The Physical Educator Vol. 72 • pp. 193–205 • 2015

ASSESSMENT

Effect of Frequent Peer-Monitored Testing and Personal Goal Setting on Fitnessgram Scores of Hispanic Middle School Students

Grant Hill and Aaron Downing

Abstract

The purpose of this study was to determine the effects of frequent peer-monitored Fitnessgram testing, with student goal setting, on the PACER and push-up performance of middle school students. Subjects were 176 females and 189 males in 10 physical education classes at a middle school with an 83.7% Hispanic student population. Students were baseline fitness tested with five classes assigned to the control group and five to the experimental group with no significant differences between the two groups in baseline fitness test performance. Students in the experimental group set personal goals and participated in peer fitness testing four times over the next 18 weeks. Results from formal teacher testing at the end of 18 weeks using MANOVA demonstrated that placement in the experimental group had no effect on pre- versus posttest scores for PACER and push-up tests compared with the control group. Students in control and experimental groups also completed the PAQ-A, with results indicating a significant positive correlation of higher weekly activity levels with push-up scores, but no significant difference for weekly activity levels and PACER scores. Results are discussed in terms of Locke's goal setting theory as well as recent research pertaining to youth fitness testing.

Grant Hill is a professor in the Department of Kinesiology, California State University, Long Beach. Aaron Downing is a physical educator at Lake Center Middle School in Santa Fe Springs, CA. Please send author correspondence to Grant Hill, Department of Kinesiology, California State University, Long Beach, 1250 Bellflower Blvd., Long Beach, CA 90840. E-mail: ghill@csulb.edu

Fitness testing has been a part of most K–12 physical education (PE) programs since the creation of the President's Council on Physical Fitness during the 1950s (Morrow, Weimo, Franks, Meredith, & Spain, 2009). Fitnessgram is one of the most frequently implemented youth fitness test programs in the United States (Keating & Silverman, 2004). Since 1995, California state law has required students in the fifth, seventh, and ninth grades in public schools to take the Fitnessgram, which has been used to reflect California students' health-related fitness profile (California Department of Education, 2003). Fitnessgram includes six subtests: sit and reach, skinfold measurement, PACER/mile run, push-ups, curl-ups, and shoulder stretch. In the state of California, fitness testing has become a high-stakes process wherein 10th grade students who do not achieve scores in the healthy zone in at least five out of six of the fitness subtests are required to take 2 additional years of PE, or each semester until they pass at least five of the six subtests (California Department of Education, 2012).

Wiersma and Sherman (2008) stated that when physical fitness testing is conducted in a motivating manner, it increases internal validity, self-efficacy, enjoyment, and overall interest in physical activity (PA) and that self-assessment is a viable means to increase competence in fitness performance. In contrast, Corbin (2009) argued that fitness test scores are related to factors other than fitness promotion strategies that may be employed in a school PE class. The Ontario Physical and Health Education Association (OPHEA, 2006) stated that since children mature at different rates, fitness test results are largely determined by physical maturity. In a feasibility study commissioned by the National Assembly for Wales (Cale & Harris, 2009), the value of fitness testing on promoting healthy lifestyles and PA was questioned, and Morrow and Freedson (1994) found a low relationship between fitness scores and PA among youth. Corbin (2002) stated that for elementary and middle school students, fitness test scores are difficult to predict from PA patterns. Consequently, any intervention in a PE class alone will unlikely result in major changes in physical fitness scores over the short term.

Harris and Cale (2007) and Rice (2007) warned that fitness testing may contribute to a diminished interest in PE and PA in general because the results undermine the confidence, self-esteem, and sense of self as a PA participant for those who either have low scores or do not experience improvement. Others (Cale, Harris, & Chen, 2007; Rice, 2007; Rowland, 1995) have called for an end to fit-

ness testing in schools altogether because they perceive it does more harm than good as it is time consuming, embarrassing to students, and not effective in promoting PA.

Because of the apparent controversy regarding the value of fitness testing, a study was deemed necessary to determine whether a specific intervention would be effective in increasing Fitnessgram PACER and push-up scores of seventh and eighth grade students in a school with a primarily Hispanic population over 18 weeks. The specific intervention was to provide frequent peer-monitored fitness testing with student-generated goal setting after each test. In this study, PACER and push-up scores of seventh and eighth grade male and female students in required PE classes who engaged in peer-monitored fitness testing and goal setting four times over 18 weeks were compared with the scores of students in a control group to determine whether either group showed significantly greater improvements in fitness scores. The results of this study are considered to be important in determining whether increasing the frequency of fitness testing with student goal setting is an effective strategy to improve youth fitness scores.

Goal setting is a positive motivational strategy that is designed to improve performance (Burton, 1992). Goal setting has been found to be effective in improving long-term self-motivation through eliciting commitment, perseverance, dedication, and effort. Goals tend to provide a focus and direction for a person's activity and permit an individual to measure performance continuously through internal processes of comparison using subjective standards to evaluate ongoing pursuits (Locke & Latham, 1990). Locke, Shaw, Saari, and Latham (1981) reviewed 110 workplace studies and concluded that 99 of them reported findings supportive of his theory that specific, difficult goals lead to higher levels of task performance than "do-your-best" goals, easy goals, or no goals. Mento, Cartledge, and Locke (1980) and Tubbs (1986) also found increased performance and productivity associated with specific goal setting. However, the effectiveness of goal setting in a sport and exercise setting has been tested in few studies (Annesi, 2002), and generally, studies on the influence of goal setting on performance of a physical skill have resulted in inconsistent findings (La Clair, 1994). Shilts, Horowitz, and Townsend (2004), in reviewing research related to the effectiveness of goal setting in regard to improving nutrition and PA practices, found that no studies had been conducted with middle school adolescents. They also stated that attempting to change the dietary

and PA behaviors of youth aged 12 to 14 years through self-set goals may be theoretically futile because children this young have not yet developed the ability to think logically about abstractions. Consequently, in this research, an issue that has not been previously examined is addressed: whether frequent peer-monitored fitness testing with student goal setting is an effective strategy to improve middle school fitness test scores.

Methods

Ten PE classes in a Southern California middle school with 176 females and 189 males aged 12 to 14 participated in the study. The published ethnic percentages for this school were 83.7% Hispanic, 6.1% Caucasian, 1.8% African American/Black, 1.2% Asian, and 7.2% Other or Unreported. Approximately 67% of the students in this school were reported as eligible to receive free or subsidized meals. Approval was secured from the district review board, and informed consent was obtained from parents of the participants. One PE teacher arbitrarily designated three classes to be in the experimental group and two classes to be in the control group. Another PE teacher designated two classes to be in the experimental group and three classes to be in the control group. The two teachers had taught at the same school for the past 3 years and reviewed testing procedures to ensure consistency in administrating the test items. Students in the control and experimental groups were formally tested by their teachers during the first week of the semester to establish baseline fitness scores for the 20-m PACER and push-up tests using testing procedures as specified in Fitnessgram (The Cooper Institute, 2010). Students in the experimental group subsequently completed the PACER and push-up tests in small groups every 3 to 4 weeks for the remainder of the semester (i.e., four peer-monitored tests). In addition to recording their own scores, students in the experimental classes set personal goals for their next test performance. Students entered their self-report scores on sheets collected by the instructors after each test. Those sheets were returned to students during each subsequent peer-monitored test so they could view their goals and progress. Six units, each approximately 3 weeks in length, were offered during the semester: flag rugby, tumbling, paddle tennis, football, soft lacrosse, and softball. At the end of the 18 weeks, students in the control and experimental groups were again formally tested by the teachers. At the end of 18 weeks, students in the control and experimental classes completed the Physical Activity Questionnaire

for Adolescents (PAQ-A), which indicates the frequency, duration, and type of PA in which the students have recently engaged (Kowalski, Crocker, & Donen, 2004). A MANOVA was used to determine if there were differences in PACER and push-up performance between the control and experimental groups for the pre- and posttests. The researchers conducted *t* tests to determine whether there were differences between the final self-report test and final PACER and push-up scores for the experimental group. An ANOVA was conducted to determine whether students with higher scores on the PACER and push-up tests also reported higher frequencies of PA using the PAQ-A.

Results

Through the MANOVA, it was demonstrated that whether students were placed in a control group or an experimental group had no effect on their pre- and posttest scores for the PACER and push-up tests ($p < .01$). In some cases, students in the control group had significant improvements between pre- and posttest scores, and in other cases, students in the experimental groups had significant improvements between pre- and posttest scores (see Tables 1 and 2). The means for the final peer-monitored push-up test and the final teacher-administered push-up test were very similar ($M = 15.69$, $SD = 6.41$; $M = 15.96$, $SD = 7.07$), and the Pearson product–moment coefficient indicated a significant correlation ($r = 0.50$, $p < .000$). The means of the final PACER peer-monitored test ($M = 14.41$, $SD = 4.37$) and final teacher-administered PACER test ($M = 23.99$, $SD = 11.7$) were also significantly correlated ($r = 0.354$, $p < .000$). Through the ANOVA, no significant differences were demonstrated between the control and experimental groups in regard to number of weekly activities (Item 1), effort and intensity of activity (Items 2 to 7), and frequency of daily activity (Item 8) reported by students ($p < .01$; see Table 3). Based on the responses of students by gender, no significant differences for any of the PAQ-A items were revealed ($p < .01$). However, students who reported the greatest frequency of daily PA (PAQ-A Item #8) achieved significantly higher scores on the push-up test (see Table 4). A comparison of the final push-up and PACER scores with PAQ-A Item 8 yielded low (0.12, .11), nonsignificant positive Pearson product–moment correlations.

Table 1

Means and Standard Deviations of Push-Up and PACER Scores for Middle School Experimental and Control Groups by Gender

	Push-ups		PACER	
Groups	**Pretest**	**Posttest**	**Pretest**	**Posttest**
Experimental				
Boys (n = 79)	17.36 (8.58)	17.92 (7.05)	25.76 (11.46)	28.18 (13.67)*
Girls (n = 72)	12.60 (5.49)	14.50 (6.48)*	19.32 (5.80)	19.94 (7.27)
Control				
Boys (n = 112)	15.81 (7.82)	18.79 (8.11)*	21.39 (10.88)	24.77 (13.24)*
Girls (n = 90)	14.06 (6.91)	14.81 (6.47)	17.45 (7.23)	20.57 (8.44)

*Posttest score significantly greater than pretest score, $p < .01$.

Table 2

Means and Standard Deviations of Push-Up and PACER Scores for Middle School Experimental and Control Groups by Grade Level

	Push-ups		PACER	
Groups	**Pretest**	**Posttest**	**Pretest**	**Posttest**
Experimental				
6th (n = 46)	15.55 (8.09)	17.07 (8.48)	20.60 (5.37)	18.47 (5.94)
7th (n = 51)	12.98 (5.99)	14.33 (6.21)	20.67 (9.33)	23.37 (11.83)*
8th (n = 57)	16.13 (8.05)	17.07 (5.98)	25.44 (11.45)	28.76 (12.73)*
Control				
6th (n = 41)	15.49 (8.65)	18.49 (9.70)*	14.53 (5.43)	17.47 (6.38)*
7th (n = 91)	12.46 (6.55)	15.66 (6.30)	24.11 (11.30)	27.92 (13.92)*
8th (n = 77)	17.85 (6.73)	17.73 (7.72)	16.31 (5.04)	18.98 (5.79)

*Posttest score significantly greater than pretest score, $p < .01$.

Table 3

Means and Standard Deviations of Reported Physical Activity Levels for Experimental and Control Groups Using PAQ-A

PAQ item #	Experimental	Control	*p*
Number of activities in past week (PAQ Item 1)	1.79 (0.68)	1.78 (0.69)	0.90
Effort and intensity (PAQ composite Items 2–7)	3.01 (0.89)	3.07 (0.94)	0.60
Frequency of daily activity (PAQ Item 8)	2.72 (1.01)	2.83 (1.17)	0.38

Note. No significant differences found between experimental and control group scores ($p < .01$).

Table 4
Means and Standard Deviations of Push-Up and PACER Scores of Students Who Reported Various Frequencies of Physical Activity Using the Physical Activity Questionnaire for Adolescents (PAQ-A)

Student reported frequency of daily activity (PAQ Item 8)	Push-ups	PACER laps
1 - None (43)	16.2 (8.4)	21.1 (11.0)
2 - Little bit (90)	15.4 (5.6)	22.3 (11.6)
3 - Medium (109)	15.9 (6.9)	22.9 (10.8)
4 - Often (64)	17.6 (6.7)	26.1 (11.8)
5 - Very often (22)	20.4 (8.5)*	23.3 (10.5)

*Significant differences for push-ups based on frequency of daily activity level at .01 level. Group 5 > than Groups 1, 2, and 3.

Discussion

The results indicate that frequent peer-monitored fitness testing with goal setting as an 18-week intervention strategy does not positively impact posttest PACER or push-up test scores for middle school boys or girls. Specifically, students in the control group experienced equal or greater gains compared to students in the experimental group in PACER and push-up scores. These findings appear to support the results of the feasibility study that was commissioned by the National Assembly for Wales, through which the value of increasing class time spent on fitness testing was questioned (Cale & Harris, 2009).

For several reasons, the students in the experimental group did not experience greater gains in push-up and PACER scores than students in the control group. First, students in the experimental group may have disliked or resented using PE time for fitness testing because they found it boring or embarrassing (Silverman, Keating, & Phillips, 2008). This may partially explain the moderate correlations between the experimental group's self-test and final teacher-administered test scores. Specifically, students in the experimental group may have performed below their ability level during the peer-monitored testing sessions due to a lack of motivation (Domangue & Solmon, 2010; Mahar & Rowe, 2008). Second, differences in group dynamics within specific classes may have affected the scores (e.g., the means in the PACER and push-up tests were significantly higher for the sixth grade boys than seventh grade boys in the control and

experimental groups). Third, given the similarity between the experimental and control groups PAQ-A scores, it appears as if the students in the experimental group were not motivated to be more physically active than students in the control group, despite the frequent peer-monitored fitness testing (Cale et al., 2007). Fourth, given the significant improvements in push-up and PACER scores for boys in the control group, it appears as if physical maturation over the 18 weeks may have played a more important role improving test scores than the intervention (Lloyd, Colley, & Tremblay, 2010).

These findings do not appear to support Locke's theory of goal setting, although Locke's research was primarily in industrial and organizational settings with adults (Weinberg, 1994). In addition, the goals the students set may not have been taken seriously or realistic because no incentive was provided for reaching those goals. The students may have also been too young to formulate realistic, motivating goals that would change their dietary and PA behaviors (Shilts et al., 2004). It is also possible that the students in the control groups informally set their own goals for the final fitness test based on their initial score (Correa & Souza, 2009). The effectiveness of the goal setting may also have been limited because the students in the experimental groups had reached the limits of their physical ability (Weinberg, 1994). In addition, according to Weinberg (2010), just setting goals does not ensure improvements in performance or productivity—certain principles and guidelines should be followed to maximize their effectiveness. The students in the experimental group of this study wrote specific and measurable goals, but those goals may not have been realistic and the students may not have formulated specific plans to reach their goals.

It is not surprising that males in both groups scored higher than their female counterparts in the push-up and PACER tests. This finding is consistent with Fitnessgram gender standards, which require males to score higher in all categories than females to reach the Healthy Fitness Zone. Males typically perform better on the push-up test than females due to genetic strength and maturation (The Cooper Institute, 2010).

Notably, students who reported the highest frequency of daily PA scored significantly higher on the push-up test, but no significant differences were found for the PACER test. However, because correlations between student-reported weekly activity (PAQ-A) and push-up and PACER test results were extremely low, these results provide only partial evidence that weekly PA is a primary factor pre-

dicting fitness score performance, findings that are congruent with those of Morrow and Freedson (1994) and others who have reported low correlations between adolescent PA levels and fitness test scores (Armstrong & Welsman, 1997; Cale & Harris, 2009; Cale et al., 2007). The results also appear to affirm Corbin (2009), who stated it is difficult to predict youth fitness scores from PA patterns for elementary and middle school students.

Conclusions and Recommendations

The results of this study are consistent with those of previous studies in which the researchers found a low correlation between youth fitness test scores and reported PA levels (Armstrong & Welsman, 1997; Cale & Harris, 2009; Cale et al., 2007). The participants in this study were primarily Hispanic; the results should not be generalized to other ethnic groups since physical activity levels and fitness scores for middle school–aged students have been found to be dissimilar among ethnic and socioeconomic groups (Grieser et al., 2008; Fahlman, Hall, & Lock, 2006; Hoelscher, Barroso, Springer, Castrucci, & Kelder, 2009; Yoo, Lounsbery, Bungum, & Gast, 2010). In addition, the validity of self-report measures of PA have been shown to have limited validity among children (Pate, 1993). A more accurate measure of weekly PA may have been realized through the use of pedometers or accelerometers (Kelly et al., 2010).

Rather than spending increased time in PE classes for fitness testing, it appears to be more important to use class time to help students engage in moderate to vigorous PA and identify ways to increase their daily PA levels (Cale & Harris, 2009; Pangrazi, 2000). In regard to goal setting, rather than having middle school students set their own fitness test goals, it may be more productive to have them use "guided goal setting," which involves having them choose from a variety of goals that the instructor has developed (Shilts et al., 2004). In addition, students should also be required to formulate specific plans to increase their activity levels specific to the Fitnessgram component tests (Weinberg, 2010). Middle school teachers should find ways to reinforce student achievement that is focused on their personal activity plans. It also appears important to inform PE teachers that it is not until adolescence that regular PA begins to override heredity, maturation, and age as primary factors affecting fitness test scores (Pangrazi & Corbin, 2008; Wrench & Garrett, 2008).

Since the link between fitness scores and healthy lifestyle is not causal, PE teachers should strive offer comprehensive curriculums

that are focused on addressing state and/or national standards in PE as opposed to narrowly seeking to justify the value of their programs by student fitness test scores (Wrench & Garrett, 2008). This may be challenging given that program accountability for PE programs appears to be linked to improving student fitness test scores. However, if students are given opportunities to master sport-related motor skills, it should positively impact their self-efficacy, which should make them more likely to participate in independent PA (Chase, 2001).

Future research should be focused on identifying the most salient ways to motivate students to increase their PA levels, particularly during the elementary school years. Since PA levels have been shown to decrease with age, particularly among females, it also appears to be important to identify specific reasons why students lose motivation to be physically active (Salvy et al., 2009).

References

Annesi, J. J. (2002). Goal-setting protocol in adherence to exercise by Italian adults. *Perceptual and Motor Skills, 94*, 453–458.

Armstrong, N., & Welsman, J. (1997). *Young people and physical activity*. Oxford, England: Oxford University Press.

Burton, D. (1992). The Jeckyl/Hyde nature of goals: Reconceptualizing goal setting in sport. In T. S. Horn (Ed.), *Advances in sport psychology* (pp. 267–297). Champaign, IL: Human Kinetics.

Cale, L., & Harris, J. (2009). Fitness testing in physical education: A misdirected effort in promoting healthy lifestyles and physical activity? *Physical Education & Sport Pedagogy, 14*(1), 89.

Cale, L., Harris, J., & Chen, M. (2007). More than 10 years after "The Horse is Dead…": Surely it must be time to "Dismount"?! *Pediatric Exercise Science, 19*, 115–131.

California Department of Education. (2003). Physical fitness testing. Retrieved January 16, 2012, from http://www.cde.ca.gov/ta/tg/pf/

California Department of Education. (2012, February 17). Questions and answers regarding high school physical education exemptions. Retrieved from http://www.cde.ca.gov/ta/tg/pf/pftqanda.asp

Chase, M. A. (2001). Children's self-efficacy, motivational intentions, and attributions on physical education & sport. *Research Quarterly for Exercise & Sport, 72*, 47–61.

The Cooper Institute. (2010, July 7). Fitnessgram/Activitygram. Retrieved from http://www.cooperinstitute.org

Corbin, C. (2002). Physical activity for everyone: What every physical educator should know about promoting physical activity. *Journal of Teaching in Physical Education, 21*, 128–144.

Corbin, C. (2009). Appropriate uses of physical fitness tests. *The CAHPERD Journal, 72*(1), 20–26, 40.

Correa, U., & Souza, O. (2009). Effects of goal difficulty and temporality in motor skill acquisition using the bachman ladder. *Perceptual and Motor Skills, 109*(3), 817–823.

Domangue, E., & Solmon, M. (2010). Motivational responses to fitness testing by award status and gender. *Research Quarterly for Exercise & Sport, 81*(3), 310.

Fahlman, M., Hall, H., & Lock, R. (2006). Ethnic and socioeconomic comparisons of fitness, activity levels, and barriers to exercise in high school females. *Journal of School Health, 76*(1), 126.

Grieser, M., Neumark-Szteainer, D., Saksvig, B., Lee, J., Felton, G., & Kubiki, M. (2008). Black, Hispanic, and White girls' perceptions of environmental and social support and enjoyment of physical activity. *Journal of School Health, 78*(6), 314.

Harris, J., & Cale, L. (2007). Children's fitness testing: A feasibility study. *Health Education Journal, 66*(2), 153–172.

Hoelscher, D., Barroso, C., Springer, A., Castrucci, B., & Kelder, S. (2009). Prevalence of self-reported activity and sedentary behaviors among 4th, 8th, and 11th grade Texas public school children: The School Physical Activity and Nutrition Study. *Journal of Physical Activity & Health, 6*(5), 535–547.

Keating, X. D., & Silverman, S. (2004). Teachers' use of fitness tests in school-based physical education programs. *Measurement in Physical Education and Exercise Science, 8*(3), 145–165.

Kelly, E. B., Parra-Medina, D., Pfeiffer, K. A., Dowda, M., Conway, T. L., Webber, L. S., . . . Pate, R. R. (2010). Correlates of physical activity in Black, Hispanic, and White middle school girls. *Journal of Physical Activity, and Health, 7*, 184–193.

Kowalski, K. C., Crocker, P. R. E., & Donen, R. M. (2004). *The Physical Activity Questionnaire for Older Children (PAQ-C) and Adolescents (PAQ-A) manual*. Saskatoon, Canada: College of Kinesiology University of Saskatchewan.

La Clair, K. W. (1994). *The influence of goal setting on individual endurance swimming performance* (Master's thesis). Retrieved from http://himpubs.uoregon.edu/

Lloyd, M., Colley, R., & Tremblay, M. (2010). Advancing the debate on 'fitness testing' for children: Perhaps we're riding the wrong animal. *Pediatric Exercise Science, 22*(2), 176.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and performance*. Englewood Cliffs, NJ: Prentice Hall.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin, 90*, 125–152.

Mahar, M., & Rowe, D. (2008). Practical guidelines for valid and reliable youth fitness testing. *Measurement in Physical Education and Exercise, 12*(3), 126–145.

Mento, A., Cartledge, N., & Locke, E. (1980). Maryland vs. Michigan vs. Minnesota: Another look at the relationship of expectancy and goal difficulty to task performance. *Organizational Behavior and Human Performance, 25*, 419–440.

Morrow, J., & Freedson, P. (1994). Relationship between habitual physical activity and aerobic fitness in adolescents. *Pediatric Exercise Science, 6*, 315–329.

Morrow, J., Weimo, Z., Franks, B., Meredith, M., & Spain, C. (2009). 1958–2008: 50 years of youth fitness tests in the United States. *Research Quarterly for Exercise & Sport, 80*(1), 1.

Ontario Physical and Health Education Association. (2006). The purpose of fitness assessment in a quality health and physical education program: Six key messages developed by OASPHE. Retrieved January 23, 2012, from http://www.ophea.net

Pangrazi, R. (2000). Promoting physical activity for youth. *Australian Council for Health, Physical Education, Recreation, and Healthy Lifestyles Journal, 47*(2), 18–21.

Pangrazi, R., & Corbin, C. (2008). Factors that influence physical fitness in children and adolescents. In G. Welk & M. Meredith (Eds.), *Fitnessgram/Activitygram reference guide* (pp. 52–60). Dallas, TX: The Cooper Institute.

Pate, R. R. (1993). Physical activity assessment in children and adolescents. *Critical Review Food Science Nutrition, 33*(4), 321–326.

Rice, C. (2007) Becoming "the fat girl": Acquisition of an unfit identity. *Women's Studies International Forum, 30*, 158–174.

Rowland, T. W. (1995) The horse is dead; let's dismount. *Pediatric Exercise Science, 7*, 117–120.

Salvy, S., Roemmich, J., Bowker, J., Romero, N., Stadler, P., & Epstein, L. (2009). Effect of peers and friends on youth physical activity and motivation to be physically active. *Journal of Pediatric Psychology, 34*(2), 217–225.

Shilts, M. K., Horowitz, M., & Townsend, M. S. (2004). Goal setting as a strategy for dietary and physical activity behavior change: A review of the literature. *American Journal of Health Promotion, 19*, 81–93.

Silverman, S., Keating, X. D., & Phillips, S. R. (2008). A lasting impression: A pedagogical perspective on youth fitness testing. *Measurement in Physical Education and Exercise Science, 12*, 146–166.

Tubbs, M. E. (1986). Goal setting: A meta-analytic examination of the empirical evidence. *Journal of Applied Psychology, 71*, 474–483.

Weinberg, R. (1994). Goal setting and performance in sport and exercise settings: A synthesis and critique. *Medicine & Science in Sports & Exercise, 26*(4), 469–477.

Weinberg, R. (2010). Making goals effective: a primer for coaches. *Journal of Sport Psychology in Action, 1*, 57–65.

Wiersma, L. D., & Sherman, C. P. (2008). The responsible use of youth fitness testing to enhance student motivation, enjoyment, and performance. *Measurement in Physical Education and Exercise Science, 12*, 167–183.

Wrench, A., & Garrett, R. (2008). Pleasure and pain: Experiences of fitness testing. *European Physical Education Review, 14*(3), 325–346.

Yoo, S., Lounsbery, M., Bungum, T., & Gast, J. (2010). Adolescent gender and ethnicity differences in physical activity perceptions and behavior. *Journal of Physical Activity & Health, 7*(6), 770–775.