

## Using Clinical Teaching to Increase Student Achievement in High-Needs, Urban, Partnership Schools

Gwendolyn Benson  
Georgia State University

William Curlette  
Georgia State University

Susan Ogletree  
Georgia State University

Robert Hendrick  
Georgia State University

**Abstract:** In this research, a clinical teaching approach in a Professional Development School (PDS) partnership was employed to prepare student teachers in urban high-needs partnership schools. Though there are many qualitative studies that indicate an increase in student achievement in PDSs, few quantitative studies have been published. The clinical teaching used Teacher-Intern-Professor (TIP) groups with an Anchor Action Research (AAR) model to help evaluate the PDS teacher intern preparation efforts by measuring student achievement on pre- and post-assessment scores. Within the PDS partnership, a meta-analysis using random effects pre-post-control (PPC) model was used to summarize achievement differences between the TIP and comparison classes. The results showed that Hedges'  $g$  effect size between the classes is .326 in favor of TIP classes, which is a typical effect size for educational interventions. Thus, this study provides quantitative research to support K-12 student academic achievement through a PDS model.

**KEYWORDS:** professional development schools, PDS, clinical teaching, student achievement

### **NAPDS NINE ESSENTIALS ADDRESSED:**

1. A comprehensive mission that is broader in its outreach and scope than the mission of any partner and that furthers the education profession and its responsibility to advance equity within schools and, by potential extension, the broader community;
4. A shared commitment to innovative and reflective practice by all participants;
5. Engagement in and public sharing of the results of deliberate investigations of practice by respective participants
6. Work by college/university faculty and P-12 faculty in formal roles across institutional settings;

### **Introduction**

PDS partnership in this research was administered through Georgia State University and our HEA partners (Albany State University, Columbus State University, and Georgia Southern University) and served both urban and rural LEAs. Another valuable partner was Learning Forward (formerly National Commission on Teaching and America's Future, NCTAF), which has

provided training and resources for the student teachers and as many as 10 partner urban and rural LEAs. Within the partnership LEAs, the PDS partnership employs an Anchor Action Research (AAR) project used within clinical teaching in Teacher-Intern-Professor (TIP) groups (Curlette & Ogletree, 2011). This PDS approach, TIP, involves a university professor, mentor teacher, and intern working together on a unit of instruction. Specific components of the TIP group include meetings of the TIP group twice per month, a monthly training session for mentor teachers, understanding and implementation of an action research project focused on student achievement, and submission of a detailed report outlining the decision-making process, action research process, and results. The mentor teacher, professor, and intern (or student teacher) have experienced an informal relationship for many years. This TIP group formalizes the relationships among the members, has an AAR project for instructional focus, and brings a greater emphasis to action research in the classroom. The TIP group unites the leadership, specialization, and instructional experience of the mentor teacher and professor with the abilities of the intern to help prepare instruction and facilitate student achievement. The TIP group also meets the five essential characteristics of a professional learning community as defined by Vescio, Ross, and Adams (2008). The five essentials include developing shared values regarding students' abilities to learn, a clear focus on student learning, focused constructive dialogue among teachers regarding student learning and instruction, making teaching public, and enhanced collaboration between and among teachers. TIP groups engage in all those essentials through the AAR project. AAR is a form of action research that is anchored in three aspects:

1. The project is assessed using a pre- and post-assessment design.
2. The project addresses the teaching and learning process in which the teaching of the intern facilitates student learning in the classroom.
3. The project has a comparison condition (Curlette & Ogletree, 2011).

As a part of the TIP group, the intern will participate in the planning and delivering of a unit of instruction that uses a pretest and posttest assessment. In addition, a class, of same subject matter within the same treatment school, will be selected as a comparison class. The comparison class will have the same pre- and post-assessments as the treatment class. However, the treatment class will receive specially designed instruction based upon the planning of the TIP group. The innovative approach to instruction within the PDS can take many forms and is dependent on the expert guidance of the mentor teacher and professor and delivered by the intern. Qualitative studies have indicated the positive effects of PDS on student achievement. However, limited evidence of positive effects using quantitative methods within PDS has been published (Vescio, Ross, & Adams, 2008).

Although some AAR studies have been presented and a summary of the initial 10 studies was published (Curlette, Hendrick, Ogletree, & Benson, 2014), no comprehensive summary, across 25 AARs, has been presented. This report is designed to summarize the overall effectiveness of TIP with AAR to address a gap in the literature related to the evaluation of this approach. Objectives of this article are as follows: (a) to present the summative mean difference effect size of the pre- to post-assessment scores between AAR treatment and comparison classrooms and (b) to provide a discussion which informs clinical teaching practices based on those findings.

### Perspective

A discussion of the benefits of clinical teaching can be found in Bohan and Many's (2011) book, *Clinical Teacher Education: Reflections from an Urban Professional Development School Network*. Another perspective in TIP with AAR is support for the action research approach which values participants' conducting research to improve teaching practices (Hendricks, 2009). The TIP group approach is consistent with Darling-Hammond and Richardson's (2009) position that asserts the importance of professional development in communities of practice.

A brief review of relevant studies establishes the context in which to interpret student achievement effect sizes results. In 2009, The Council of Chief State School Officers (CCSO) released a study in which a number of meta-analysis projects "analyzed evidence on the effects of mathematics and science teacher preparation and development initiatives on student achievement" (p. 3). The CCSO found a mean effect size of .21 between mathematics classes using a pre-and post-assessment model and a mean effect size of .05 between science classes using a pre-and post-assessment model. Many of the studies in this meta-analysis used norm referenced or criterion referenced tests to evaluate student performance. Therefore, the professional development provided to teachers may or may not be cogent to the assessment. "Studies that utilized student measures that are closer to the heart of what the professional development is intended to impact, do report larger effect sizes" (CCSO, 2009, p. 17). This addresses the 2012 NCME paper by Popham and Ryan that examines the "instructional sensitivity" of high-stakes tests. Most high-stakes tests are not sensitive to the precise pedagogy employed in the classroom; however, in some instances, educational decision makers may use students' high stakes test scores to evaluate teaching quality. Both the CCSO report and Popham and Ryan's article caution that, for the purposes of teacher evaluation, instructionally sensitive tests yield more applicable data.

Another recent meta-analysis, conducted by Yoon, Duncan, Lee, Scarloss, and Shapley (2007) for the Southwest Regional Education Laboratory, indicated that PDSs in the elementary grades are more effective than control classes at increasing student achievement. The reported effect size was .54, which is typical of elementary educational intervention studies. For this meta-analysis over 1300 studies were prescreened and nine studies met the What Works Clearinghouse (WWC) evidence standards. In examining other evidence, Sipe and Curlette (1996) conducted a large meta-synthesis of 103 meta-analyses related to education and student achievement that was consistent with Hattie (1992). The findings of both studies indicated effect sizes of .375 for the Sipe and Curlette meta-synthesis and .40 for Hattie. Even though common meta-analyses accounted for only about 10% of overlap between these two large meta-syntheses, the findings of the meta-syntheses were similar.

### Process

The clinical teaching experience for the student teachers is extended to include yearlong placements in urban partnership schools. The student teacher is paired with the mentor teacher during the school's preplanning period. The student teacher to mentor partnership continues through the school's post planning period. During the fall, the intern in conjunction with the other members of the TIP group will choose an AAR unit. The TIP group will discuss and plan the theoretical instructional approach, activities, goals, and duration of the unit. Typically, the unit is 2 to 3 weeks in duration and the intern delivers all instruction during the AAR unit. The mentor

teacher provides daily formative feedback while the professor provides pedagogical consultation. The pretest and posttest are typically equivalent or similar in content and scope. The tests reference subject area goals set forth by the state curriculum, and are consistent with learning assessments that are familiar to the students. A comparison condition is selected in the same subject area, grade, performance level, and student socioeconomic level to help control potential confounding variables. The teacher of the similar unit for the comparison class will not alter the instructional plan for the comparison group, but he or she will teach in a manner that is consistent with effective pedagogy and quality instructional practices. The goal of the TIP group in this AAR unit is to be as effective at influencing student performance as the quality instruction of a veteran teacher in the comparison group. Therefore, the mean gain effect size of student performance in the AAR class would be similar to the mean gain effect size of student performance observed in the comparison class.

Typically, high-stakes tests are designed to assess whether a student understands a defined set of related concepts, which are deemed by experts as appropriate to predict an understanding of the overall subject area at an appropriate level of precision. High-stakes tests generally assess concepts that are considered central to the reasoning, performance, and understanding of the defined subject area. As such, the presentation of test items is limited to those that collectively indicate an overall grasp of the subject matter to minimize testing time and maximize the predictive nature of the test. While this may be an effective manner to assess the performance of the student, it is not designed to assess the quality of the teacher's pedagogy. This is one reason that unit pre- and posttests are used in AAR.

Using this model, each AAR with comparison would typically measure the achievement of between 20 and 60 students depending upon class size, consent, and assent rates. The relatively small sample size of a single AAR limits the statistical power and generalizability of the AAR; however, when multiple AAR studies are analyzed using meta-analysis techniques, the samples from a number of AAR studies are aggregated, providing increased statistical power and greater generalizability.

### **Methods and Data Sources**

The goal for each of the AAR projects was to implement a quasi-experimental design: a pre- and post-assessment involving a treatment group and a comparison group. Some of the interns taught in rural school districts. In these settings, comparison classes were an issue because the teacher was in many cases the only subject matter teacher for a specific grade in the district. Though comparison classes could be assigned, the matching criteria could not be adequately applied to meet the WWC evidence standards. Also, some of the remaining interns were in Special Education assignments, where locating a matching comparison class was problematic and WWC evidence standards were not met because of the lack of a matching control group. From the initial group of AAR studies, we eliminated several studies because the research design did not meet the WWC evidence standards with reservations.

The remaining 31 studies were coded according to the following criteria: (a) the AAR project had a comparison condition, (b) the pre- and post-assessment used the same instrument in both AAR and comparison classrooms, (c) the instruments comprised objective questions that pertained to the targeted AAR unit, and (d) the comparison classroom was similar on student achievement level, gender balance, ethnic composition, and student socio-economic level. A quantitative and qualitative approach accessing AAR reports and data from LiveText was

augmented by conducting student and teacher interviews (Silverman, 2010). Of the 31 studies, three were eliminated because the number of participating students with signed consent and assent documents was low. In two studies, the instrumentation for the pretest and posttest was not equivalent, and in one study the focus for the instructional unit was not reflective of the assessment, or there was a lack of instructional sensitivity (Popham & Ryan, 2012). Therefore, 25 AAR quasi-experimental studies met the WWC evidence standards with reservations and were included in the meta-analysis with each study having a different intern in each AAR.

The method used to analyze the data was a random effects meta-analysis (Cooper, Hedges, & Valentine, 2009; Morris, 2008). In total, 817 individual students' pre- and post- assessment scores were included within the 25 AAR studies. Throughout the considered treatment and comparison classrooms, the number of students ranged from 12 to 52. Studies selected for this report have met the aforementioned criteria confirmed by LiveText, records of interviews, and other documentation.

Qualitative interviews with interns and mentor teachers as well as a collection of artifacts indicated that interns were focused on student engagement, relationship building, relevance of the lesson, effectively scaffolding learning strategies, and how to use action research in the classroom to improve student academic achievement. Qualitative data sources included 45- minute telephone interviews conducted by trained staff members of the Center for Evaluation and Research Services with NET-Q and CREST-Ed Mentor Teachers and Interns, document analysis of Mentor Teacher Monthly Training Sessions, document analysis of bi-weekly Teacher Intern Cohort Meetings, and analyses of class reflection papers submitted through Live Text by the interns. Interviews were transcribed and coded allowing for categories and themes to emerge.

## Results

A comparison of the teaching effectiveness, in terms of student achievement, was analyzed by comparing the relative gains on the pre-to post-assessment scores of the AAR treatment class with the similar gains in scores made by the comparison class. The overall mean difference effect size for the random effects meta-analysis is .326 with confidence interval (.073 to .578) as seen at the bottom of Table 1. The effect size, .326, is a substantial and a statistically significant effect size in favor of the AAR outcome. This finding supports the qualitative research that PDSs have a positive influence on student achievement and is consistent with previous meta-syntheses examining student achievement. Our goal was to show that the PDS teacher preparation using the TIP model and AAR will produce beginning teachers who are as effective as or slightly more effective in facilitating student achievement than teachers in comparison classrooms in a unit of instruction.

Table 1. Statistics for Each AAR Study

Study	Hedge's g	Standard error	Variance	Lower limit	Upper limit	Z-value	p-value
AAR1	0.183	0.29	0.083	-0.38	0.75	0.633	0.527
AAR2	-0.542	0.35	0.120	-1.22	0.14	-1.562	0.118
AAR3	0.456	0.39	0.154	-0.31	1.23	1.161	0.246
AAR4	0.449	0.27	0.073	-0.08	0.98	1.661	0.097
AAR5	0.700	0.30	0.088	0.12	1.28	2.355	0.019
AAR6	0.468	0.30	0.089	-0.12	1.05	1.569	0.117
AAR7	0.373	0.35	0.123	-0.31	1.06	1.066	0.286
AAR8	0.869	0.32	0.100	0.25	1.49	2.754	0.006
AAR9	0.457	0.32	0.104	-0.18	1.09	1.413	0.158
AAR10	0.248	0.28	0.076	-0.29	0.79	0.897	0.370
AAR11	0.502	0.47	0.225	-0.43	1.43	1.059	0.290
AAR12	1.108	0.39	0.150	0.35	1.87	2.862	0.004
AAR13	1.697	0.60	0.359	0.52	2.87	2.833	0.005
AAR14	0.211	0.39	0.149	-0.55	0.97	0.548	0.584
AAR15	-0.553	0.42	0.175	-1.37	0.27	-1.320	0.187
AAR16	0.913	0.36	0.131	0.20	1.62	2.523	0.012
AAR17	-0.582	0.55	0.306	-1.67	0.50	-1.051	0.293
AAR18	-0.538	0.39	0.156	-1.31	0.23	-1.364	0.172
AAR19	1.110	0.31	0.097	0.50	1.72	3.557	0.000
AAR20	-1.852	0.45	0.199	-2.73	-0.98	-4.155	0.000
AAR21	0.024	0.48	0.232	-0.92	0.97	0.050	0.960
AAR22	0.164	0.32	0.102	-0.46	0.79	0.514	0.607
AAR23	-0.001	0.25	0.062	-0.49	0.49	-0.003	0.998
AAR24	0.577	0.48	0.227	-0.36	1.51	1.211	0.226
AAR25	1.528	0.40	0.158	0.75	2.31	3.849	0.000
Random model	0.326	0.13	0.017	0.07	0.58	2.523	0.012

A forest plot (Figure 1) illustrates the weight and mean gain comparison of the 25 studies within the random effects meta-analysis. As shown within the forest plot, there was a statistically significant negative change in only one (AAR 20) of the 25 AAR treatment and comparison class studies and a statistically significant positive change in seven of the studies (AAR 5, 8, 12, 13, 16, 19, 25). Further examination of each of the 25 AAR studies show a majority indicating a positive main effect between pretest and posttest, averaging a gain of 31 percentage points for the treatment group and 26 percentage points for the comparison group. In most cases, the percentage gains from pretest mean to posttest mean were similar for both groups. In the studies with negative effect sizes, typically the pretest mean for the comparison group was more than 10 percentage points lower than the treatment pretest mean. Though, for those studies, the posttest means were about the same, greater growth was indicated by the comparison group leading to a negative effect size. The same is not the case when examining the studies with positive effect sizes. In those studies, the majority of the cases show that the treatment posttest mean is an average of 5 percentage points higher than the comparison posttest mean, which typically results in a positive effect size. The

majority of studies demonstrate a main effect increase in the treatment group. The comparison group main effect shows a matched counterfactual to reference the amount of increase in student performance.

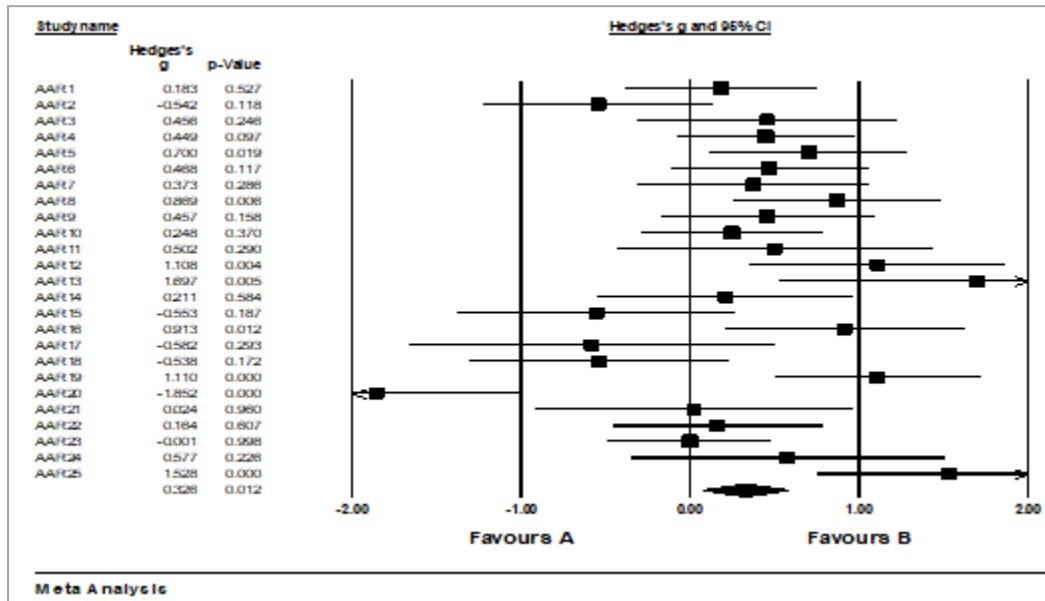


Figure 1. Forest plot showing weight and mean gain.

The importance of a matched counterfactual is noted in AAR 7, in which the overall change for the treatment, from pretest to posttest mean, was negative. This would seem to indicate that the intern had not influenced a gain in student performance. The counterfactual change in the comparison group was also negative, which may indicate an issue with the topic or group. A closer inspection of artifacts relating to AAR 7 shows the students in both the treatment and comparison classes were socially promoted during the previous two times they had attempted this specific course and were being given a remedial curriculum during the AAR 7 that assumed the students had foundational mathematics skills (e.g. multiplication, and division). Many of the students were not successful in the course, just as they had not been successful in that course for the previous two years. An analysis of the change in pretest and posttest scores between treatment and comparison groups resulted in a positive effect size (0.384) and a non-significant result in the individual study due to the low sample size. However, the importance of a matched comparison group is illustrated in this specific study. The random effects meta-analysis assumes the observed estimates of the AAR treatment effect can vary across studies because of different teaching strategies used within each AAR in each study as well as some variability within each class. Such heterogeneity in treatment effects is caused by uncontrolled differences in the target classes, interventions received (teaching strategy), length of the unit, and other factors (Riley, Higgins, & Deeks, 2011).

The establishment, training and maintenance of TIP groups with AAR was important to the success of the intern program. Initial training was provided through a summer research class dedicated to the intern cohort. In this class, the TIP model was presented and interwoven throughout the course classwork with the data collection. The unique research cohort was given an “in progress” using the fall semester to implement the AAR in their intern classrooms. Grades

were given to the interns upon the successful completion of the AAR project. While the interns were being trained through the research cohort class, mentor teachers were also provided with mentor training that included monthly meetings with the university mentor trainers. Mentors were provided with Mentor Modules produced to support mentors regardless of their use of AAR.

Each TIP group was given latitude in choosing their instructional area of focus, which depended largely on the unique needs of the area of study and placement of the intern. Once the instructional focus was determined, the TIP groups focused on understanding data and their varied uses such as for differentiation of instruction, measuring student engagement, and teacher professional identity. The TIP groups also included discussions around becoming better consumers of research and the use of technology for collecting, cleaning, analyzing, and reporting findings for dissemination purposes. One intern stated “The data I collected helped me to improve my lessons, understand how students learn and change delivery to respond adequately to the needs of my students. I have more insight into what I am doing [in the classroom] because of my AAR experience.” A second intern saw the TIP/AAR experience as helping to establish her professional identity, “I grew enormously. I now view myself as a professional educator. I feel more confident in my ability to shift mid lesson if needed.”

Mentor teachers cited several ways in which their participation in the mentor training and the TIP/AAR experience has positively impacted their self-efficacy when mentoring interns. Mentor teachers cited increased levels of confidence in their reflexive ability when working with interns from different backgrounds, openness to new ideas around models of management and instruction techniques, as well as being better able to give feedback in supportive ways. Using AAR provided a space for mentor teachers to practice these confidence-building techniques. One mentor teacher stated, “I feel that is it important for new teachers to see and use Action Research. New teachers need to be able to try new strategies and then take the data to see what did or did not work. Action Research is a practice.” A second mentor teacher stated, “I see the difference being made in student achievement. It is beneficial to the intern in that it helps them to compare where students are and where they need to be...you can see the growth.”

The formative evaluation feedback was valuable to a majority of the interns as the lesson plans were flexible and could be modified to promote student learning in several ways. One modification, duration of the lesson, was somewhat fixed by the class schedule and the curricular pacing. Approximately 60% of those interviewed commented that more time for the teaching process would help the interns by allowing for remediation of previously taught skills and concepts that were presented to students in previous years, but not mastered by many students. The majority of learning at the middle and high school levels is cumulative, that is simple concepts and skills are needed to solve more complex problems and those simpler concepts are combined and more complex algorithms are formed from these building blocks in learning. Students who are missing critical understanding may be missing a building block in the learning process that requires some effective remediation to reach a more complex understanding of the subject.

Another observation by some of the interns concerns the self-confidence of the learner. Some of the interns indicated an additional need to address the self-confidence of the learner. As described in the reflections and observations of the interns, the students would seem to grasp the concept and successfully apply the problem-solving steps during the lesson, only to fail to perform successfully on a formative assessment of the same concept. This phenomenon may be a reason that some untrained adults attribute this problem to the urban environment. However, to an intern trained in urban teaching at GSU, this was an issue that could be addressed. The interns who cited



low self-confidence chose pedagogical modifications to bolster confidence in the students. For one class, the intern used argumentation, in which the students had to defend their positions about solving problems. In other classes, the interns used graphic organizers to firmly describe the problem-solving steps or creating a project, which ensured the students internalize the concepts within the lesson. These interns demonstrated teaching strategies that showed that every student could learn because, in part, to the yearlong placement and the excellent guidance in the urban partnership districts.

### Conclusions and Significance

The goal of the AAR program is to inform the PDS process of utilizing the TIP groups in preparing beginning teachers who are as effective as or more effective at facilitating student achievement as comparison teachers. The statistically significant .326 effect size produced by the random effects meta-analysis results of these 25 studies closely relates to the effect sizes (.375 and .40) referenced by large meta-syntheses conducted during the past two decades related to educational interventions and student achievement (Sipe & Curlette, 1996; Hattie, 1992). This is noteworthy because the typical intern in the TIP model does not just tie the comparison group teacher in student achievement but outperforms the comparison group teacher in the unit of instruction evaluated. The findings of the PDS partnership using clinical teaching through TIP with AAR compared to other overall effect sizes for educational interventions show an effect size (i.e., .326) which is typical in the published literature, thus, providing evidence for a PDS approach for improving student achievement.

### References

- Bohan, C., & Many, J. (2011). *Clinical teacher education: Reflections from an urban professional development school network*. Charlotte, NC: Information Age Publishing.
- Cooper, H., Hedges, L., & Valentine, J. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation.
- Curlette, W., Hendrick, R., Ogletree, S., & Benson (2014). Student achievement from anchor action research studies in high-needs, urban professional development schools: A meta-analysis. In J. Ferrara, J. Nath, & I. Guadarrama (Eds.), *Creating visions for university-school partnerships*. Charlotte, NC: Information Age Publishing.
- Curlette, W., & Ogletree, A. (2011). An approach to increasing student achievement: teacher-intern-professor groups with anchor action research. In C. Haeussler Bohan & J. Many (Eds.), *Clinical teacher education: Reflections from an urban professional development school network*. Charlotte, NC: Information Age Publishing.
- Darling-Hammond, L., & Richardson, N. (2009). Teacher learning: What matters? *Educational Leadership*, 66, 46-53.
- Hattie, J. (1992). Measuring the effects of schooling. *Australian Journal of Education*, 36, 5-13.
- Hendricks, C. (2009). *Improving schools through action research: A comprehensive guide for educators* (2<sup>nd</sup> ed.). Upper Saddle River, NJ: Pearson.
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 364, doi: 10.1177/1094428106291059.

- Popham, W. J., & Ryan, J. M., (2012, April). *Determining a high stakes test's instruction sensitivity*. Paper presented at the annual meeting of the National Council on Educational Measurement, Vancouver, B.C., Canada. Retrieved from [http://www.minnetonka.k12.mn.us/administration/teachingandlearning/staffdev/2012%20Handouts/June%202012/Popham\\_article\\_DETECTING%20THE%20INSTRUCTIONAL%20SENSITIVITY.pdf](http://www.minnetonka.k12.mn.us/administration/teachingandlearning/staffdev/2012%20Handouts/June%202012/Popham_article_DETECTING%20THE%20INSTRUCTIONAL%20SENSITIVITY.pdf)
- Riley, R. D., Higgins, J. P. T., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *BMJ*, 342, doi: 10.1136/bmj.d549
- Silverman, D. (2010). *Doing qualitative research* (3<sup>rd</sup> ed.). Thousand Oaks, CA: Sage.
- Sipe, T. A., & Curlette, W. L. (1996). A meta-synthesis of factors related to educational achievement: A methodological approach to summarizing and synthesizing meta-analyses. *International Journal of Educational Research*, 25(7) 583-660.
- The Council of Chief State School Officers. (June 2009). *Effects of teacher professional development on gains in student achievement: How meta-analysis provides scientific evidence useful to education leaders*. Washington, DC. Retrieved from <http://www.ccsso.org>
- Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, 24(1), 80-91.
- Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from <http://ies.ed.gov/ncee/edlabs>

*Gwendolyn Benson serves as the Associate Dean for School, Community and International Partnerships in the College of Education and Human Development at Georgia State University.*

*William Curlette serves as the Chair of Educational Policy Studies in the College of Education and Human Development at Georgia State University.*

*Susan Ogletree is the Director for the Center for Evaluation and Research Services in the College of Education and Human Development at Georgia State University.*

*Robert Hendrick works in the Center for Evaluation and Research Services as a Research Associate in the College of Education and Human Development at Georgia State University.*

*Authors' Note: This meta-analysis was made available through the Collaboration and Resources for Encouraging and Supporting Transformations in Education (CREST-Ed, award number U336S140036) and the Network for Enhancing Teacher Quality (NET-Q, award number US336S09137) grants by the U.S. Department of Education. The contents of this article do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.*