J Pers Eval Educ (2007) 20:117–128
DOI 10.1007/s11092-008-9058-7

Using Action Research and Local Models of Instruction to Enhance Teaching

Robert J. Marzano

Received: 11 February 2008 / Accepted: 12 February 2008 /
Published online: 11 March 2008

Keywords Action research · Instructional models · Teacher autonomy · Master teachers · Teacher observation · Instructional strategies · Teacher feedback · Teacher evaluation

Education has long pursued a knowledge base to guide instructional practices. In their 1993 article in *Review of Educational Research* entitled "Toward a Knowledge Base for School Learning," Wang et al. (1993) combined the results of a content analysis of 86 chapters from annual research reviews, 44 handbook chapters, 20 government and commissioned reports, and 11 journal articles to develop a list of specific strategies that have research and theoretical support for their effectiveness. That same year Lipsey and Wilson (1993) analyzed the findings from 302 meta-analyses and estimated the overall effect of psychological and educational interventions to be.34 in standardized mean difference terms (i.e. 34 of a standard deviation). Hattie (1992) analyzed the results from 134 meta-analyses of specific instructional techniques and computed an overall effect size of.40 (i.e..4 of a standard deviation) for such interventions. Most recently Seidel and Shavelson (2007) calculated an average effect size of.45 (i.e. 45 of a standard deviation) for domain specific instructional strategies on cognitive outcomes.[1]

While these effects might not be considered large by some, one can make a case that they are substantial when compared to effect sizes considered large in other fields. Rosenthal and Rubin (1982) use the Binomial Effect Size Display (BESD), to demonstrate this very point. They explain that when Smith and Glass (1977) summarized the results of some 400 studies on the effects of psychotherapy, they computed an overall correlation of.32. A Pearson correlation is comparable to a standardized mean difference effect size in that the former is generally about one-half the size of a standardized mean difference. In this case, the correlation of.32 translates to a standardized mean difference of.68 (i.e..68 of a standard deviation)[2].

[1] Seidel and Shavelson (2007) reported their results in terms of a Fisher Z transformed correlation. This was transformed to a Pearson correlation and then transformed to a standardized mean difference effect size.

[2] The typical formula used to transform a Pearson correlation to a standardized mean difference effect size is 2r/(1-r^2)^.5.

R. J. Marzano (✉)
7127 S. Danube Court, Centennial, CO 80016, USA
e-mail: Robertjmarzano@aol.com

An eminent critic of psychotherapy (Rimland 1979) immediately argued that the results of their study sounded the "death knell" for psychotherapy because a correlation of.32 is considered by many to be a moderate effect size and by some a small effect size. After all, a correlation of.32 means that psychotherapy accounts for only about 10% of the variance in patient outcomes. However when this effect size is represented using the BESD a different interpretation is warranted. This is depicted in Fig. 1.

The BESD depicts a correlation in terms of a hypothetical population that is dichotomized on the dependent variable—in this case mentally healthy versus mentally unhealthy. As Fig. 1 demonstrates given a correlation of.32 between psychotherapy and mental health one would expect 66% of subjects (in a population where 50% are mentally healthy and 50% are mentally unhealthy) who had received psychotherapy to be mentally healthy and 34% of subjects who had not received psychotherapy to be mentally unhealthy. Even though there is a 50/50 split in the general population regarding mental health, when that population is reexamined from the perspective of psychotherapy versus no psychotherapy important differences are made evident. Clearly, from this vantage point it seems that psychotherapy has a profound effect on patient mental health.

To further dramatize the importance of seemingly small effect sizes, Rosnow and Rosenthal (1989) comment on the report of the Steering Committee of the Physicians' Health Study Research Group (1988). They note that on December 18, 1987, researchers decided to end prematurely a randomized double blind experiment on the effect of aspirin on reducing heart attack on the grounds that the findings were so important that it would be unethical to continue giving half the subjects a placebo. The effect size for aspirin on the incidence of heart attack was r=.034. Stated differently, the coefficient of determination for incidence of heart attack regressed on use of aspirin was.0011 which means that aspirin accounts for dramatically less than one percent of the variance in the dependent measure. Figure 2 represents these findings in BESD terms:

Figure 2 indicates that there is an expected 3% difference in heart attack when subjects who take aspirin are compared with those who do not take aspirin within a hypothetical distribution where the incidence of heart attack is 50%. Rosnow and Rosenthal emphasize that this seemingly small difference is considered important when the dependent measure is something as consequential as incidence of heart attack. In comparison, consider the research findings for instructional strategies presented at the beginning of this article. Lipsey and Wilson (1993) standardized mean difference effect size of.34 translates to a correlation of.17; Hattie (1992) standardized mean difference effect size of.40 translates to a correlation of.20; Seidel and Shavelson (2007) standardized mean difference effect size of.45 translates to a

	Mentally Healthy	Mentally Unhealthy
Psychotherapy	66	34
No Psychotherapy	34	66

Fig. 1 BESD for psychotherapy meta-analysis. Computed from data reported in Smith and Glass (1977)

	No Heart Attack	Heart Attack
Aspirin	48.5%	51.5%
No Aspirin	51.5%	48.5%

Fig. 2 BESD for aspirin study. Computed from data reported in Rosnow and Rosenthal (1989)

correlation of.22.[3] All of these correlations are much larger than the.034 correlation between taking aspirin and incidents of heart attack. From this perspective one would conclude that educational research most certainly provides information about instructional strategies that should be taken seriously by practitioners.

While it is certainly true that educational research provides us with guidance as to the nature of effective teaching, I strongly believe that there is not (nor will there ever be) a formula for effective teaching. This is not an unusual claim. Many researchers and those who try to apply research (a category into which I place myself) would most probably agree. Commenting on educational research in the 1970s and 1980s, Willms (1992) notes "I doubt whether another two decades of research will…help us specify a *model for all seasons* [original emphasis]—a model that would apply to all schools in all communities at all times" (p.65). A similar sentiment is credited to the famous mathematical statistician George Box who is reported to have said that all mathematical models are false but some are useful (DeLeeuw, 2004). In effect, Box warned that mathematical models which form the basis of all quantitative research are approximations only of reality, yet they can help us understand the underlying dynamics of a specific situation. Reynolds and colleagues (2000) address the issue in the following way:

> Sometimes the adoption of ideas from research has been somewhat uncritical; for example, the numerous attempts to apply findings from one specific context to another entirely different context when research has increasingly demonstrated significant contextual differences. (p. 216)

While the comments of Willms and Reynolds and colleagues address the broader issue of school reform they are quite applicable to research on classroom instruction. No amount of further research will provide an airtight model of instruction.

Riehl (2006) offers an interesting perspective as a result of her contrast of educational research with medical research. She notes that medical research employs a variety of methodologies that range from randomized clinical trials to single subject case studies. But the findings from these studies are anything but absolute. She explains: "Even the seemingly most determinant causal association in medicine (such as the relationship between smoking and lung cancer) is really just a probability" (p. 26). Riehl comments:

> When reported in the popular media, medical research often appears as a blunt instrument, able to obliterate skeptics or opponents by the force of its evidence

[3] The typical formula used to transform a standardized mean difference (*d*) into a Pearson correlation is d/(d^2+4)^.5.

> and arguments…Yet repeated visits to the medical journals themselves can leave a much different impression. The serious medical journals convey the sense that medical research is an ongoing conversation and quest, punctuated occasionally by important findings that can and should alter practice, but more often characterized by continuing investigations. These investigations, taken cumulatively, can inform the work of practitioners who are building their own local knowledge bases on medical care. (pp. 27–28)

The individual medical practitioner must sift through a myriad of studies and opinions to build his local knowledge base with which to interact with patients. So too must the practitioner in education. Educational research is not a blunt instrument that shatters all doubt about best practice. Rather it provides general direction that must be interpreted by individual districts, schools, and teachers in terms of their unique circumstances. Practitioners must determine which strategies to employ with the right students at the right time. Consequently, it is up to local school and districts to develop a knowledge base of effective teaching strategies. Fortunately, there is a powerful tool that can be used in this endeavor—action research.

Using Action Research

Over the years a number of models of effective pedagogy have been proposed (e.g. Hunter, 1984). While a case can be made that a district or school should simply adopt a model, a case can also be made that "off-the-shelf" interventions are typically short lived in K-12 education. This point was made by Cuban (1987) who chronicled the fate of a number of interventions all of which were basically sound and had supporting research. Some of the more visible ones that have not endured are programmed instruction, open education, the Platoon System, and flexible scheduling. A viable alternative to adopting an instructional model wholesale is to use action research to develop a local school or district approach.

The concept of action research has become quite popular in the last few decades. Nolen and Putten (2007) note that action research was first introduced as a methodology in education research in the mid-1950s. They explain that it "surfaced in response to the growing need for more relevant and practical knowledge in the social sciences: It bridged the gap between academic research and day-to-day applications" (p.401). Chiseri-Strater and Sunstein (2006) describe action research conducted by teachers in terms of praxis. They note that praxis involves:

> …connecting our ideas with our actions, deriving theories from our practices… [It] gives us the power to understand teaching as a kind of scholarship and resists ideas that confuse our common sense. We reclaim internal agency for ourselves as inquisitive, successful professionals when we take the time to ask what works and then try to answer it." (p. xxii)

The comments of Chiseri-Strater and Sunstein provide a unique perspective on research. Indeed, they note that their book: "expands the definition of teacher research by stretching the potential genres that people traditionally accept as 'research'" (p. xxii).

Action research for the purposes discussed here begins with the identification of specific instructional techniques that are to be studied. This typically means selecting strategies from existing list of effective practices. For example, relative to instructional strategies, Marzano, Pickering, and Pollock (2001) identified the following nine instructional strategies:

1. identifying similarities and differences
2. summarizing and note taking
3. reinforcing effort and providing recognition
4. homework and practice
5. nonlinguistic representations
6. cooperative learning
7. setting objectives and providing feedback
8. generating and testing hypotheses
9. cues, questions, and advance organizers

Relative to classroom management strategies, Marzano, Pickering, and Marzano (2003) identified the following four areas:

1. rules and procedures
2. disciplinary interventions
3. teacher–student relationships
4. teacher mental set

Other similar lists of effective strategies have been developed by Good and Brophy (2003) and Mayer (2003).

Once a reference list of strategies has been identified teachers throughout a school or district can conduct action research projects on a voluntary basis. Action research projects can be quite informal or formal. At an informal level teachers might simply try strategies in their classrooms and record their impressions of how well they worked. At a more formal and more rigorous level teachers can design and carry out studies involving experimental classes (classes in which a specific strategy is employed) and control classes (classes in which the selected strategy is not employed). Over the last five years I have been involved in over 100 action research projects that employed experimental/control classes and controlled for previous knowledge using pretests as covariates (see, for example, Marzano & Associates 2005). Results from these studies are reported in Fig. 3.

In Fig. 3 effect sizes are reported in terms of standardized mean differences as opposed to correlations. The average effect size in Fig. 3 is.39 and the median is.28. One interesting aspect of these results is that they were produced with very little if any professional development. Specifically, the vast majority of the 113 teachers either participated in a one-day or half-day in-service professional development workshop regarding specific instructional strategies, read a brief description of a specific instructional strategy, or both. Even a moderate average effect of.39 is considerable under these conditions. Specifically, Prentice and Miller (1992) make the case that effect sizes must be judged as important not simply on their magnitude but also on the resources that are required to produce the effect. In this case the required resources were minimal (i.e. attend a professional development day, read a brief description).

Mean	.39
Median	.28
Range	6.67
10^{th} percentile	-.43
20^{th} percentile	-.17
25^{th} percentile	-.06
30^{th} percentile	.06
40^{th} percentile	.15
50^{th} percentile	.28
60^{th} percentile	.33
70^{th} percentile	.61
75^{th} percentile	.70
80^{th} percentile	.90
90^{th} percentile	1.44

Fig. 3 Distribution of 113 effect sizes (standardized mean difference)

Design a Local Model or Language of Instruction and have Teachers Interact Using that Model

Once instructional strategies have been studied via action research, a school or district is in a position to design a model or "language of instruction. An instructional model should not be misconstrued as an attempt to constrain teachers to one particular approach to teaching. Rather, it should be interpreted as a necessary vehicle for communication between teachers. In effect, the model should constitute an agreed upon way of describing and discussing effective teaching. This idea has been espoused by many. For example, the importance of a common language is addressed implicitly and explicitly by those who promote the importance of Professional Learning Communities (PLCs) (Stoll et al. 2006).

Typically schools and districts begin with a very limited instructional model in that they identify a few strategies only that have been validated by their action research. For example a school or district might identify the following four strategies in their initial model:

- Strategy 1: Designing clear learning goals and providing feedback
- Strategy 2: Previewing new content

- Strategy 3: Using notes and graphic organizers to explain content
- Strategy 4: Reviewing content using comparison, contrast, classifying, analogies and metaphors

Over time as new strategies are studied using action research, the model is expanded to include a variety of strategies for instruction and classroom management.

A model of instruction is powerful only if used as a vehicle for communication—as the basis for conversations about effective teaching. Louis et al. (1995) refer to such conversations as reflective dialogue. Little (2002) notes that naturally occurring interactions among teachers regarding instruction are a useful tool in terms of informally enhancing instruction. While naturally occurring interactions should be supported, it is also important to provide a structure for these interactions that does not turn into what Hargreaves (1994) refers to "contrived congeniality." Dimmock (2000) notes that providing teachers with the time and space to interact about instruction are critical to effective interaction; however, time and space are not sufficient. A format and structure for such interactions should be developed.

To illustrate, one district with whom I have worked uses "late starts" on a monthly basis. During late starts teachers meet in small grade level or subject matter teams to discuss instructional issues. Between late start meetings teachers are asked to record their reactions to instructional techniques they have tried from the district model. The record keeping is kept to a minimum for teachers. Teachers simply spend a few minutes after a particular lesson in which they tried a strategy recording their perceptions of the effectiveness of the strategy. During their late start meetings teachers discuss their recorded observations using the following protocol:

- describe the strategy or strategies you tried
- describe its effect on student learning and the evidence for your conclusions
- describe what you did
- describe areas for improvement on your part

Have Teachers Observe Master Teachers Applying Instructional Strategies

Teachers systematically talking about instruction will go a long way to creating a culture of effective teaching. However, nothing will put effective pedagogy in the spotlight as well as teachers observing teachers. Louis et al. (1995) note that ultimately PLCs must foster the "deprivatization of practice." This is perhaps one the most difficult aspects of PLCs to implement. In his book *A Place Called School* which summarized data from 1,350 elementary and secondary teachers, Goodlad (1984) noted that teachers generally report that they would like to observe others: "…approximately three quarters of our sample at all levels of schooling indicted that they would like to observe other teachers at work" (p. 188). This same issue is addressed by Flinders (1988) in his article "Teacher Isolation and the New Reform." He notes that teacher isolation is an adaptive strategy teachers engage in because of the cellular organization of schools which are unwittingly designed

to minimize interaction. Unfortunately, the by-product of this isolation is lack of shared practical knowledge. Similar sentiments have been expressed by Shulman (2004).

Here I recommend a very specific approach to teachers observing teachers. It begins with the identification of "master teachers." It is important to note that a master teacher is defined as one who produces substantial gains in student learning. This is in contrast to defining a master teacher as one who employs specific instructional strategies. Although this might seem counterintuitive, it has a strong logic. Given the complexity of the teaching/learning process, it is safe to say that no model of instruction or7 set of instructional strategies could completely define effective teaching. This sentiment has been expressed directly or indirectly by many researchers and theorists (Willms, 1992; Reynolds et al. 2000; Berliner 1986; Marzano 2007a, 2007b). Different teachers employing the same instructional techniques might produce very different results in student learning. Consequently, overall effectiveness in teaching must be defined in terms of the one indisputable criterion—student learning.

Once master teachers have been identified using the criterion of student learning, each master teacher's strengths can be identified in terms of the model of instruction that has been designed by the school or district. For example, one master teacher might be quite skilled at designing clear learning goals and providing feedback. Another master teacher might be recognized as particularly skilled at previewing new content and so on. Presumably each master teacher would be skilled in a number of components of the instructional model, although master teachers would probably different skill profiles in terms of their strengths. On a voluntary basis, teachers would then sign up to observe master teachers for specific design questions. For example, if a teacher wanted to observe an expert in designing clear learning goals and providing feedback she would seek out one of the district experts on this issue. Ideally, the expert teacher would also visit the classroom of the teacher seeking assistance.

Provide Feedback to All Teachers on the Effectiveness of Their Instruction

To fully implement a local model of instruction every teacher must be provided with feedback regarding the effectiveness of their instruction with the intent of capitalizing on strengths and improving on weaknesses. One important reminder is useful here. As started previously, the criterion for effective teaching should be student learning not the rigid use of strategies identified in the instructional model. The instructional model is a means to an end, not the end in itself. As Fullan (2001) explains the purpose of teacher observation is to produce shared knowledge through interaction which can be applied by teachers to address real-world issues in their classrooms. Hord (1997) echoes these comments noting that shared knowledge regarding instruction should translate into practical tools that can be used by teachers to enhance student achievement.

While it is true that teachers should be allowed flexibility in the instructional strategies they employ, it is also true that all teachers should be expected to produce "learning" in their classrooms. Stated differently, teachers should be allowed to

exhibit wide variation (i.e. have different profiles) as to various aspects of the instructional model they emphasize. However, there should be no variation in expectations about student learning from teacher to teacher.

To this end I recommend that data are systematically collected on students as well as teachers. For teachers, data would be collected on the extent to which they employ elements of the instructional model. Such data might be collected via supervisor observations and teacher self report. To this end a rubric like should be developed for each element of the instructional model. In previous works I have discussed the format for the design of effective rubrics (see Marzano 2006). Briefly, though, I recommend a 0–4 scale for which a score of three is the criterion score or the acceptable level of competence for a specific strategy.

Using rubrics for each element of the instructional model a profile of each teacher can be compiled through teacher self-report and observations by supervisors. Regarding self-reports teachers can rate themselves on a systematic basis and compile these ratings. Additionally, supervisors can make systematic observations of teachers. These two sources of data can be combined to construct a profile for each teacher regarding their use of the instructional model. This is shown in the first 11 columns of Fig. 4.

Figure 4 depicts feedback to teachers on an instructional model that has 10 components. S1 stands for strategy 1, S2 for strategy 2 and so on. A report like that in Fig. 4 would be generated for each school within a district. Each pair of rows in Fig. 4 represents the self-report and supervisor report data for a specific teacher regarding the 10 design questions for the model shown in Fig. 4. To illustrate consider the first two rows in Fig. 4. The teacher has provided self-report scores for each of the 10 strategies in the instructional model. The teacher simply rated himself on each strategy using rubrics that have been designed by the district or the school.

Teacher	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Pre/post Gain	Engagement	Student Learning
Teacher 1	2.5	2.5	3.0	2.0	2.0	3.0	4.0	3.5	3.0	2.5	.76	2.11	3.09
Supervisor			2.5			3.0			2.5				
Teacher 2	2.0	3.0	2.5	3.0	3.5	2.5	3.0	2.0	2.5	3.0	1.33	3.23	3.06
Supervisor		2.0			2.5	2.5							
School Average	2.34	2.46	3.01	2.77	3.11	2.67	2.84	2.91	3.01	3.12	.53	2.33	2.56
District Average	2.67	2.87	2.84	3.01	2.78	2.41	2.86	2.73	3.08	3.04	.47	2.77	2.41

Fig. 4 Summary data for teacher profiles. Copyright: Marzano & Associates, 2008

Supervisor ratings are right below teacher ratings. Note that supervisor ratings are not reported for all 10 strategies in the model. This is because a supervisor would not have the time to make valid observations on all components of a model in a single year. Consequently, in a given year, a supervisor in consultation with a given teacher would identify a few strategies to observe. Over the years, scores on all strategies would be obtained from supervisors. Also note that teacher names are not used. A composite report like that depicted in Fig. 4 lists all teachers but only individual teachers and supervisors know the identity of specific teachers. Finally note that the last two rows for each design question report district and school averages respectively. This allows comparison of individual teacher profile data with that for the entire school and the entire district.

To complete the profile for each teacher student data must be collected. I recommend at least three types of student data. The first is pretest/posttest data from a specific unit of instruction. These pretest/posttest scores should all use the same metric. I recommend a 0 through 4 scale that can be applied to any subject area (see Marzano 2006). The pretests and posttests can either be teacher made assessments or common assessments designed by the district. Another type of data is student self-report engagement data. Again I recommend a 0 through 4 scale. The final type of data is student self-report data on their learning again using a 0 through 4 scale. The last two columns in Fig. 4 show class averages for each teacher on the three types of student self-report data. As before the last two rows contain school and district averages for the student data.

It is certainly not the case that every teacher should be expected to meet or exceed district or school averages in all measures. Each teacher is unique in his or her instructional profile. However, comparison between individual teacher profiles and school or district averages should be the basis for discussion between teachers and supervisors. More specifically, individual teachers in consultation with their supervisors should identify specific goals for improvement. The focal point of such deliberation should always be student learning and engagement. An individual teacher might set a goal of raising her pretest/posttest achievement gain from.5 to.6 and raising the average level of student engagement by one half a scale point by the end of the year. The teacher might elect to focus on one specific instructional strategy to achieve these goals. For example, after examining her instructional profile as compared to the profiles of others in the district, the teacher might select designing clear learning goals and providing feedback on which to focus throughout the year. That selection might be made because the teacher notes that her scores on that particular strategy are significantly below the school or district average.

Conclusions

This article has made a case for districts and schools designing their own instructional models using action research. Once an instructional model is defined it should be used initially as the basis for teachers systematically discussing effective teaching. Next the instructional model should be used as the basis for observing master teachers who are operationally defined as those who consistently produce learning in their classrooms. Finally, data should be collected systematically on each

teacher regarding their use of the instructional model and its effect on student learning and engagement. Using this data teachers and supervisors would set goals regarding areas of improvement with the eye toward increasing student achievement.

References

Berliner, D. C. (1986). In pursuit of the expert pedagogue. *Educational Researcher*, *15*(7), 5–13.

Chiseri-Strater, E., & Sunstein, B. S. (2006). *What works? A practical guide for teacher research*. Alexandria, VA: Association for Supervision and Curriculum Development.

Cuban, L. (1987, July). *Constancy and change in schools (1880s to the present)*. Paper presented at the Conference on Restructuring Education: Keystone, CO.

De Leeuw, J. (2004). Senior editor introduction. In R.A. Berk (Ed.) *Regression analysis: A constructive critique* (pp. xi–xv). Thousand Oaks, CA: Sage Publications.

Dimmock, C. (2000). *Designing the learning-centred school: A cross cultural perspective*. London: Falmer Press.

Flinders, D. J. (1988). Teacher isolation and the new reform. *Journal of Curriculum and Supervision*, *4*(1), 17–29.

Fullan, M. (2001). *The new meaning of educational change (3rd edition)*. New York and London: Teachers College Press and RoutledgeFalmer.

Good, T. L., & Brophy, J. E. (2003). *Looking in classrooms (9th edition)*. Boston, MA: Allyn & Bacon.

Goodlad, J. I. (1984). *A place called school*. New York, NY: McGraw-Hill.

Hargreaves, A. (1994). *Changing teachers, changing times: Teachers' work and culture in the postmodern age*. London: Cassell.

Hattie, J. (1992). Measuring the effects of schooling. *Australian Journal of Education*, *36*(1), 5–13.

Hord, S. M. (1997). *Professional learning communities: Communities of continuous inquiry and improvement*. Austin, TX: Southwest Educational Development Laboratory.

Hunter, M. (1984). Knowing, teaching, and supervising. In P. Hosford (Ed.) *Using what we know about teaching* (pp. 169–192). Alexandria, VA: Association for Supervision and Curriculum Development.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis. *American Psychologist*, *48*(12), 1181–1209.

Little, J. W. (2002). Locating learning in teachers' communities of practice: opening up problems of analysis in records of everyday work. *Teaching and Teacher Education*, *18*(8), 917–946.

Louis, K. S., Kruse, S. D., & Associates (1995). *Professionalism and community: Perspectives on reforming urban schools*. Thousand Oaks, CA: Corwin Press Inc.

Marzano & Associates (2005). *Blue Springs School District instructional strategies project: Final report.* Centennial, CO: Author.

Marzano, R. J. (2006). *Classroom assessment and grading that works*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2007a). *The art and science of teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2007b). *Three critical interventions for district and school improvement*. Centennial, CO: Marzano & Associates.

Marzano, R. J., Pickering, D. J., & Marzano, J. S. (2003). *Classroom management that works: Research-based strategies for every teacher*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

Mayer, R. E. (2003). *Learning and instruction*. Upper Saddle River, NJ: Merrill, Prentice Hall.

Nolen, A. L., & Putten, J. V. (2007). Action research in education: addressing gaps in ethical principles and practices. *Educational Researcher*, *36*(7), 401–407.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin*, *112*(1), 160–164.

Reynolds, D., Teddlie, C., Hopkins, D., & Stringfield, S. (2000). Linking school effectiveness and school improvement. In C. Teddlie, & D. Reynolds (Eds.) *The international handbook of school effectiveness research* (pp. 206–231). New York: Falmer Press.

Riehl, C. (2006). Feeling better: a comparison of medical research and education research. *Educational Researcher*, *35*(5), 24–29.

Rimland, B. (1979). Death knell for psychotherapy? *American Psychologist*, *34*, 192.

Rosenthal, R., & Rubin, D. B. (1982). A simple general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, *74*(2), 166–169.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276–1284.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in past decades. *Review of Educational Research*, *77*(4), 454–499.

Shulman, L. S. (2004). *Teaching as community property: Essays on higher education*. San Francisco, CA: Jossey-Bass.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*, 752–760.

Steering Committee on the Physicians' Health Study Research Group (1988). Preliminary report: findings from the aspirin component of the ongoing physicians' health study. *The New England Journal of Medicine*, *318*, 262–264.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: a review of the literature. *Journal of Educational Change*, *7*, 221–258.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, *63*(3), 249–294.

Willms, J. D. (1992). *Monitoring school performance: A guide for educators*. Washington, DC: The Falmer Press.