

Methods Forum

Conducting Quasi-Experimental Research in an Urban School District: A Closer Look at the Challenges and Implications

Joanne Genova Carman

University of North Carolina at Charlotte

Andria Parrott

*University of San Diego
Institute for Entrepreneurship in Education*

In fall of 2014, a large, urban public school district partnered with a foundation to pilot a health and wellness program in ten elementary schools. Through this partnership, health and wellness instructors were trained, hired, and deployed to each of the ten schools to lead students through a health and wellness curriculum, designed to promote student physical and mental health and wellness. This methods forum article presents a summary of the challenges and lessons learned during the first year quasi-experimental research study of the health and wellness program, including challenges with recruitment, missing data, reliability and validity issues. The article also describes changes and improvements that were made for evaluation study in the second year, and highlights the implications for future school-based research.

Keywords: quasi-experimental research, health and wellness, evaluation

In reflecting upon trends in evaluation, Hatry, Newcomer, and Wholey (2010) describe how policymakers and administrators will, very likely, continue to focus on research and evidence-based practices. This is particularly true with respect to education evaluation and research in schools, as there continues to be a growing interest in identifying effective, evidence-based programs to implement in schools (Kennedy, 2016; Ritter, Barnett, Denny & Albin, 2009; Slavin & Lake, 2008; Slavin, Lake, Chambers, Cheung & Davis 2009; Stemler & DePascale, 2016). Yet, research in the school setting can be fraught with challenges (Spada, 2005). The purpose of this methods forum article is to describe the challenges and lessons learned during the first year of a three-year research study of a large-scale health and wellness program, as well as discuss the implications for future school-based research studies.

INTRODUCTION

Given the applied nature of evaluation research, it is not surprising that the success of an evaluation research study is dependent upon a number of factors, such as: the strength of the study's research design (Babbie, 2012; Jung & Pirog, 2014; Murnane & Willett, 2011), the study's size and scope (Bamberger, Rugh, Church, & Fort, 2004; Posavac, 2011), and the support of organizational leaders and other stakeholders (Alaimo, 2008; Bryson & Patton, 2010; Thomas, 2010). Previous research suggests that conducting evaluation research in the school setting is particularly challenging (Samaroo, Dahya, & Alidina, 2013; Sen, Mullick, Chakrabarti, & Das, 2015). As Spada (2005) observes "Schools are complex places, with participants who have different roles and responsibilities and a wide range of goals, needs, and interests that are not necessarily compatible with those of the researcher" (p. 328).

In addition, many research studies of school-based interventions have small sample sizes and they are limited in terms of their size and scale, typically having short-term programmatic interventions, limited tracking or follow up over time, or being very narrow in focus (Budd & Volpe, 2006; Fitzgibbon, Stolley, Schiffer, Van Horn, Kaufer-Christoffel, & Dyer, 2005; Griffin, Holliday, Frazier & Braithwaite, 2009; le Roux, Swartz & Swart, 2014; Rahill & Teglassi, 2003). Other school-based studies have encountered challenges in recruiting and retaining study participants or suffered from selection biases (Bowes, Marquis, Young, Holowaty, & Isaac, 2009; Jung & Pirog, 2014; Nabors, 2003; Robinson, Ruch-Ross, Watkins-Ferrell, & Lightfoot, 1993; Ward-Begnoche, Gance-Cleveland, Harris, & Dean, 2008). The purpose of this article is to describe the challenges and lessons learned during the first year of a three-year research study of a large-scale health and wellness program.

Context of the Study

In fall of 2014, a large, urban public school district partnered with a foundation to pilot a health and wellness program in ten high poverty K-5 elementary schools. A fulltime health and wellness instructor was hired at each of the ten schools and trained to deliver a health and wellness curriculum (developed by the foundation) designed to promote student physical and mental health and wellness. For a minimum of 30 minutes, twice a week, health and wellness instructors led students through mindfulness practices, breathing strategies, and yoga-based movement. The theory behind the program is that ongoing yoga and mindfulness practice ultimately reduces students' anxiety and increases their ability to self-regulate (Napoli, Krech, & Holley, 2005; Semple, Reid & Miller, 2005; Wall, 2005).

While researchers have studied the benefits of health and wellness programs in various contexts with adults (Alexandera, Innes, Selfe, & Brown, 2013; Morgan, 2014; Shirey, 2007; Smith & Pukall, 2009), the findings from studies of health and wellness programs for youth have been mixed. While there have been randomized-control trials conducted, reviews of these studies describe how the studies have been small in size and the evidence is still preliminary (Birdee, Yeh, Wayne, Phillips, Davis & Gardiner, 2009; Ferreira-Vorkapic, Feitoza, Marchioro, Simões, Kozasa, & Telles, 2015; Graziano, Reavis, Keane, & Calkins, 2007; Gumora & Arsenio, 2002; Khalsa, Hickey-Schultz, Cohen, Steiner, & Cope, 2012; Pascoe & Bauer, 2015; Pilkington, Kirkwood, Rampes, & Richardson, 2005). Other studies have only looked at short-term health and wellness interventions (Benavides & Caballero, 2009; Uebelacker, Tremont, Epstein-Lubow,

Gaudio, Gillette, Kalibatseva, & Miller, 2010), and some studies were more qualitative in nature (Conboy, Noggle, Frey, Kudesia, & Khalsa 2013; Spinazzola, Rhodes, Emerson, Earle, & Monroe, 2011). Though findings from many of these studies were positive, the general consensus is that more research is needed (Greenberg & Harris, 2012; Serwacki & Cook-Cottone, 2012). To that end, this study was intentionally designed to be large in terms of size, scale, and duration, with 10 intervention schools, five comparison schools, and three years of research, and it was intentionally designed to be both qualitative and quantitative in nature.

Purpose of the Study

Typically, researchers acknowledge the limitations of their studies, but these descriptions are often brief. In contrast, this paper looks closely at the first year of the health and wellness research study, and presents the findings for two research questions: 1) “What are the challenges associated with conducting a large, quasi-experimental research study of a health and wellness program in an urban school district?” 2) What lessons were learned that have implications for future school-based research? By sharing these findings, the goal is for future research studies of similar (or even different) school-based programs to be strengthened.

RESEARCH METHODS

To understand the impact of the health and wellness program, the foundation provided funding for a three-year research study designed to track the implementation of the program and measure student outcomes. The program has three years’ worth of dedicated funding, supportive school leaders, and a flexible evaluation team that is working with teachers and students in the classrooms to collect quantitative and qualitative data. Of the 1,674 students participating in the study during the first year, 1,269 were students in the intervention group (at 10 schools) and 405 students were students in the comparison group (at five schools).

Quantitative methods. A quasi-experimental, pre-post research design with matched comparison schools is being used to study the program’s outcomes over time, using data gathered by the school district and surveys of students. The initial design for the study incorporated a combination of primary and secondary quantitative data including: standardized reading and math test scores; attendance data; numbers of in-school and out-of-school suspensions; surveys of students’ response to stress; physical fitness assessment data; and resting heart rate data gathered from the students.

Qualitative methods. A qualitative, case study design was used to document the health and wellness program’s implementation during the first year of the program, and focus groups, interviews, and observations were used to study the program’s implementation across the ten schools. The design incorporated primary data including: interviews with district personnel, principals, and health and wellness instructors; focus groups with classroom teachers, health and wellness instructors and parents; and classroom observations to assess program fidelity.

FINDINGS

With respect to the first research question about challenges associated with conducting a large, quasi-experimental research study in an urban school district, four challenges emerged relating to recruitment, missing data, reliability and validity.

Challenges with Recruitment

School district personnel intentionally selected specific schools to participate in the study because they were led by principals who were amenable to implementing the program. They were also geographically dispersed throughout the district. Initially, the research team matched the ten intervention schools with ten comparison schools using enrollment, academic, demographic, socioeconomic, and geographic data maintained by the district. While ten comparison schools were identified and invited to participate in the study, only five of those schools agreed to participate, limiting the number of comparison students for the study.

Within each intervention school, a full-time health and wellness instructor was committed to recruiting students to participate in the study offering a dedicated on-site person invested in the program's success. At comparison schools, however, school administrators solicited teachers to recruit students to participate, and the effort in which comparison school teachers offered to recruit students varied from school to school and from class to class. As a result, response rates at intervention schools ranged from 70-100 percent, while response rates at comparison schools ranged from 24-95 percent. This difference left researchers with even fewer comparison students for the study.

For some of the wellness and comparison school matches, the number of students participating in the study was similar (e.g., 60 third grade students at an intervention school and 66 third grade students at the comparison school). For others, the numbers varied considerably (e.g., 95 third grade students at an intervention school and 22 third grade students at the comparison school). Ratios between intervention and comparison students ranged from a low of .23 and a high of 1.10 for third grade students, and a low of .14 and a high of .83 for fifth grade students. Overall, only 37 percent of eligible students were recruited at the comparison schools compared with 86 percent at the intervention schools.

Moreover, the small samples from the comparison schools were not representative of the school's population. Means tests and analysis of variance tests revealed small, but significant differences between the intervention students and the comparison students with respect to important demographic characteristics (i.e., ethnicity, gender, economically disadvantaged), as well as key baseline measures (i.e., math achievement, reading achievement, and attendance).

Challenges with Missing Data

During the first year of the study, the researchers had to deal with missing data in all datasets. Data were missing for four reasons: 1) recruitment challenges restricted the timeline with which researchers could collect data; 2) students were missing during scheduled data collection visits due to absenteeism, transiency, or other school-related matters; 3) students accidentally or

intentionally missed items on the survey; or 4) data collected and provided by the district was incomplete.

Throughout the first year of the study, the researchers worked with school administrators and teachers to allocate time to administer a survey to students about their experiences in school. Scheduling at the intervention schools allowed for students to be surveyed three times over the course of the school year (once near beginning of the school year, once during the middle of the school year and once near the end of the school year). Delays and scheduling issues with the comparison schools made it so that comparison students were only surveyed twice (once two months into the school year and once near the end of the year). This resulted in slightly different times for data collection between intervention and comparison participants, as well as the absence of a midyear collection period for comparison schools.

At the end of the first year of data collection, survey data (data collected at the beginning of the year and data collected at the end of the year) were matched using a unique identifier for each student. During data analysis, researchers discovered that student absenteeism (or unit nonresponse) was quite problematic. More than 20% of participating students were missing at least one survey. Furthermore, among the students for whom there were two surveys, 30% were missing data for key questions that comprised the subscales embedded into the instrument (or item nonresponse) (Ritter & Sue, 2007). While imputation procedures were used to ameliorate the item nonresponse (substituting the mean value when possible), researchers did this only under certain circumstances.¹

Additionally, some of the data that were supposed to be collected or provided by the school district simply weren't available. Fitnessgram data, for example, were supposed to be collected by school personnel for all students in the study. At the conclusion of the school year, researchers received data for only a small sub-sample of students from three schools at different points in time. Follow up inquiries with district personnel revealed that these data weren't collected for most students, so they were ultimately not available for the study. Further, for attendance, behavior, and academic achievement outcomes, researchers did not receive "pre" program measures, as expected, so there was no way to determine if there had been any changes with school-based outcomes during the first year of the program. Finally, there was missing data due to misunderstandings between school personnel, district personnel, and the researchers with regarding data collection scheduling.

Challenges with Reliability

In addition to incomplete or missing data, several reliability issues were discovered during the first year of the study. These issues primarily involved the measures researchers used to assess the outcomes of the health and wellness program.

Early into the data collection process, for example, researchers realized that the reliability of the resting heart rate data was going to be an issue. Pulse Oxymeters were being used to measure the students' resting heart rate by placing the device on students' index finger, with three measurements being taken while researchers were in classrooms with students

¹ For example, each subscale was comprised of three questions (e.g., Q1, Q2, Q3). If the response to one of the three questions was missing, the researchers calculated the average of the other two questions and used that value. (e.g., if Q1 = missing, Q2 = 1, and Q3 = 2, then Q1 = 1.50). If two responses were missing, researchers did not use the survey.

administering the survey. Throughout the school year, variations in the timing of the measurements resulted when teachers and administrators requested schedule changes, students changed classrooms, or activities preceding data collection were introduced (i.e. physical education or recess) that affected student heart rate. There were also questions about the accuracy of the measurements due to the frequent presence of outliers and/or data coding errors.

Reliability was also an issue for some of the subscales on the survey as it was administered to both third and fifth grade students during the first year of the study. While the reliability for the two overarching scales (involuntary engagement and primary control engagement coping scale) was fairly good (for third grade, Cronbach's alpha levels = .863 and .702; for fifth grade, Cronbach's alpha levels = .889 and .760), the reliability for select subscales, such as physiological arousal, emotional regulation, and emotional expression (key measures of interest), were much lower for both third and fifth graders (ranging from .443 to .600). This led researchers to interpret subscale findings with caution going forward.

With respect to the behavioral data, the study's initial design planned for relying on district collected behavioral data, specifically the frequency of in-school and out-of-school suspensions to understand potential changes in student behavior. Data analysis of these data revealed that the frequency of behavior incidents was very low. Four of the 15 schools reported having no in-school or out-of-school suspensions throughout the year. Among the schools that did have in-school or out-of-school suspensions, these were infrequent and typically could be attributed to just a few students. When there were greater numbers of suspensions, the findings showed that the suspensions were concentrated at just two schools.

After comparing the district collected behavioral data with observational data, contradictions emerged. Researchers noted that students were frequently placed in "in-school suspension" rooms or taken out of the classroom quite frequently leaving them unavailable for survey administration or absent from health and wellness class. Yet, these "in-school suspensions" happened at the schools where no suspension incidents were reported by the district. This suggests the decision to report behavior incidents is left to school personnel, and data can vary from school to school or, in some cases, not reflect the students' behavior at a particular school.

Challenges with Validity

Issues relating to internal validity including attrition, contamination and seasonality emerged during the first year of the study. Throughout all ten schools, the issues of attrition and contamination emerged (Fife-Shaw, 2012; Mark & Reichardt, 2009). For example, some students moved in and out of the district during the course of the school year. Researchers also encountered students at one intervention school who they had initially surveyed at another intervention school. A few students also moved from the comparison schools to the intervention schools, and from the intervention schools to the comparison schools.

As noted previously, seasonality was also an issue because of the inconsistent timing of the survey data collection (Mark & Reichardt, 2009). While the students in the intervention schools were surveyed at the beginning of the school year in September, challenges in recruiting comparison schools caused the first wave of data collection at the comparison schools to take place in December.

LESSONS LEARNED

After the first year of data collection, the researchers implemented a number of strategies to improve data collection and manage these challenges.

Improving Recruitment

To address the recruitment issue, comparison schools were financially compensated to continue participating in the study. In order to preserve the comparative research design and improve recruitment, comparison schools were compensated by the foundation in the second year of the study. Compensation was set at a sliding scale, where the top recruiters (recruiting between 70%-100% of the eligible students) received \$5,000 for the second year of the study and \$6,000 for the third year of the study. The researchers also worked more closely with the teachers to recruit more comparison students. For the second year of the study, recruitment increased by 20 percent. With larger sample sizes, the reliability on the subscales for the survey should also increase. Data collection at the intervention and comparison schools is now occurring at the same time periods due to improved coordination and greater buy-in from the comparison schools.

Decreasing Missing Data

To address the item and unit nonresponse, the researchers began to closely track the number of surveys being completed during data collection periods. When students were unavailable during periods of survey administration, researchers met with school administrators to document why students were unavailable, confirming whether they were absent, withdrawn from the school, or elsewhere that hindered their participation. Additional research staff was also hired to review the surveys for completeness ensuring that any items a student misses were done so intentionally.

To alleviate the limitation of not having baseline academic achievement and fitness data, researchers followed up with district leaders to address the matter for subsequent years of the study. District personnel confirmed that Fitnessgram data will be collected from all intervention and comparison schools during the second year of the study, and district personnel forwarded math and reading standardized test scores, as well as attendance and behavioral data, from the year before the study so that these data can be used to establish a baseline for the longitudinal component of the study.

Improving Reliability

To address the challenge with the reliability of behavior data, the researchers piloted a different way to measure positive behavior change. Teachers at schools (one intervention school and one comparison school) consented to participate in monthly survey rating the behavior of the classes they taught within a day. For one week each month, teachers will fill out a brief survey assessing the behavior of different classes (e.g., they are indicating whether the students were very well behaved, well behaved or poorly behaved on each day) and offering notes to provide context. In addition, the student survey was modified, adding five questions, to measure the extent to which

students are using yoga, mindfulness and breathing tools in and outside of the school environment. With respect to the resting heart rate measures, the researchers decided to discontinue collecting those data.

Improving Validity

To improve the validity of the study, the researchers have worked more closely with the schools to ensure that data are collected at the same time in the intervention and comparison schools within each data collection cycle. The researchers are also working more closely with teachers to identify new students and students who have moved in order to understand patterns of attrition and track the potential for contamination.

IMPLICATIONS FOR FUTURE RESEARCH

While there is growing support for adopting health and wellness programs in schools in order to mitigate obesity, physical fitness, and mental health issues (Agron, Berends, Ellis & Gonzalez, 2010; Hagen & Nayar, 2014; Hyde, 2012), the research about the benefits of these programs has been mixed and well-designed, large scale studies are needed (Serwacki & Cook-Cottone, 2012). Yet, large-scale studies can be challenging because they are not conducted “under controlled laboratory conditions” (Hatry & Newcomer, 2010, p. 557). The experiences from the first year of this study illustrate how challenging research in the school setting can be, especially for large scale studies in an urban school district.

The challenging experience of recruiting comparison schools and participants illustrates the importance of offering financial incentives. In addition, having a full-time program or research advocate on the school’s campus can help to generate more favorable recruitment efforts. When working with so many schools and classrooms, being consistent with data collection procedures can be difficult. Data collection procedures may need to be improved, measures may need to be changed, and staffing may need to be changed in order to improve the reliability, validity and credibility of the study (Hatry & Newcomer, 2010). Moreover, asking children to complete surveys and offer self-reported data is impacted by a number of factors, many of which are related to their development. Given absenteeism is a challenge in large, urban districts, adopting strategic data collection procedures to minimize missing data and keeping record of why data are missing can help researchers to better understand the populations with which they are working and the data they are analyzing.

Finally, the experiences of this study illustrate how difficult it is to identify “what works” on a larger scale. Collecting data from large groups of students in schools, over time, can be labor and time-intensive, and it takes coordination and considerable support from teachers, principals, parents, and students. This study, however, illustrates how some relatively easy adjustments, such as investing more in staffing the data collection efforts, having stakeholders understand the importance of capturing baseline data, changing out measures, if need be, and compensating comparison groups can go a long way toward improving school-based research.

REFERENCES

- Agron, P., Berends, V., Ellis, K. & Gonzalez, M. (2010). School wellness policies: Perceptions, barriers, and needs among school leaders and wellness advocates. *Journal of School Health, 80*(11), 527–535.
- Alaimo, S. P. (2008). Nonprofits and evaluation: Managing expectations from the leader's perspective. (pp. 73–92). In J. Carman & K. Fredericks (Eds.) *Nonprofits and Evaluation: New Directions for Program Evaluation*, 119, San Francisco, CA: Jossey Bass.
- Alexander, G.K., Innes, K.E., Selfe, T.K., & Brown, C. J. (2013). "More than I expected": perceived benefits of yoga practice among older adults at risk for cardiovascular disease. *Complementary Therapies in Medicine, 21*(1),14-28.
- Benavides, S. & Caballero, J. (2009). Ashtanga yoga for children and adolescents for weight management and psychological well being: An uncontrolled open pilot study. *Complementary Therapies in Clinical Practice, 15*(2), 110–114.
- Babbie, E. (2012). *The practice of social research*. (13th Edition) Cengage Learning.
- Bamberger, M., Rugh, J., Church, M., & Fort, L. (2004). Shoestring evaluation: Designing impact evaluations under budget, time and data constraints. *American Journal of Evaluation, 25*(1), 5–37.
- Birdee, G. S., Yeh, G. Y., Wayne, P. M., Phillips, R. S., Davis, R. D. & Gardiner, P. (2009). Clinical applications of yoga for the pediatric population: A systematic review. *Academic Pediatrics, 9*(4), 212-220.
- Bowes, D., Marquis, M., Young, W., Holowaty, P., & Isaac, W. (2009). Process evaluation of a school-based intervention to increase physical activity and reduce bullying. *Health Promotion Practice, 10*(3), 394-401.
- Bryson, J. M. & Patton, M. Q. (2010). Analyzing and engaging stakeholders. (pp. 30-54). In J. Wholey, H. Hatry, & K. Newcomer (Eds.). *The handbook of practical program evaluation*. (3rd Edition). San Francisco, CA: Jossey-Bass.
- Budd, G. M. & Volpe, S. L. (2006). School-based obesity prevention: Research, challenges, and recommendations. *Journal of School Health, 76*(10), 485–495.
- Conboy, L.A., Noggle, J.J., Frey, J.L., Kudesia, R. S. & Khalsa, S. B. S. (2013). Qualitative evaluation of a high school yoga program: Feasibility and perceived benefits. *Explore, 9*(3), 171-180.
- Ferreira-Vorkapic, C., Feitoza, J. M., Marchioro, M., Simões, J., Kozasa, E., & Telles, S. (September 28, 2015). Are there benefits from teaching yoga at schools? A systematic review of randomized control trials of yoga-based interventions. *Evidence-based Complementary & Alternative Medicine (ecam)*, 1-17, Retrieved: <http://dx.doi.org/10.1155/2015/345835>
- Fife-Shaw, C. (2012). Quasi-experimental designs. In G. M. Breakwell, J. A. Smith, & D. B. Wright (Eds.). *Research methods in psychology* (4th edition.). pp. 75-92. Thousand Oaks, CA: Sage Publications.
- Fitzgibbon, M. L., Stolley, M. R., Schiffer, L., Van Horn, L., Kaufer-Christoffel, K., & Dyer, A. (2005). Two-year follow-up results for Hip-Hop to Health Jr.: A randomized controlled trial for overweight prevention in preschool minority children. *The Journal of Pediatrics, 146*(5) 618-625
- Graziano, P., Reavis, R., Keane, S., & Calkins, S. (2007). The role of emotion regulation and the student-teacher relationship in children's academic success. *Journal of School Psychology, 45*(1), 3-19.
- Greenberg, M., & Harris, A. (2012). Nurturing mindfulness in children and youth: Current state of research. *Child Development Perspectives, 6*(2), 161-166.
- Griffin, Jr., J.P., Holliday, R. C., Frazier, E. & Braithwaite, R. L. (2009). The BRAVE (Building Resiliency and Vocational Excellence) program: Evaluation findings for a career-oriented substance abuse and violence preventive intervention. *Journal of Health Care for the Poor and Underserved, 20*(3), 798-816.
- Gumora, G. & Arsenio, W. (2002). Emotionality, emotional regulation, and school performance in middle school children. *Journal of School Psychology, 40*(5), 395-413.
- Hagen, I., & Nayar, U. S. (2014). Yoga for children and young people's mental health and well-being: Research review and reflections on the mental health potentials of yoga. *Frontiers in Psychiatry, 2*, 5-35.
- Hatry, H.P. & Newcomer, K.E. (2010). Pitfalls of Evaluation. In K. E. Newcomer, H. P. Hatry, & J.S. Wholey, (Eds.), *The handbook of practical program evaluation*. (3rd Edition). (pp. 557-580). San Francisco, CA: Jossey Bass.
- Hyde, A. M. (2012). The yoga in schools movement: using standards for educating the whole child and making space for teacher self-care. In J.A. Gorlewski, B. Porfilio & D.A. Gorlewski. *Using standards and high-stakes testing for students: Exploiting power with critical pedagogy*. New York, NY: Peter Lang Publishing, Inc.
- Jung, H. & Pirog, M. A. (2014). What works best and when: Accounting for multiple sources of pureselection bias in program evaluations? *Journal of Policy Analysis and Management, 33*(3), 752–777.

- Kennedy, M. (2016). How does professional development improve teaching? *Review of Educational Research*, Retrieved from: <https://msu.edu/~mkennedy/publications/docs/Teacher%20Learning/HowDoesPDfoosterTL/KennedyInPressRERHowDoesPD.pdf>
- Khalsa, S. B. S., Hickey-Schultz, L., Cohen, D., Steiner, N., & Cope, S. (2012). Evaluation of the mental health benefits of yoga in a secondary school: A preliminary randomized controlled trial. *The Journal of Behavioral Health Services & Research*, 39(1), 80-90.
- le Roux, M. C., Swartz, L., & Swart, E. (2014). The effect of an animal-assisted reading program on the reading rate, accuracy and comprehension of grade 3 students: A randomized control study. *Child & Youth Care Forum*, 43(6), 655-673.
- Morgan, M. (2014). The feasibility of a holistic wellness program for HIV/AIDS patients residing in a voluntary inpatient treatment program. *Journal of Holistic Nursing*, 32(1), 54-60.
- Murnane, R. J. & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research*. New York, NY: Oxford University Press.
- Mark, M.M. & Reichardt, C.S. (2009). Quasi-Experimentation. In L. Bickman and D. Rog (Eds.), *The SAGE handbook of applied social science research methods*. (182-213). USA: SAGE Publications, Inc.
- Nabors, L. A. (2003). Evaluation in school-based health centers. *Psychology in the Schools*, 40(3), 309-320.
- Napoli, M., Krech, P., & Holley, L. (2005). Mindfulness training for elementary school students: the attention academy. *Journal of Applied School Psychology*, 21(1) 99-125.
- Rahill, S. A. & Teglassi, H. (2003). Processes and outcomes of story-based and skill-based social competency programs for children with emotional disabilities, *Journal of School Psychology*, 41(6), 413-429.
- Ritter, L. A. & Sue, V. M. (2007), Conducting the survey. *New Directions for Evaluation*, Issue 115, *Using Online Surveys in Evaluation* pp. 47-50.
- Ritter, G. W., Barnett, J. H., Denny, G. S. & Albin, G. R. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A meta-analysis. *Review of Educational Research*, 79(1), 3-38.
- Robinson, W.L., Ruch-Ross, H.S., Watkins-Ferrell, P., & Lightfoot, S. (1993). Risk behavior in adolescence: Methodological challenges and school-based research. *School Psychology Quarterly*, 8, 241-254.
- Pascoe, M. C. & Bauer, I. E. (2015). A systematic review of randomised control trials on the effects of yoga on stress measures and mood. *Journal of Psychiatric Research*, 68, 270-282.
- Pilkington, K., Kirkwood, G., Rampes, H. & Richardson, J. (2005). Yoga for depression: The research evidence. *Journal of Affective Disorders*, 89(1), 13-24.
- Posavac, E. (2011). *Program evaluation: Methods and case studies* (8th edition), Boston, MA: Pearson.
- Samaroo, Dahya, & Alidina, (2013). Exploring the challenges of conducting respectful research: Seen and unforseen issues within urban school research. *Canadian Journal of Evaluation*, 36(3), 438-457.
- Semple, R., Reid, E., & Miller, L. (2005). Treating anxiety with mindfulness: An open trial of mindfulness training for anxious children. *Journal of Cognitive Psychotherapy: An International Quarterly*, 19, 379-392.
- Sen, P., Mullick, P. D., Chakrabarti, G., & Das, A. (2015). Emerging challenges and consequences of evaluation process in school education: An ethnographic insight. *International Journal of Multidisciplinary Approach*, 2(1), 276-288.
- Serwacki, M.L. & Cook-Cottone, C. (2012). Yoga in the schools: A systematic review of the literature. *International Journal of Yoga Therapy*, 22(1), 101-109.
- Shirey, M. R. (2007). An evidence-based solution for minimizing stress and anger in nursing students. *Journal of Nursing Education*, 46(12), 568-571.
- Slavin, R. E. & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research*, 78(3), 427-515.
- Slavin, R. E., Lake, C., Chambers, B., Cheung, A. & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research*. 79(4), 1391-1466.
- Smith, K. B. & Pukall, C. F. (2009). An evidence-based review of yoga as a complementary intervention for patients with cancer. *Psychooncology*, 18(5), 465-475.
- Spada, N. (2005), Conditions and challenges in developing school-based SLA research programs. *The Modern Language Journal*, 89(3), 328-338
- Spinazzola, J. Rhodes, A.M., Emerson, D., Earle, E. & Monroe, K.(2011). Application of yoga in residential treatment of traumatized youth. *Journal of the American Psychiatric Nurses Association*, 17(6) 431- 444.
- Stemler, S. S. & DePascale, M. (2016). Aligning mission and measurement. (pp. 57-92). In A. A. Lipnevich, F. Preckel & R. D. Roberts (Eds.). *Psychosocial skills and school systems in the 21st century: Theory, research, and practice*. Switzerland: Springer International Publishing.

- Thomas, J. C. (2010). Outcomes assessment and program evaluation. (pp. 401-430). In D. Renz (Ed.). *The Jossey Bass Handbook of nonprofit management and leadership*. San Francisco, CA: Jossey Bass.
- Uebelacker, L. A., Tremont, G., Epstein-Lubow, G., Gaudiano, B. A., Gillette, T., Kalibatseva, Z., & Miller, I. W. (2010). Open trial of Vinyasa yoga for persistently depressed individuals: Evidence of feasibility and acceptability. *Behavior Modification, 34*(3), 247–264.
- Wall, R. B. (2005). Tai chi and mindfulness-based stress reduction in a Boston public middle school. *Journal of Pediatric Health Care, 19*, 230–237.
- Ward-Begnoche, W. L., Gance-Cleveland, B., Harris, M. M., & Dean, J. (2008). Description of the design and implementation of a school-based obesity prevention program addressing needs of middle school students. *Journal of Applied School Psychology, 24*(2), 247-263