

An exploratory examination of practices and behaviors of teachers in
relationship to their teacher effect scores
by Pamala J. Carter

Research on teacher effectiveness has its beginnings in the 1920s with much of the earlier works being framed around the administrator perspective (Dunkin & Biddle, 1974; Gage, 1965). The 1930s and 1940s gave rise to presage-process-product studies where various characteristics of teachers were examined for their relationship with teacher effectiveness on student learning (Campbell, Kyriakides, Muijs, & Robinson, 2003). These studies considered the formative experiences, called presage variables, of the teacher that were present prior to entering the classroom. Variables that represented actions that occurred in the classroom, i.e. activities of the teacher and the students, were called process variables. The product variables were the outcomes that most often related to the progress determined as a result of the process implemented (Dunkin & Biddle, 1974). Experimental studies were popular during the 1940s through the 1960s, including investigations of different teaching styles such as formal and informal and progressive and traditional (Mitzel, 1960).

From the 1960s forward, much of the attention on teacher effectiveness research has concentrated on teacher knowledge and beliefs and their relationships to student progress (Campbell, et al., 2003). Medley and Mitzel (1963) purport that an important factor in determining student learning is the teacher. They further state that factors attributing to this effectiveness, once they have been identified, can be taught in teacher education programs to enhance the practice of new teachers. A more concrete and measurable definition of effectiveness will enable a program of research and analysis that

will lead to a better understanding of what it means to be an effective teacher as well as a better understanding of how to teach those that want to become effective.

The premise of this study assumes that teaching's purpose is to produce student learning, defined by academic progress. A scientific approach linking teacher behaviors to student academic gains will provide evidence of teaching behaviors that promote student learning. Although there is a growing consensus in the role the teacher plays in student learning, to date, there is no agreement on a single definition of teacher effectiveness. Research has progressed to its current use of student standardized test score gains, relative to a particular teacher's class, to determine teacher effectiveness (Sanders & Horn, 1994; Thum, 2003; Webster & Mendro, 1997).

The purpose of this exploratory research is to investigate the effectiveness of various instructional behaviors specifically among math, reading and language arts teachers in grades 3-8, defined by their (a) planning and preparation and (b) instructional practices. Through structured observation and survey assessments, the researcher will examine defined aspects of the teacher to collect data on behaviors associated with each dimension. The Tennessee Value-Added Assessment System (TVAAS) model's teacher effect is the dependent variable defining effectiveness. The researcher will analyze data from the observation and survey, seeking practices that discriminate effective from ineffective teachers.

The importance of finding a model that adequately addresses these complexities, including the universe of classroom behaviors and learning as expressed through student achievement is imperative to future research. No longer are the rating scales and models that were subject to administrator competence and bias acceptable. The question of what

makes a teacher effective has yet to be answered. However, if the motive is the success of the student and that success is to be determined by how well they present their learning outcomes on an annual standardized exam, then a model that considers student-tested learning a function of teacher effectiveness is appropriate.

How best to determine effective teaching practices presents a plethora of shortcomings and complexities, most of which are methodological challenges. According to Millman (1997), psychometric problems with the high-stakes tests as well as the reliability of the measurements used to determine teacher effectiveness should be causes for concern. Other shortcomings include the generalizability of many of the measures, the accuracy of the measures when weighed against the curriculum being taught, weak research designs and weak measures to capture the data (Cruickshank, 1990). These challenges remain but are much more manageable than they were fifty to sixty years ago. Pioneering researchers have grappled with and solved many of the issues that once were major stumbling blocks.

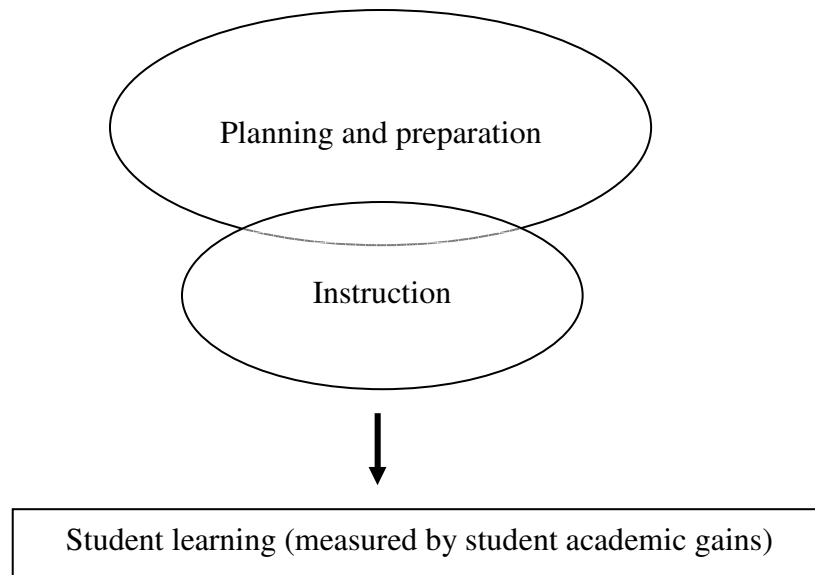
Utilizing value-added models, researchers have begun to discriminate the patterns of teacher effectiveness and have found that some teachers are consistently better at producing higher gains than others (McCaffrey, Lockwood, Koretz & Hamilton, 2003; Thum, 1997; Sanders & Horn, 1998; Webster & Mendro, 1997). These patterns have raised questions about the possibility of being able to improve student learning by studying specific cases. If there are differences among teachers that produce higher gains than their counterparts who do not and these differences can be proven, documented and more importantly, duplicated, the impact on student learning would be substantial.

As researchers continue to examine the relationship between student achievement and teacher practice, many have declared teacher effectiveness to be the single largest determinant of student achievement (Darling-Hammond, 2000; Goldhaber & Brewer, 2001; Hanusek, Kain, & Rivkin, 1998; Mosteller, Light & Sachs, 1996; Sanders & Rivers, 1996; Schmidt, McKnight, & Raizen, 1997; Wenglinsky, 2000; Wright, Horn & Sanders, 1997). Studies (those listed above) that support the claim of the importance of the teacher in determining the academic success of the student capture the multifaceted aspects of undertaking such a complex topic. Methodological challenges abound. The need for further explanation of the appropriate class size, the true variation accounted for by the teacher, heterogeneity among the students, and other issues continue to push researchers to seek out ways to improve current and future research on teacher effectiveness. The urgency of the topic and the imperative need of finding a meaningful method drive this study's search for classroom instructional behaviors that are found to make a difference in teacher effectiveness.

The framework (see Figure 1) will guide the study through the data collection and analysis phases, as well as in the written dissemination of the findings.

Figure 1

The study's model of teacher effectiveness*



**Adapted from Hay McBer, 2000.*

Evidence of the two domains of study was captured through survey and observational data. Planning and preparation, for the purpose of this study, included the perceptions of the teacher towards teaching, instructional preparation and interactions with colleagues and were captured in the teacher survey. Behaviors and practices exhibited in the preparation and planning domain also were collected and documented by trained observers during a scheduled observations.

Observers reported on aspects of planning and preparation as they were present during an observed lesson. Evidence was collected that would support the stated objectives, the lesson's planning and organization and the activities and strategies used to provide evidence of an awareness of diversity and equity issues among students in the classroom.

Through factor analysis, the many variables examined were reduced to a manageable number of interpretable factors for use in subsequent analyses. Regression analysis is designed to determine the relationship between the multiple variables derived from factor analysis and the dependent variable, the teacher effect score. This study examined the relationship between these predictor variables and the teacher effect.

The major research question addressed in this study was: What are the differences in (a) planning and preparation and (b) the implementation of instructional practices of teachers at varying levels of effectiveness? This question will be addressed using math, reading and/or language arts classes in grades three through eight in one Tennessee school district.

Criteria for eligibility for participation in the study included (a) the teacher must teach at least one of the three subjects being investigated (math, reading and/or language arts) and (b) he must have a teacher effect score. Forty-six participants were purposively chosen from the 764 teachers eligible within the Hamilton County school district in Tennessee. The participants included 3rd through 8th grade math, reading, and language arts teachers, representing rural, suburban and urban schools in grades three through eight,

Limitations

A survey was used to collect data on the teachers' perceptions of various aspects of their instruction, beliefs and practices. Teachers are not always aware of their behaviors in the classroom and often articulate that they are doing something that may not be observed in the same way by others. The researcher recognizes this limitation and

will cautiously report survey findings as perceptions, not proven practices and/or behaviors.

The observational protocol presents another limitation in the amount of time allotted to observing instruction. Given the scope and resources of the researcher, only one observation per participant was possible. Given the exploratory nature of this study, the usefulness of the findings in changing policy are limited. However, the findings will provide constructive information for future observational research of classroom behaviors and their relationship to increased student achievement gains.

The study is not generalizable beyond the participant pool. The study's purpose was to begin exploring possible relationships between varying levels of teacher effectiveness and the planning, preparation and instructional practices of teachers examined. The findings will be useful for this purpose.

FINDINGS

Demographics of the participants

Most of the teachers (68%) received degrees from Tennessee institutions, with 47% graduating from the University of Tennessee at Chattanooga located in Hamilton County, Tennessee. Sixty percent hold a Master's degree. The majority of participants (97%) were female.

Teacher effect scores of this group ranged from -9.18 to 7.14, with the mean and median teacher effect scores of 0.42 and 0.49, respectively. The teaching experience for participants in the study ranged from four to thirty-six years. The categories, as defined in table 2, will be used throughout the analysis to describe groups of teachers by varying levels of experience.

Table 2

Number of years taught

	Years experience	Range of teacher effect score	N	Percent	Cumulative percent
Group 1	less than 5 years	-9.18 to 1.75	12	19.35	19.35
Group 2	5-10 years	-0.93 to 4.04	9	14.52	33.87
Group 3	11-19 years	-0.47 to 3.08	10	16.13	50.00
Group 4	20-25 years	-1.38 to 7.14	14	22.58	72.58
Group 5	26-36 years	-7.31 to 3.19	17	27.42	100.00
Total			62	100.00	

Subjects observed included reading, language arts and math. Seventeen observations were made of reading lessons, seventeen of language arts lessons and twenty-eight of math lessons. All but four of the math lessons observed were general math lessons with four teachers observed teaching algebra. The total number of lessons observed was sixty-two. Several teachers were observed teaching more than one subject.

All participants held Tennessee state certifications, with all but one having a standard teaching certificate. Three were not certified in areas in which they were teaching and one participant did not respond to this statement. A majority (75%) of the participating teachers received their certifications as part of their undergraduate studies. Eight percent earned their certifications through Master’s programs, 10% through alternative certification programs after they had begun teaching and 7% earned them through other professional development programs.

Findings from the teacher survey

Categorical and descriptive findings of the teacher survey are reported in the previous section on demographics on participants. The following section will describe

findings from the teacher survey specific to the variables that were used in the factor and regression analyses. Variables of the teacher survey were designed to measure specific aspects of the instructional behaviors exhibited by the classroom teacher. Some were reported descriptively but most were analyzed using factor analysis. Following the factor analysis, regression analyses were performed on the factors and the teacher effect scores. Experience levels are considered in this analysis to determine its impact on the relationship between the study's factors and teacher effect scores. These years of experience are treated as categorical variables and grouped as such: (a) Group 1: less than 5 years experience; (b) Group 2: 5-10 years experience; (c) Group 3: 11-19 years experience; (d) Group 4: 20-25 years experience; and (e) Group 5: more than 25 years experience.

The factor analysis produced 9 major sets of variables or constructs. The variables of these constructs were coded and themed (see table 3). Twenty-seven factors emerged from these nine constructs. Variables included in the construct were chosen if the correlation was 0.60 or higher. Of the 27 factors, not all were found to be statistically significant. Seven main effects across levels of experience and six interactions were found. Twenty-four of the 27 factors were found to be related to years experience defined previously by the five experience groups. These factors will not be reviewed in this section but will be discussed in the final section of this report.

Table 3
 Constructs of the teacher survey

Constructs	Variables included
Student engagement in the work	2-20; 31-50
Teacher actions involved in implementation	21-30
Teacher preparedness for specific strategies	51-57; 79-95 (odd numbers)
Importance of various types of assessment	61-67; 69-77; 78-94 (even numbers)
Teacher attitudes towards teaching	96-105; 181-203
Teacher attitudes towards the school environment	106-119
Interaction with peers/colleagues	168-180
Professional development (site, district, personal)	204-236; 240-263
Attitudes of types of professional development	265-269; 279-297

Seven factors were found to have main effects where the factor was a strong and statistically significant predictor of the teacher effect scores across levels of experience. Six additional factors were found to have interactions between the teacher effect scores, the years experience groupings and the factor queried. For factors to have statistically significant interactions, the *p* value of the type II analysis had to be less than .05.

Of the six factors found to have interactions, two were also found to have main effects and will be discussed in the following section. Most of the interactions found clustered around beginning teachers and those with the most experience. There was very little, if any, interaction occurring between the factors, the teacher effect scores and years of experience for teachers representing Groups 2 and 3 (5-19 years of experience). Of interest in the interactions, most times the more experienced teachers' and beginning teachers' responses were contrary to each other. The exception to this was found in one factor where both beginning and more experienced teachers were found to have teacher effect scores positively related to higher agreement on school-based professional development variables.

Findings from the teacher survey report that for all teachers examined, there is a negative relationship with the teacher effect scores in classrooms where teachers report (a) students engaged in verbal interactions; (b) teachers encouraged high level interactions; (c) [teachers] involved in collaborative problem-solving with peers/colleagues and (d) district provided professional development opportunities. There is a positive relationship among all teachers reporting on the teacher survey in the following areas: (a) [teacher] providing clear expectations; (b) [teacher] support for instruction; and (c) school-based professional development opportunities.

Other results are found to be significant among groups of teachers defined by their levels of experience. There are several interactions with the factor and the teacher effect score, dependent on years of experience. Most of the significant findings relate to the least experienced group of teachers.

Negative interactions for Group 1 (less than 5 years) include the factors “students are engaged with materials” and “teacher attitudes towards the school environment.” In classrooms where teachers report higher agreement or engagement in these areas, the teacher effect scores are lower. For “students are engaged with materials,” teachers with less than 10 years experiences (including Groups 1 and 2) are found to be negatively related to the teacher effect score, meaning teacher effect scores are lower in classrooms with higher levels of student engagement.

Positive interactions for Group 1 include the factors “peer observations” and “professional development characteristics.” Teachers with less than 5 years experience that reported the importance of and/or higher levels of agreement with these factors are found to have higher teacher effect scores.

A few conflicting results were found among the most experienced and the least experienced teachers. Evidence of student engagement with materials is found to be negatively related to the teacher effect score for teachers with the least experience. For teachers with the most experience, the opposite is true. Teachers with the most experience that report high levels of student engagement with materials in their classrooms are found to have higher teacher effect scores. The same contradiction is seen with attitudes towards “professional development characteristics” seen as important to the teacher. Those with the least experience are positively related to this factor while teachers with the most experience are found to be negatively related with the teacher effect scores. A third significant result for the most experienced teachers is with the factor, “implementation of performance-based instruction.” The most experienced teachers that report higher frequency of such instruction are found to be less effective, i.e. have lower teacher effect scores.

Findings from the teacher survey confirm the complexities involved in the individual perspectives of teachers as they relate to the measurement of teacher effectiveness. Perspectives of instructional behaviors of effective beginning teachers were often found to be in direct contradiction to those same behaviors for more experienced teachers. A particular instructional behavior might be found to relate positively to student learning for one group while negatively related to the other. This discovery however does not suggest that the findings are not significant and crucial to uncovering aspects of teaching that lead to increased student academic gain. It supports the need for closer examination and better measurements to tease out any differences that might exist among experience levels of teachers. This finding also supports the need for varying the

investigation to include observational methods for collecting information on the instructional behaviors.

Findings from the observation process

The observational system used to record data collected in the classroom includes a sequential set of instruments. First, the observer, prior to observing the lesson, sketched the classroom, collected information on various aspects of the classroom's physical structure. Also, prior to the observation, an environmental checklist was completed, reporting the condition of various materials in the classroom. A narrative was completed during the lesson, collecting specific points determined prior to the observation. Following the observation, the behavior indicator checklist was completed. Observers used the behavior indicator checklist as a guide to specific comments documented in the narrative. After having completed all of these instruments, the observer then completed the summative observation assessment (SOA) form. This form served to combine the evidence collected on all instruments completed prior to, during and immediately following the observation.

Summary. A limited number of instructional behavior statements reviewed were found to produce statistically significant results when analyzed. The interactions presented in the preceding section all include interactions with the behavior, the teacher effect and Group 5, the most experienced teachers. Two interactions were found to be negatively related to the teacher effect score. For teachers in Group 5, those observed using varied assessment and classrooms where there was evidence of implementation of performance-based instruction had lower teacher effect scores than teachers where these behaviors were not observed. The third interaction involving Group 5 was a positive

relationship with the teacher effect score and classrooms where the students were engaged with materials. Interestingly, for that same behavior (students engaged with materials), responses from teachers in Groups 1 and 2 were found to be negatively related with their teacher effect scores. Teachers from Groups 1 and 2 where it was observed that students were engaged with materials had lower teacher effect scores. The last interaction noted involved teachers in Group 4 and the implementation of performance-based instruction. Although teachers in Group 5 were found to have a negative relationship between this behavior and the teacher effect scores, teachers in Group 4 were found to have a positive relationship.

The dearth of statistically significant findings among the behavioral statements observed should be carefully noted and consideration given to why so many behaviors did not yield more significant results.

IMPLICATIONS, RECOMMENDATIONS AND CONCLUDING THOUGHTS

In the examination of the relationship between measures of teachers' instructional behaviors and measures of their students' academic progress, the findings can be summed up in three major points. First, in reviewing a teacher's instructional behaviors and the relationship of those behaviors to student academic progress, some instructional behaviors were found to be significant predictors of the teacher effect scores among all teachers examined in this study. Second, in reviewing these instructional behaviors and the relationship of those practices to student academic progress, some instructional behaviors were found to be significant predictors of teacher effect scores, but only when experience was held constant. Lastly, most instructional behaviors examined in this study were found to be limited and lacking in their ability to predict the teacher effect scores of

participants reviewed. The implications of these findings, the recommendations for consideration and concluding thoughts are presented in this final chapter.

Implications

Just as important as the findings are the implications of these findings to what we determine to be “best practices” noted in classroom instruction. Implications of these findings include: (1) maybe there is not AN effective instructional behavior; (2) possibly a sequence of actions is more representative of effective behavior; (3) perhaps clusters of behaviors, instead of individual behaviors, make the difference; (4) perhaps effective behaviors are immeasurable; and/or (5) the approach in the strategy used to capture these findings was inadequate.

It is not the intent of this research to suggest that there is a single behavior that *must* be present among effective teachers. There may not be a behavior that is common among all but maybe there *is* an effective practice. To further explain, it may be that behaviors are nested under some larger umbrella of macro-behaviors and/or actions. The overarching behavior, combined with other specific actions may in turn create an effective practice. The overarching behavior may actually be based in a personality trait. Examining individual behaviors or traits in isolation may not be an appropriate way to determine these differences.

Individual effective behaviors are often elusive and sometimes missed in an observation. Consideration might be given to the practices viewed in a sequence instead of independently. Do effective practices follow some unknown sequential order? Is there an order in which these practices should be learned and/or exhibited in the classroom? Would it be appropriate to teach preservice teachers how to differentiate for various

reading levels prior to teaching them how to use various reading strategies independently? One would presume that preservice teachers would need to know something about the various teaching strategies *before* they would be able to execute them well, especially when differentiating the strategies according to the needs of the students they serve. Would it be appropriate for that teacher to implement these various differentiated reading strategies to students prior to assessing the class to see which strategy would be best for each? Again, the order in which these behaviors are not only taught but the order in which they are implemented in the classroom may contribute to the effectiveness. Maybe the sequence spoken to here is being overlooked. Maybe there *is* an effective practice, but *not* an effective behavior.

The idea of behaviors presented in a particular sequence is interesting. There could be a set of behaviors that when implemented simultaneously may be very effective, but when implemented individually are not. Could it be that if a few of the behaviors in a set are removed and others added, the set of behaviors once deemed effective no longer works? Maybe there is the “cluster” of behaviors that create an effective practice relative to the type of lesson being taught. If this notion were to be considered, how then do we determine what this cluster of behaviors looks like? It could be that over time clusters of behaviors might be captured by examining numerous lessons. This may lead to other questions. Would these clusters then distinguish themselves to particular types of lessons? Would the cluster of behaviors for an introduction to a new idea look the same as would that of a review lesson? Do these questions imply that behaviors, if examined within the context of the lesson and activity being observed, naturally cluster in groups?

It has often been stated that good teachers are born, not taught or trained. Another implication of this study's findings might be that behaviors that are most important are unobservable and perhaps not measurable. Could it be that the difference is in the innate abilities of the effective teacher to know when, how and where to implement effective behaviors? Could ineffective teachers be going through the same motions mechanically but not linking them in productive ways? This might explain why it was difficult in this study to discriminate distinguishable differences. The often unobservable aspects of teaching, the innate abilities of the individual teacher, may play a major role in determining practices that are most successful at promoting academic growth. Research procedures to date are unable to tap these effectively.

In an effort to systematically capture predetermined behaviors in a scientific and formulated manner, the validity of the instrument could have been jeopardized. Maybe there were effective practices occurring in the classrooms examined, but the instrument used to capture those practices was insensitive to them. Since student learning is the underlying definition of the effectiveness of the teacher in this study, it could be that the student should be the focus of the research. Another alternative might be that we are examining the wrong behaviors altogether. In this study many behaviors that had previously been suggested as best practices that were found to be limited or lacking in their ability to predict differences in teacher effect scores. This may speak to the need to reexamine all aspects of the design to determine if a different plan and system might lead to different results.

Findings from the study suggest there are several areas that could be further examined and show promise as being predictors of teacher effect scores. It may be that a

particular instructional behavior may not fit all teachers, but may be better suited for teachers within a specific level of experience or perspective. Also, it might be that the important differences in these teachers may not be related to their behaviors, but to something else. Implications from the teacher survey and observation instrument lead to several recommendations.

Recommendations

Given the divergent implications presented in the previous section, recommendations for research can be summed up in three categories. First, consider other factors that might be congruent with best practices. Second, consider the behaviors within the context of the lesson and class involved. Third, investigate the impact of experience on the amount of student learning.

A study examining best practices of the teacher, in isolation of other aspects of the classroom, might be limited in its ability to accurately provide findings that predict positive student learning outcomes. We might find that best practices are a place to start, but from the findings of this study best practices were seen in classrooms of both effective and ineffective teachers. Future research should place more emphasis on the person(s) being examined. A design that incorporates study of the teacher as an individual, his/her personality as it relates to how (s)he teaches may be helpful in teasing out differences between those that are effective and those that are ineffective.

It also might be beneficial to consider students in the design frame. If the outcome is the academic growth of the student, aspects of how the student learns, interactions that are found when the learning occurs and relationships that foster such learning should not be minimized. There is a dearth of previous research on teacher-student interactions. This

research should be re-examined for its efficacy and possible use in a modern study of interactions, using the teacher effect score as the predictor variable.

A more contextualized observation system would consider the teacher and the student in various aspects of the classroom: teaching; learning, relationships and context. It is possible that personality traits and other characteristics which teachers bring to the classroom are being overlooked, or at least minimized, and should be carefully studied before discounted.

Second, simply capturing that a behavior occurs, apart from the context of that lesson, is not sufficient. The appropriateness of the behavior exhibited also should be more closely examined. Findings of this study suggest that some behaviors were independently strong predictors of teacher effect scores while others were dependent on experience and others showed no relationship at all to teacher effect scores. This implication requires the researcher to step back and examine the process of data collection.

As suggested by Barr (1929), the appropriateness of the context seems to play an important role when considering factors that relate to teacher effectiveness. For example, in Barr's study, teachers were asked to present a lesson. An observer documented practices and behaviors exhibited in this classroom. However, the lesson's content was not prescribed or even suggested for that matter. Some teachers responded that the focus of the lesson was review, others as an introductory lesson, while others reported lessons as clarifications or further discussions on previously discussed topics. These differences in delivery could be compared to Barr's consideration of "appropriateness within the context" (Barr, 1929). Teachers would likely not employ the same practices and

strategies in an introductory lesson that they might once the material has been introduced and students are given an opportunity to practice what has been taught. A limitation of this study (no defined lesson) would strongly suggest that lessons be observed within a specific context (i.e., introduction, practicing concepts, reviewing ideas, etc.). A more carefully designed study that considers the appropriateness of the behaviors within the context of the lesson observed might better capture distinguishable differences in instructional practices.

Third, for 25 of the 27 factors of the teacher survey there was a statistically significant difference among the five groups of experience analyzed. The two factors in which experience was not significant include “students engaged in verbal interactions in the classroom” which had a statistically significant negative main effect and attitudes of “teachers helping teachers” which was not found to have any significance. Throughout the findings of this study, experience played a significant role. To consider the instructional behaviors absent the experience of the teacher is clearly misleading. Any further study of teacher effectiveness should consider this seriously.

Incorporating suggestions from recommendations one through three, researchers should consider creating a new instrument that would be better at capturing differences in teachers that are most effective and those that are ineffective in terms of student academic gains measured by valued-added scores. If these behaviors are indeed measurable, the need for a more scientifically designed observational instrument is crucial to accurately capture variables that might be predictors of teacher effectiveness. The current process has proven ineffective. A worst case scenario would be one in which researchers become content with the findings of date and push for policies to change the way we prepare

teachers based on a flawed design. This study suggests that there is little basis on which to push for any specific behaviors.

Much of the literature on teacher effectiveness comes from the extensive body of literature on teacher evaluation (Beecher, 1949; Charters & Waples, 1929; Danielson, 1996; McBer, 2000; Merriam, 1905). These evaluations may be the problem in the design of many studies on teacher effectiveness. Most teacher evaluations are the result of observation. This observational information then is crafted into formal rating systems that are used in research studies to examine teacher effectiveness. These rating and ranking systems are met with apprehension when used as instruments to define teacher effectiveness. As evaluations, they are met with criticisms of subjectivity and bias. As observation instruments, understandably, they are confounded by what was previously known as best practices. With the academic gains (i.e., student learning) as the outcome measure for the predictor, effective teaching, it is imperative that a valid instrument be used to capture student learning. Careful attention should be given to what is happening in the classroom. It is not sufficient to use a checklist to capture a group of best teacher behaviors that could or could not be predictors of academic success for the students that are recipients of what we have determined to be “best practices.”

Particularly in the age of accountability, with pay for performance knocking on the doors of many states, the use of a proven, robust observational instrument, or combination of companion instruments is imperative not only to validate the assessments adopted by the policymakers, but to provide fairness and equity to the educators being evaluated by this instrument. It is time to reconsider our observational assessment

processes and work to craft an instrument that is more successful in capturing interactions and classroom delivery of instruction which lead to greater student learning.

Concluding thoughts

Continued efforts are needed to address the deficits of observational instruments used in determining the effective practices and behaviors in the classroom. It is apparent that findings of this study are held within the confines of the observation variables reviewed. Is it possible that other variables might be better at predicting teacher effectiveness? Have these variables been determined, researched and/or considered in such research? It would be careless to report that there are no predictors of teacher effectiveness except those found in this study.

It is the belief of the researcher that an adequate observational instrument has not been found. Much work has been done and we have come a long way. Yet more work is necessary before we put children in jeopardy of being recipients of poorly constructed instructional strategies that are not founded on a series of sound and rigorous scientific inquiries. Without a clear connection between teaching and learning, recommendations have little grounding. It is imperative that educational researchers continue to search and question tools being used to determine instructional practices that are most effective at promoting student learning.

Findings from this study provide useful information in exploring the complexities involved in observing effective teaching that promotes student learning. There are some factors that, in this study, have proven to make a difference in teacher effectiveness. There are other factors that have proven to make a difference when experience has been considered. There are, no doubt, other factors yet to be explored that may contribute as

much or more to the equation of “best” instructional behaviors proven to increase student progress.

Teacher preparation programs are struggling with some of the same issues in trying to determine which practices should be taught in their colleges to provide the best training for future teachers. For the legitimacy of the profession and the strength of our future, we must take this matter seriously and rigorously determine what it is about teaching that makes a difference in achievement.

More than thirty years after Barr’s study (1929), Wallen and Travers (1963) wrote of the impact of various teaching methods. They purported that when considering teacher effectiveness, various methods made little difference and one method could not be favored over another. A statistically significant difference in the means between those observed using the behavior and those not using the behavior was found in only three of the twenty-eight instructional behaviors reviewed in the observational component of this study. This study largely does this. It is time for a full reconceptualization of teaching and how it is connected to learning.

REFERENCES

- Barr, A. S. (1929). *Characteristic differences in the teaching performance of good and poor teachers of the social studies*. Bloomington, Illinois: Public School Publishing Company.
- Campbell, R. J., Kyriakides, L., Muijs, R. D. & Robinson, W. (2003). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education*, 29(3), 347-362.
- Charters, W. W. & Waples, D. (1929). *The Commonwealth teacher-training study*. Chicago, Illinois: The University of Chicago Press.
- Cruickshank, D. R. (1990). *Research that informs teachers and teacher educators*. Bloomington, IN: Phi Delta Kappa.
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1), 1-40.
- Dunkin, M. & Biddle, B. (1974). *The study of teaching*. New York: Holt, Rinehart & Winston.
- Gage, N. L. (1965). Desirable behaviors of teachers. *Urban Ed*, 1, 85-95.
- Goldhaber, D. D. & Brewer, D. J. (2001). Evaluating the evidence of teacher certification: A rejoinder. *Educational Evaluation and Policy Analysis*, 23(1), 79-86.
- Hanushek, E. A., Kain, J. F. & Rivkin, S. G. (1998). *Teachers, schools, and academic achievement* (pp.1-50). Cambridge, MA: National Bureau of Economic Research.
- McBer, H. (2000). A model of teacher effectiveness. Report to the Department for Education and Employment. Retrieved from <http://www.teachernet.gov.uk/teachinginengland/detail.cfm?id=521>
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M. & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND.
- Medley, D. M. & Mitzel, H. E. (1963). Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), *Handbook of Research on Teaching* (pp. 247-328). Chicago: RAND McNally.

- Merriam, J. L. (1905). Normal school education and teaching efficiency. In D. E. Beecher, *The evaluation of teaching, backgrounds and concepts* (p. 5). Syracuse NY: Syracuse University Press.
- Millman, J. (1997). *Grading teachers, grading schools*. Thousand Oaks, CA: Corwin Press.
- Mitzel, H. E. (1960). Teacher effectiveness. In C. W. Harris (Ed.), *Encyclopedia of Educational Research* (3rd edition). New York: Macmillan.
- Mosteller, F., Light, R. J. & Sachs, J. A. (1996). Sustained inquiry in education: Lessons in skill grouping and class size. *Harvard Education Review*, 66, 797-842.
- Sanders, W. L. & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 12(3), 247-256.
- Sanders, W. L. & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville, TN: University of Tennessee at Knoxville.
- Sanders, W. L. & Horn, S. P. (1998). Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, 12(3), 247-255.
- Schmidt, W. H., McKnight, C. C. & Raizen, S. A. (1997). *A splintered vision: An investigation of U. S. science and mathematics education*. Boston: Kluwer Academic Press.
- Thum, Y. M. (2003). Measuring progress toward a goal: Estimating teacher productivity using a multivariate multilevel model for value-added analysis. *Sociological Methods and Research*, 32(2), 153-207.
- Wallen, N. E. & Travers, R. M. W. (1963). Analysis and investigation of teaching methods. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 448-505). Chicago: Rand McNally.
- Webster, W. J. & Mendro, R. L. (1997). The Dallas value-added accountability system. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 81-99). Thousand Oaks, CA: Corwin Press.
- Wenglinsky, H. (2000). *How teaching matters: Bringing the classroom back into the discussions of teacher quality*. A policy information report. Princeton, N.J.: The Milken Family Foundation and Educational Testing Service.

Wright, S. P., Horn, S. P. & Sanders W. L. (1997). Teacher classroom and context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11, pp. 57-67.