | **0.00 (21.32)** | 0.000 | 1.000 | 0.00 |
| Asking related questions | 75.00 (26.11) | 66.67 (19.46) | -8.33 (28.87) | 1.00 | 0.339 | 0.36 | **60.42 (27.09)** | **66.67 (26.83)** | **6.25 (26.38)** | -0.821 | 0.429 | 0.23 |
| Composite self-report desirable behavior score | 69.98 (9.22) | 65.25 (7.52) | -4.73 (8.82) | 1.86 | 0.09 | 0.56 | **64.73 (10.63)** | **69.88 (11.74)** | **5.15 (11.80)** | -1.51 | 0.159 | 0.46 |

Note. Bold indicates treatment phase. Class A received the therapy balls treatment during the $T_2$ to $T_3$ period. Difference scores were calculated by subtracting the $T_2$ observation period data from the $T_3$ observation data ($T_3 - T_2$). Therefore, positive scores indicate an increase in students' self-report of desirable behavior from $T_2$ to $T_3$, and negative scores indicate a decrease. *indicates a significant difference at the .05 level and **indicates a significant difference at the .01 level.