

# MID-WESTERN EDUCATIONAL RESEARCHER

Official Publication of the Mid-Western Educational Research Association

Volume 24, No. 1—Winter 2011



The University of Akron  
College of Education, Zook Hall

## Inside this Issue:

Notes from the Editors' Roundtable

Politics, Economics, and Testing:  
Some Reflections

Coaching Conversations:  
Enacting Instructional Scaffolding

2011 Annual Meeting  
Call for Proposals

The Maintenance of Teacher Autonomy  
in a Policy-Driven Era

The Perfect Storm: How Policy,  
Research, and Assessment  
Will Transform  
Public Education

Facilitating a Faculty  
Learning Community:  
Determining Consensus  
using Q Methodology

MWERA



[www.mwera.org](http://www.mwera.org)

## *On the Cover*

It is a pleasure to welcome the *Midwest Educational Researcher* to its new home at The University of Akron College of Education. Akron is located in the heart of the so-called “Rust Belt”, but like many other cities in our region, has had to “re-invent” itself from a manufacturing economy to one that is vested in medicine, technology, and advanced scientific techniques (e.g., corrosion prevention). In order to accomplish this, our surrounding communities have relied heavily on The University of Akron to provide leadership in the STEM areas, including STEM education. Our College of Education is proud to be a co-sponsor of the National Inventors Hall of Fame STEM Middle School, and is working with the Akron Public Schools to launch a STEM high school in 2012. We were recently awarded a Woodrow Wilson Foundation Grant to re-tool STEM education at the master’s level, and look forward to a similar transformation for our undergraduate programs.

The foundation of our work and the work of many other educators is informed and driven by the types of publications that find their way to MWER. Applied and theoretical pieces broaden the knowledge bases that are, in turn, used to address the myriad of educational problems we face. The College of Education is honored to be associated with MWER and the work it is advancing.

Mark D. Shermis, Ph.D.  
Dean, College of Education  
University of Akron

*Cover photo taken by University of Akron student, Adrienne Janke, graphic designer and photographer in the Cleveland/Akron area.*



---

## *Notes from the Editors' Roundtable*

As the new editorial team for the Mid-Western Educational Researcher, we would like to take a moment to welcome back old friends, greet new ones, and share our vision for the journal over the next three years. We are delighted to have been selected to serve MWERA and the journal and believe strongly that there is an important and synergistic relationship between the organization and its journal. A strong, scholarly journal draws attention to an organization by promoting membership and participation in the annual conference as well as disseminating relevant and significant scholarship to the field of education. Likewise, a strong and vibrant MWERA helps to provide a presence for MWER. We believe that readership and membership are closely linked. Through our efforts to increase readership of MWER, membership and participation in MWERA will likewise increase offering a vibrant and collegial space for more educational researchers, teachers, and stakeholders to come together and share their work.

Guiding our work over our term as editors are four overarching goals. We intend to diversify the content of the journal, increase readership through online journal publication, publish high quality research-based manuscripts, and streamline and expedite the review process. We would like to broaden the scope of MWER to include other types of publications including book reviews, invited articles, editorial dialogues, and position papers. We anticipate this transition to multiple sections will take some time, and we invite you to review the new Call for Manuscripts in this issue for a description of the new sections and criteria for submitting manuscripts.

At the 2010 annual meeting in Columbus, Ohio, we began discussions about moving the journal to an online format. This

medium offers significant possibilities to increase readership, diversify and expand content, and illuminate our presence and contribution in the field regionally, nationally, and internationally. At the center of this decision is a focus on maintaining excellent scholarship, a strong and efficient peer-review process, and a high quality journal. As your editorial team, we want to assure you that we will be diligent in attending to the conventions and expectations of scholarly work in academe.

We have been hard at work since beginning our term officially in October at the annual meeting. We appreciate the confidence of the organization in our work and leadership with the journal. We look forward to serving you, our readers, and contributors for the next three years. Should you ever have any questions, comments, or feedback, please contact us at [MWER@uakron.edu](mailto:MWER@uakron.edu). We hope that you will consider submitting your work to the *Mid-Western Educational Researcher*, and continue to see it as a quality publication for your teaching and research.

Warmest regards,

Editorial Team

Jennifer L. Milam, Ph.D., Managing Editor

Kristin L. K. Koskey, Ph.D.

Susan N. Kushner Benson, Ph.D.

Xin Liang, Ph.D.

Sandra Spickard Prettyman, Ph.D.

**MWER Editors**

Kristin L. K. Koskey  
Susan N. Kushner Benson  
Xin Liang  
Jennifer L. Milam  
Sandra Spickard Prettyman  
*The University of Akron*

**MWER Editorial  
Advisory Board**

Shereen Abdel Kader  
*Harris-Stowe State University*

Doug Feldman  
*Northern Kentucky*

Melissa Grabner-Hagen  
*Indiana University Kokomo*

Rafa Kasim  
*Kent State*

Matt Priest  
*Bowling Green State University*

Kenneth Royal  
*American Board  
of Family Medicine*

LaTisha Smith  
*Harris-Stowe State University*

Penny Soboleski  
*University of Findlay*

Kristine Still  
*Cleveland State University*

Sharon Valente  
*Savannah College of Art  
and Design*

Rachel Vannatta  
*Bowling Green State University*

Jennifer Weber  
*University of Kentucky*

Judith Zimmerman  
*Bowling Green State University*

**Layout and Design**

Judy E. Teska  
*JET Desktop Solutions*

**Presidential Address**

***The Maintenance of Teacher Autonomy in a Policy-Driven Era***

*Doug Feldmann, Northern Kentucky University*

2

***Coaching Conversations: Enacting Instructional Scaffolding***

*Sharan A. Gibson, San Diego State University*

5

***2011 Annual Meeting—Call for Proposals***

21

***Luncheon Keynote Address***

***Politics, Economics, and Testing: Some Reflections***

*Michael J. Feuer, George Washington University*

25

***Facilitating a Faculty Learning Community:  
Determining Consensus using Q Methodology***

*Susan Ramlo, The University of Akron*

30

***The Perfect Storm: How Policy, Research,  
and Assessment Will Transform Public Education***

*James H. McMillan, Virginia Commonwealth University*

39

***Call for Manuscripts***

48

**MWER Publication Address**

Jennifer L. Milam  
The University of Akron  
321 Zook Hall  
Akron, OH 44325-4205  
Phone: 330-972-8538  
e-mail: MWER@uakron.edu

**MWER Subscription &  
MWER Membership Information**

Sharon McNeely, Ph.D.  
P. O. Box 34421  
Chicago, IL 60634-0421  
Phone: 773-442-5518  
Fax: 773-736-7033  
e-mail: mwera\_ed@yahoo.com

## The Maintenance of Teacher Autonomy in a Policy-Driven Era

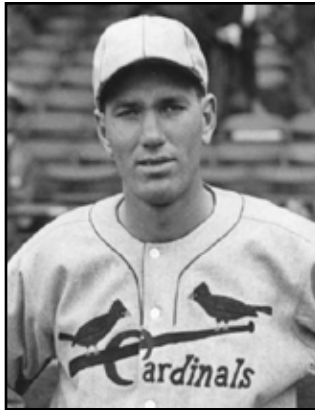
Doug Feldmann  
Northern Kentucky University

Colleagues, it is my pleasure to address you and the MWERA membership. Over the past several years, I have been honored to serve with several tremendous individuals on the organization's board of directors, and I look forward to being involved with MWERA for many years to come.

Baseball pitching great Dizzy Dean once said of his abbreviated formal education, "I dropped out of school in the second grade—but I didn't do too well in the first grade, either." To be sure, "dropping out" has unfortunately become a phrase to which educators are accustomed. However, in recent decades, the phrase has been more often attributed to *the educators* themselves.

I submit that those going into the education field in contemporary times are no longer overly concerned with traditional discouragements such as low salaries; rather, what appears to be desired today by teachers—both novice and veteran—is the *freedom* to make reasonable curricular and instructional choices, and the *administrative support* to enact such choices. When widespread public education began to emerge in the United States in the latter half of the 1800s, teachers often had complete freedom, as the materials and methods utilized often consisted of only what teacher possessed, and the reading material often consisted of what the students could bring from home; now, it may appear that we are inching toward the other end of that spectrum, with the day-to-day operations of the school setting being highly-regimented and compartmentalized. Few administrators would argue with the idea of providing their teachers with as much freedom and support as possible—yet, how does one balance these pursuits with an equal measure of alignment and responsibility?

This lack of freedom and support, in my opinion, has contributed to increasing attrition rates in the teaching profession. In an anecdotal example, a curricular study I conducted of a rural Midwestern school system a few years ago displayed this sentiment, as teachers relayed their amazement at the lack of confidence their students had in the continuity of the staff. "It's unbelievable how many kids come up to me and ask, 'Are you going to be here next year?'" one such teacher said to me. "They just expect you not to come back, and I don't think that's very good... I never asked my teachers if *they* were coming back; I just expected them to be there." (Feldmann, 2003, p. 144).



Dizzy Dean

To me, this suggests a dire need for deeper and broader research into the idea of *Teacher Well-Being*. Politicians and administrators rightfully display concern about *student* well-being (and they certainly look after *parent* well-being very carefully, as parents are voters), but there is little empirical

evidence to suggest there is widespread concern about whether teachers find fulfillment in what *they* do. To this end, there are three questions I wish to pose to you today, along with (hopefully) a few possible solutions:

- How do we define "Teacher Well-Being?"
- What consists of a "professional community" to help promote this well-being?
- Are we currently using research and policy to impact these ideas?

With each passing year, an increasing amount of policy pressure appears to be applied to public school educators. Not only are teachers expected to adapt to perpetual curricular changes, but now are on the verge of being even more acutely assessed by student performance on standardized tests (masked in a contemporary euphemism known as the "Value-Added Model"). But as with any career, finding well-being for teachers begins with finding a true, discernible, professional *purpose* to what they do.

But unfortunately, at very few times in the history of the United States has there been an open, honest discussion—and let alone, subsequent agreement—on what we specifically want our public schools to be, to do, and to produce. In the 1840s, Horace Mann launched such a discussion with his movement for the common schools; in the 1920s Franklin Bobbitt offered a very clear vision of what he saw as the purpose of schooling, as he suggested that society be scientifically studied as the source of objectives for the curriculum. At the height of the Great Depression in 1932, George Counts attempted to jolt the Progressive Education Association into a clearer purpose with his famous treatise, "Dare the School Build a New Social Order?" And as another example in 1983, American society was thrust into a similar academic discussion with the issuance of *A Nation at Risk*, the scathing report on the nation's scholastic operations in which our educational practices appeared to be falling behind many other countries.

Yet, within almost all of these conversations, those who were leading the discussions nonetheless encountered the difficulty—as do educators today—in convincing the

---

general public that much good can come from a student's school experience that falls **outside the realm of his or her standardized test score**. If the American public could finally agree on some simple, attainable concepts and goals on which schools and teachers could focus—and then leave the educators alone to do their jobs, a professional courtesy given to professionals in almost every other line of work—the goals could certainly be more easily and more quickly attained.

Furthermore, I am of the opinion that if we are going to have teacher performance tied to student test data, as suggested by the Value-Added Model—and have that “performance” made public—we should also make public the data on how well each community supports its school system. While it is granted that some communities' resources for financial support are limited, it would nonetheless be interesting to note which communities had the “best scores” in other aspects of supporting their local schools—such as the percentage of parents or guardians who attend scheduled meetings with teachers, the percentage of volunteerism at the local school, etc.—I think this would make for very intriguing data to be shared in the local paper as well! To be frank, it is long past time that the general public halts its abuse of teachers.

Thus, to accomplish a sense of *purpose* in the profession of teaching—and hence to balance *teacher freedom with teacher responsibility*, and hopefully to therefore achieve some means of *well-being* for those in the profession of teaching—I suggest that the following must be pursued:

### **Cultivate Creativity**

within a school for teachers and students

Reasonable limits must be placed on school administrators in their voracious appetites for “common syllabi,” “common assessments,” and the like. If administrators are truly to be advocates for their classroom teachers, they will more often than not (and when possible) defer to the teacher's professional judgment on curricular issues. For while having the entire Social Studies department subscribe to the exact same syllabus, assignments, and assessments might make for better politics and public relations in terms of “alignment” of the junior-year American History course, a greater good would be accomplished in permitting the classroom educator—those uniquely qualified to teach the subject, *as approved by the state and hired by that very same administrator*—to make a certain amount of instructional choices.

In many states, teachers are required to earn a master's degree; but I find it ironic that, in many of these same states, these same people with master's degrees are not given the basic courtesy and respect of making reasonable curricular and instructional choices in their chosen profession. It is the perfect example of public school teachers being used as political “footballs,” punted back and forth under the whim and exploitative nature of bureaucrats.

To empower teachers to experience a greater sense of autonomy and professional latitude, I also suggest the idea

of school districts hosting a “Night of Teacher Creativity.” While often conducting “open houses” that display student work for the public, districts could also have an evening in which educators can share innovative teaching strategies with the public they are utilizing in their classrooms. This, I feel, could more clearly convey the message to the local citizen that many valuable lessons for students—again, outside the realm of the standardized test—can be achieved when a teacher's creativity is released.

### **Cultivate Aspiration**

within a school for teachers and students

If a teacher is confined to the stipulations of common syllabi or common assessments, it is difficult to imagine such a teacher feeling compelled to aspire to higher levels within his/her profession. Can one truly see avenues for growth in one's career if each moment and decision is dictated from above? This question must be embraced by the building administrator on behalf of his/her teachers, as well as the classroom educators themselves. This is not to suggest that every teacher must pursue administrative positions, for leadership and growth within one's profession can take many forms. The same idea of aspiration must be cultivated in students to meet this challenge as well, as those *receiving* the education must envision goals for its use later in life.

### **Cultivate a Work Ethic**

within a school for teachers and students

Closely along the lines of aspiration, a work-ethic culture must also be furthered by a school system for teacher and student well-being. Nearly each of us above the age of 40 worked a job for an entire summer as a young person; how many of us know of such a young person today? Having a strong work ethic was part of our idea that, as children, we were told we could be “whatever we want to be when we grow up.” Now, we offer the same deal to the current generation of students; however, do today's children (and even the parents or guardians responsible for them) understand the large gulf of hard work that must be crossed before these future dreams become reality? Perhaps the other half of the story we need to tell them is the path required *to get to those goals*—perhaps attained by affixing a second statement to that quote, such as “the price of success is hard work,” as the famous football coach Vince Lombardi once said. Or, as my own college football coach always told us, “You're always either getting better, or getting worse.” And he wasn't talking about football—he was talking about all aspects of life.

A hard-work ethic, in both teachers and students, will help to foster the sense of purpose in a school which I discussed earlier. In the high school in which I taught, I felt this sense of purpose had strongly permeated the entire building; and in looking to contribute to this attitude, I posted a sign above my classroom door for my students to consider:

“DO IT            DO IT RIGHT            DO IT RIGHT NOW”

---

A popular contemporary mindset preventing the advancement of creativity, aspiration, and work ethic is the *fear of failure*. An example of this mindset is seen in the typical little league baseball game, in which players are patted on the head and told that it's "okay" to strike out, as we are afraid to damage the child's self-esteem. *Actually, it is not okay!* Striking out is indeed a failure in baseball; but Babe Ruth himself struck out 1,330 times. He learned from these failures, however, to become perhaps the greatest hitter the game has ever known. Those of us in education understand the greater destruction that can occur in children when we praise and reinforce failure, such as telling a student it's okay to say that "2 + 2 = 5." Moments of failure can be among the most instructive moments of our lives; in a like manner, school districts must permit teachers to learn from their mistakes, and must understand that it is part of professional growth—again, just as in any other profession.

I am blessed to have two parents who are educators. In a letter he wrote to the local newspaper upon his retirement in June 1986, my father—most succinctly and effectively—summarized the pursuit I have outlined today in attempting to strike a balance between professional autonomy and responsibility in teaching:

*Dear Editor:*

*On June 8 of this year, it will mark my last day in the \_\_\_\_\_ Public Schools as a teacher, coach, administrator, and counselor. There have been many pleasant years, and I want to share with the public a few genuine feelings before I walk away.*

*First of all, my appreciations are in order for a district which has promoted an educational climate; one to cherish; one that extended me considerable freedom but still required a certain amount of responsibility.*

*To the students of \_\_\_\_\_, \_\_\_\_\_, and \_\_\_\_\_ High School, I will miss and remember these associations at school as I reflect on a goal I set a long time ago: to be a positive influence on everyone I knew, even perhaps in an incidental way. Whether this came to be I will never know, but I would like to think so.*

The "incidental ways" in which my father was a positive influence, I believe, were the result of a combination of his goodness and the freedom extended to him by the administrative teams in the buildings in which he was fortunate to work. He cites this balance by stating in the middle paragraph that the district "extended me considerable freedom but still required a certain amount of responsibility." ...Were it so for every district!

As we experience the perpetuation of change in public education, let us leverage it in the direction of those who have dedicated their lives to the job—the classroom teachers. We have to adapt to change, for it is imminent in our profession; we cannot rely on the mindset of the past if we wish to attract a new generation of talented individuals into careers in teaching.

For as Dizzy Dean also said when he was struggling to maintain his pitching greatness in the twilight of his career with a sore arm, "I ain't what I used to be—but what is?"

## References

Feldmann, D. (2003). *Curriculum and the American rural school*. Lanham, MD: University Press of America.

---

# *Coaching Conversations: Enacting Instructional Scaffolding*

Sharan A. Gibson  
San Diego State University

## ***Abstract***

*This study analyzed coaching conversations and interviews of four coach/teacher partnerships for specific ways in which kindergarten and first-grade teachers, and coaches, conceptualized instructional scaffolding for guided reading. Interview transcripts were coded for coaches' and teachers' specific hypotheses/ideas regarding instructional scaffolding. Coaching session transcripts were analyzed for coaches' and teachers' actual use, or enactment, of instructional scaffolding. Significant tensions were evident between hypotheses describing the need for high levels of instructional support versus opportunities for students to read independently. Teachers' expertise for effective instructional scaffolding appeared to be assisted by coaching conversations that enacted instructional scaffolding, demonstrating an analytic, evidence-based approach to instructional problem solving.*

Receiving instruction from knowledgeable, skilled classroom teachers is one of the most powerful interventions available to young or naive readers (Darling-Hammond, 1999). Teaching reading well is a complex set of tasks requiring commitment and self-reflection grounded within a problem-solving stance (Frager, 1994; Gibson, 2010), in-depth knowledge of specific instructional strategies and underlying theories of development, and the ability to interact effectively with students' literate thinking during instruction (Ross & Gibson, 2010). In recognition of these challenges, reform efforts rely on the promise of coaching, calling for a critical mass of highly accomplished teachers in leadership roles (Taylor, Pearson, Clark, & Walpole, 2002).

Research documenting the ability of coaching to effect positive change in teachers' instructional practices, however, is mixed (Neufeld & Roper, 2003; Poglinco, Bach, Hovde, Roseblum, Saunders, & Supovitz, 2003). Technical "level 3" coaching (e.g., lesson observation and feedback; Bean, 2004) is one of the most challenging roles faced by coaches (Poglinco, Bach, Hovde, Roseblum, Saunders, & Supovitz, 2003). Limited information is available on the specific ways in which effective versus ineffective coaches interact with teachers following lesson observation. The Self Assessment for Literacy Coaches (Literacy Coaching Clearinghouse, 2009), for example, addresses crucial issues related to coaches' knowledge and leadership abilities, but it does not describe coaching conversations beyond planning, pre-meeting, observation protocols, and reflective dialogues.

Research regarding the conversational interaction between coaches and teachers following lesson observation has not yet clearly delineated the specific expertise needed by effective coaches. Strong and Baron (2004) found that veteran mentor teachers rarely provided direct advice to novice teachers during coaching conversations. The researchers speculated that this approach placed a heavy demand on novice teachers' ability to construct instructional behaviors and activities independently. Peterson, Taylor, Burnham, and Schock (2009) found that literacy coaches were able to utilize data documenting student responses during instruc-

tion to engage teachers in self-reflection. The researchers, however, did not discuss coaches' differential success with this process. Gibson (2006) documented a literacy coach's maintenance of an expert stance for coaching conversations implemented within a collaborative framework. This single case study did not contrast the specific ways in which coaching conversations varied for more or less effective coaches.

The ability to engage teachers in conversations that are focused on analysis of students' immediate responses to instruction is central to the work of coaches (Bean, 2004; Rodgers & Rodgers, 2007). Teacher educators and coaches need information on the characteristics and outcomes of such conversations. The study described here meets this critical need by investigating one important aspect of the interaction of reading coaches with teachers following lesson observation, and it presents a rubric for coaches' self-evaluation and professional development. This study investigated the ways in which coaches and teachers both conceptualized and enacted instructional scaffolding during coaching conversations within the context of teachers' guided reading lessons.

## *Instructional Scaffolding*

Instructional scaffolding consists of assistance provided by a more expert person for the performance of a particular task, resulting in the learner's internalization of ways of conceptualizing and acting (Roehler & Cantlon, 1997; Tharp & Gallimore, 1988; Wood, Bruner, & Ross, 1976). Instruction is most efficacious when more help is provided as soon as a student is struggling, and the teacher then either withdraws support or raises the level of challenge contingent on student success (Wood, 1998). Instructional scaffolding constitutes a fundamental shift in teachers' interaction with students; what the teacher does is dependent on what students do on a moment-by-moment basis (Rodgers, 2004). Scaffolding is also a dialogic, socially based joint venture (Meyer & Turner, 2002; Palincsar, 1986). Teachers support students in aspects of tasks that they cannot yet complete on their own, while students become more adept at judging the results of their performance and arranging for assistance (Wood, 2003).

---

Learning how to provide effective instructional scaffolding to students requires more than simply knowing what scaffolding is. Many, Dewberry, Taylor and Cody (2009) found, for example, that preservice teachers' implementation of reflective instructional scaffolding was dependent on their level of understanding of language and literacy development and ability to make complex instructional decisions using multiple knowledge sources.

*Guided reading lessons.* Instructional scaffolding is an important requirement for effective guided reading instruction. Guided reading is a small group context that supports students' development of strategies for successful processing of increasingly difficult texts (Fountas & Pinnell, 1996). This instruction is effective in developing children's ability to use graphophonetic information, produce semantically and syntactically acceptable miscues, and retell texts cohesively and accurately (Altwerger, et al., 2004). Teacher decision-making (both preplanned and "in-the-moment") is crucial. Teachers select a "just right" text for each lesson and provide a book orientation and prompting during students' independent reading, directly targeted to students' immediate needs. Within each guided reading lesson, the teacher's decisions moderate the level of difficulty. Teachers consider students' emerging expertise, utilizing modeling, questioning and explanation in order to teach appropriate strategic behavior. Virtually all decisions made for guided reading instruction, then, concern the type and/or amounts of scaffolding provided to students. Although they may do so with more or less effectiveness (and with or without explicit intent), it seems clear that reading coaches will address aspects of teachers' use of instructional scaffolding as they engage in coaching conversations.

*Coaching conversations.* One of the primary ways in which coaches may provide assistance to a teacher is through supportive conversations following the coach's observation of a lesson. These conversations are typically conducted between the coach and teacher shortly after lesson observation and include discussion of what actually occurred within the lesson and how instruction might be refined to meet the needs of students.

Instructional scaffolding is an important aspect of both the instruction provided to students during guided reading lessons and the discussion between a coach and teacher during a subsequent coaching session. Influential literacy teachers are able to monitor their students' literate thinking during instruction and use this information to adjust instruction and scaffold students' new learning (Ross & Gibson, 2010; Ruddell, 2004). Coaches also provide scaffolding for tasks for which *teachers* are not yet expert, implementing modeling and assistance to teachers on a contingent basis. For guided reading, teachers' analysis of students' immediate responses to instruction and their ability to connect this information directly to needed changes in instruction is key to their instructional effectiveness. Similarly, where a coach is enacting instructional scaffolding during coaching conversations, analysis of students' strengths and weaknesses directly

connected to proposals for improving instruction would be evident in the talk of the coach and/or teacher.

This study describes the ways in which four coach/teacher partnerships conceptualized and enacted instructional scaffolding, following the coach's observation of guided reading lessons. Data sources included three cycles of individual interviews and observation of coaching sessions. Transcripts were qualitatively coded, and coding was examined to ascertain coaches' and teachers' viewpoints on instructional scaffolding over time. Data collection and analysis were based on the following research questions:

1. What hypotheses and/or ideas about instructional scaffolding do coaches and teachers articulate during coaching conversations and individual interviews?
2. In what ways did coaches' and teachers' enactment of scaffolding within coaching conversations support teachers' ability to reflect on students' need for instructional scaffolding during guided reading lessons?

## Method

### *Participants*

A small urban public school district that had recently implemented a district-wide, long-term professional development program for K-2 literacy instruction was the site of this study. This school district was chosen as a purposeful, critical case based on the district's implementation of significant training and on-going support for a literacy coach at each elementary school in the district. A classroom teacher at each of seven elementary schools in the district had transitioned to a full time literacy coach position, charged with implementation of a K-2 instructional framework (i.e., interactive read aloud, shared and interactive writing, word study, content area connections, guided reading, independent language and literacy work, and writing workshop). Each of these coaches had received university-based training and support for this role, which included (a) teaching a 40-hour, on-site class on literacy instruction for K-2 teachers at their own school site, (b) providing in-class coaching to all K-2 teachers, and (c) continuing to teach children for at least 90 minutes per day. The coaches' expertise was supported indirectly through professional development sessions that included feedback following viewing of videotapes of their own instruction, as well as directly through presentation of a coaching framework (e.g., pre-conference, note taking, feedback and coaching, and written plan of action for next lessons). All seven reading coaches in this district were experienced classroom teachers who had completed seven weeks of university training over a one-year period, as well as a subsequent field year in their literacy coaching positions. A convenience sample was utilized for this study, consisting of the four coaches who agreed to participate. All seven coaches had participated together in the same university training experiences, and they were equally experienced in the coaching role.

Each of the four coaches recruited a kindergarten or first-grade teacher for whom he/she was providing individual coaching sessions following observation of guided reading lessons. The choice of a kindergarten or first-grade teacher was made by each coach, based on her perception of the ability of a teacher to participate in the study without undue stress. The two kindergarten teachers who participated in the study had 3 and 25 years of experience and were teaching full day kindergarten (see Table 1). Neither teacher had taught guided reading groups prior to the year of this study. The two first-grade teachers who participated in the study were both in their second year of teaching. These two teachers had limited experience teaching guided reading lessons prior to the start of the study.

### Data Collection

Three cycles of observation and interviews were conducted between November and April with each coach/teacher partnership. Each data collection cycle consisted of (a) video recording of a classroom guided reading lesson, (b) observation and audio recording of a coaching session, (c) video recording of a second guided reading lesson, and (d) separate, audio taped interviews with the classroom teacher and coach. Each of the four coaches and four teachers was individually interviewed within each of the three cycles of data collection. These 24 individual interviews were structured both as stimulated recall (Keith, 1988) and in a standardized open-ended interview format. As a stimulated recall interview, a short segment of the videotape of the guided reading lesson or audiotape of the coaching session was played for the teacher or coach, who was asked to comment on his/her decision-making. The standardized open-ended format used with stimulated recall for each interview consisted of a predetermined sequence of standard questions asked of each coach or teacher (Patton, 1990). The following are examples of the stimulated recall and standardized questions asked of each teacher during the second cycle of data collection:

1. What I would like to do now is to learn more about your current thinking processes related both to your teaching of guided reading lessons and of coaching. I'm going to show you a short video clip from your lesson. As you watch it, I want you to reflect on how your teaching for guided reading has been going over the last month or

so. Then I'd like you to tell me about your thinking and the decisions you made today as you taught this lesson.

2. What do you do well/not well in your teaching of guided reading lessons?
3. You and [the coach] have been talking about and reflecting on \_\_\_\_\_ during your coaching sessions. What is your current thinking on that now and how are you doing with that in your teaching?

The following are examples of the stimulated recall and standardized questions asked of each coach during the second cycle of data collection:

1. What I would like to do now is to learn more about your current thinking processes for your work as a coach. First, I'd like you to listen to a portion of your coaching session from yesterday. As you listen, I want you to reflect on and talk about your thinking and the decisions you made as you interacted with [the teacher] about guided reading lessons.
2. In the past, you and [teacher] have talked about \_\_\_\_\_. How is that going for him/her?
3. What would you like him/her to come to understand now about guided reading instruction, and how do you feel that you can be of best help to [the teacher]?

### Data Analysis

All coaching session and interview transcripts were analyzed to determine the ways in which instructional scaffolding was discussed and enacted by both coaches and teachers. Each of the following data analysis steps used in the study is defined and illustrated in further detail in the sections below:

1. First, all segments within coaching session and interview transcripts that addressed instructional scaffolding were identified.
2. Each of these segments that had been identified as scaffolding-related was then coded for general topic and specific hypothesizes/ideas.
3. A chart was created for each coach/teacher partnership summarizing a timeline of ideas and hypotheses over time.
4. All conversational turns within each of the coaching session transcripts were coded for the type of interaction

Table 1  
*Participants*

Grade Level	Coach	Teacher	Teaching Experience	Grades Taught
First Grade	Kate	Sherrie	19 years 2 years	First, third, fifth, sixth First
First Grade	Rose	Charles	31 years 2 years	Kindergarten, first, second First
Kindergarten	Kristi	Melinda	12 years 3 years	Kindergarten, first, second Kindergarten
Kindergarten	Carol	Daniel	17 years 25 years	First, second, third Kindergarten, fourth

---

that occurred, defined by coaches' and teachers' enactment of instructional scaffolding.

5. Each coded coaching session transcript was then evaluated for coaches' and teachers' overall enactment of instructional scaffolding.

Grounded theory and open coding procedures (Strauss & Corbin, 1998) were used for data coding. The intent of grounded theory is to insure that all theories or findings emerge from the researcher's systematic coding/categorization of data. Open coding breaks data down into discrete segments, which are then carefully examined and compared for similarities and differences. Grounded theory was an appropriate choice for data analysis for this study based on the need to derive in depth understanding of each coach and teacher's concepts, hypotheses, and/or ideas about instructional scaffolding over time.

*Analysis of interview transcripts.* First, all statements within interview transcripts that addressed instructional scaffolding were identified. These scaffolding-related statements addressed types of instructional support needed/not needed by students or teachers, based on the definition of scaffolding as more help provided as soon as a student is struggling and withdrawal of teacher support or increase in challenge contingent on student success (Wood, 1998). Because each interview was conducted in response to a specific guided reading lesson and subsequent coaching session, most interview statements addressed the topic of scaffolding. In the following statement, for example, the teacher described a decision she had made regarding both preplanned and in the moment instructional scaffolding in support of her students' decoding of the word *asked* (within the new book read that day):

I had in my plans, originally, to have [students] predict and locate the word *asked*. But then I just abandoned [this task] because I thought, "They're going to just get that [word]." And if they don't, we'll use that as a teaching point [after reading].

This statement was identified as scaffolding-related because the teacher described a decision she had made to provide less instructional support for students' decoding of the word *asked* based on her prediction that such support would not be needed. In the following example, the coach described her analysis of students' success with the text the teacher had selected for that lesson:

I really didn't think that the level of difficulty [of the new book], I mean they [i.e., the students] breezed right through the book. And so there weren't a lot of things, I guess, trouble spots [for students as they read the new book].

This statement was identified as scaffolding-related because the coach described observed evidence demonstrating that students needed additional scaffolding from the teacher. A teacher's choice of a new text for guided reading lessons constitutes scaffolding, as a strong book choice provides a context within which students will need to use emerging

strategies but will be able to do so successfully with teacher support.

All statements identified as scaffolding-related were then coded for general topic and specific ideas concerning instructional scaffolding expressed by the participant. The following statement, for example, was coded as the topic of teacher prompting during text reading and the hypothesis that teachers need to use a wide variety of prompts in order to provide more effective support to students:

[The teacher] does have the prompt sheet [e.g., a list of such teacher prompting language as "Does that word look right?"] right in front of her. And as she was asking yesterday, "Which one should I be using? Because the [prompting language] I'm using doesn't seem to be effective." So as you learn more about those prompts and [in] one of our upcoming classes I want to tape guided reading and have them look at the prompts the teacher uses. To see, "Would you have used the same prompts? What else could you use?" Not just stuck with one or two of their own favorites. And I know from my own experience that that can easily happen. I'm thinking, "Why is this child stuck? Why isn't [my prompting] working?" And as you look at yourself you think, "Well because I'm throwing the same prompt at him every time."

This example was coded for the topic of teacher prompting because the coach described her ideas about teachers' choice of prompting language to use when students encounter word identification difficulties during text reading and the hypothesis that teachers need to use a wide variety of prompts in order to provide more effective support to students because the coach stated that that students' progress will stall if teachers use the same prompt repeatedly.

All coding for the general topic and specific ideas concerning instructional scaffolding expressed within each transcript were charted by coach/teacher partnership and data collection cycle. A chart was created for each coach/teacher partnership summarizing a timeline of prominent ideas regarding instructional scaffolding articulated by each coach and teacher across the three data collection cycles. This process addressed the study's first research question: What hypotheses and/or ideas about instructional scaffolding do coaches and teachers articulate during coaching conversations and individual interviews?

*Analysis of coaching session transcripts.* Each conversational turn within coaching session transcripts was coded on three levels: (a) general topic, (b) interaction function, and (c) the specific idea(s) concerning instructional scaffolding expressed by the participant. The following segment from a coaching session transcript, for example, was coded as (a) teacher prompting during text reading, (b) coach's replaying of the teacher's delivery of instruction, and (c) the proposition that high levels of teacher-provided support are appropriate for the decoding of proper nouns:

- T: So it sounds like the teaching point thing is something...
- C: Well, I'm thinking that where you, right in the very beginning, you prompted her to think about what's happening in the story and when she had difficulty with the names you popped right in there and provided that support for her so she wasn't stuck on trying to sound out *Michael* and *Anna*.

This example was coded for (a) teacher prompting during text reading because the coach "replayed" the ways in which the teacher had prompted a student to pay attention to the events in the story, (b) coach's replaying of the teacher's delivery of instruction because the coach described how the teacher had provided "tolds" for several characters' names from the text, and (c) the proposition that high levels of teacher-provided support are appropriate for the decoding of proper nouns because the coach stated that the type of support provided prevented the student from getting stuck on proper nouns.

Coding for interaction functions, then, was based on analysis of each coach and teacher's enactment of instructional scaffolding during coaching conversations. The researcher evaluated the function (i.e., asking, telling, analysis, description, summarizing) of each conversational turn. Coding categories that emerged during this process included the coach's or teacher's analyses or requests for information regarding observed student literacy behavior, replaying of instructional moves, analysis of the effectiveness of instructional scaffolding provided, and description of needed instruction.

Criteria were then developed to allow for characterization of the type of conversational interaction that occurred during coaching sessions, analyzing for coaches' and teachers' overall enactment of instructional scaffolding: If the coaches and teachers enacted instructional scaffolding, what processes would they have engaged in?

- Did the coach ask the teacher to analyze the instructional support that had been provided in the lesson?
- How extensive were the coach and teacher's descriptions of students' literacy behavior?
- Were any descriptions of needed instruction tied explicitly to evidence of student literacy behavior?

Coaching session analysis, then, determined the ways in which the coach and/or teacher's examination of students' strengths and weaknesses (directly connected to recommendations for improving instruction) was evident/not evident during each coaching conversation. These results, considered in concert with the timeline of prominent ideas regarding instructional scaffolding across the three data collection cycles, informed the study's second research question: In what ways did coaches' and teachers' enactment of scaffolding within coaching conversations support teachers' ability to reflect on students' need for instructional scaffolding during guided reading lessons?

The coaches in this study were able to engage teachers in conversations about reading instruction. Each of the coaching conversations addressed specific aspects of the observed guided reading lesson. Coaches, for example, discussed book selection and introduction, prompting for students' use of strategies during reading, and recommended changes in instruction. In spite of this achievement, analyses identified potentially consequential, largely unresolved differences in the ways that coaches and teachers conceptualized instructional scaffolding. Significant tensions were evident between hypotheses describing the need for high levels of instructional support versus opportunities for students to read independently. The coaches were generally not able to implement coaching conversations that both enacted and resolved teachers' understanding of instructional scaffolding.

The following sections address the two research questions for the study by presenting examples that exemplify prominent hypotheses and ideas that emerged from analysis of each coach/teacher partnership's propositional statements regarding instructional scaffolding and characterize the ways in which each partnership enacted instructional scaffolding during their coaching conversations. Prominent themes that emerged from data analysis (as described below) included how teaching for student independence is constructed within guided reading lessons, whether instructional scaffolding is a harmful crutch or needed support for students, what kinds of support will result in students' 'real' reading, and how support for students' problem-solving during reading is best structured. Results from two of the coach/teacher partnerships are presented in some detail, while results for the remaining two partnerships are summarized only briefly.

#### *Teaching for Independence: Kate (coach) and Sherrie (first-grade teacher)*

Analysis of coded interaction revealed that Kate (pseudonyms used throughout) maintained a focus on instructional scaffolding in all three coaching sessions and did request analysis from Sherrie of the degree to which her instructional decisions had been effective:

Okay. When we had talked about Stephanie and Moriah before, you had concerns about, especially Moriah, being visually balanced. And you talked about, "What are the prompts that you are using?" That was what you wanted me to look at. So as you think back to your lesson, how do you think it went?

Kate's requests for analysis were typically connected to specific concerns that she had identified during her observation of the lesson. Neither the coach nor teacher engaged in extensive description of students' literacy behavior. These findings are also illustrated in the coded coaching session example provided in Table 2.

Kate's recommendations generally suggested specific instructional interaction:

Sherrie: That's one [her student's difficulty reading the word *shouted*], I was trying to think, because you can't look at the picture....so I was at a loss for that.

Kate: If you think about the prompts...I may have taken her back to the previous page and talked about what has happened here and how he felt about that.

Sherrie: So getting her sense of story, that...

Kate: Yes. And you were at a loss, so you just [told] it to her, which is fine. But you might try pulling her back to that book. "Just stop and tell me what happened so far in this story. Now how do you think he felt?"

Try and check the picture."

Sherrie herself contributed minimal description of needed instruction during these coaching sessions, but consistently requested clarification on effective prompting: "How does that work with compound words and explaining that to them? To start with the second [part of the word]. Can you say that? Just find that main chunk and then start the next sound?"

Sherrie felt her students needed to utilize visual cues and word knowledge more independently in order to progress to harder texts. She commented during the third coaching session that she had deliberately limited her scaffolding because

Table 2  
Coded Coaching Session Excerpt: Kate and Sherrie

Participant	Transcript	Coding
Kate:	As we talked yesterday and you had asked me to help you look for those opportunities to find those teaching points, as you think back through your lesson and you may not be able to recall this, if you can't that's okay. Did you see any particular point where there was an opportunity for a teaching point?	Coach providing replaying Previous coaching  Coach requesting analysis Needed scaffolding
Sherrie:	For the whole group? So that it's directly to the whole group or individuals?	Teacher requesting explanation Recommended instruction
Kate:	Either. Either way.	
Sherrie:	Gosh. You know I'd have to look at my notes and I didn't bring them back.	Teacher requesting analysis Student literacy behavior
Kate:	Okay. Well as I was watching, one of the things that I saw with, they had trouble with the word <i>inside</i> .	Coach providing description Student literacy behavior
Sherrie:	Oh yeah. They got the first part [of the word] but they didn't, couldn't go on from there.	Teacher providing description Student literacy behavior
Kate:	Okay, and [the word] <i>in</i> is on your word wall.	Coach providing replaying Instructional moves
Sherrie:	Um hmm.	
Kate:	So as I was sitting there, I was thinking that you could either stop right there at that point, and look back at the word wall and have them, "You know the chunk <i>in</i> , you've got that, it's on our word wall. Go back and read, check the picture. Because the picture, they're now inside and go past [the word part] <i>in</i> ... and see."	Coach providing description Recommended instruction
Sherrie:	Okay.	

---

she wanted to see if students could “get [the words] on their own.” Kate, in response, emphasized the role of supportive book introductions:

The purpose of going through the pictures is to plant the whole idea of what the story is about. But also, simultaneously, is an opportunity for you to connect in context words that you feel you want to deal with.

During coaching sessions Sherrie requested practical information on how to improve her students’ use of graphophonic information, while Kate emphasized students’ use of meaning.

This tension was also evident during their discussions of book introductions. In the cycle one interview, Sherrie expressed concern that students might rely too heavily on instructional support: “I’m trying to wean them a little bit on my introductions where they’re doing more of the thinking.” She continued her reflection on the balance between book introductions and teaching for independence within the cycle two interview:

When I did the book with them on Monday and Kate observed, I think that they brought up a few memories and then I just gave them the book. And let them look through [the pages] and then let them start reading. I didn’t go through the pages like I’ve done in the past. And Kate was a little concerned that I had done that. And you know my thinking, that that’s giving them a chance to try to develop that comprehension independently. Giving them a little bit of independence.

Sherrie also emphasized that she needed to hear each student’s retelling of the story’s events in sequence, instead of providing book introductions. Kate advised Sherrie, however, to continue providing supportive introductions and to teach comprehension after read aloud sessions.

*Summary.* Kate and Sherrie presented opposing hypotheses for several important issues: (a) explicit instructional scaffolding versus assessment and independence, (b) attention to meaning versus the development of graphophonic knowledge, and (c) support for comprehension through book introductions versus retelling assessment (see Table 3 for a summary of these findings). During these coaching sessions, Kate typically requested input from Sherrie when she had identified an area of concern in her teaching, and neither the coach nor teacher related explicit, in-depth discussion of students’ responses to needed changes in instruction.

#### *Crutch or Needed Support: Rose (coach) and Charles (first-grade teacher)*

Across the three cycles, Rose and Charles also articulated a set of contradictory ideas about instructional scaffolding. Charles stated that scaffolding is a crutch that will interfere with student progress, and that students should be given hard tasks and taught explicit phonics skills. Rose responded that stronger instructional scaffolding would result in better progress, especially for the use of meaning. These findings

are summarized in Table 4. The interaction between Rose and Charles during coaching sessions, revealed by coding for interaction function, consistently emphasized Rose’s presentation of recommended instruction rather than analysis of teaching decisions or students’ literacy behavior.

#### *Defining “Real” Reading: Kristi (coach) and Melinda (kindergarten teacher)*

Kristi and Melinda discussed issues of instructional scaffolding from opposing viewpoints. Kristi emphasized highly supportive, preplanned book introductions. Melinda focused on the need for students to “really read” texts without just copying the teacher’s language and/or a textual pattern. Prominent ideas discussed by Kristi and Melinda are summarized in Table 5. Kristi and Melinda engaged in analysis of students’ success and difficulties only in very general terms, and they did not tie this information explicitly to description of recommended instructional interaction.

#### *Problem Solving: Carol (coach) and Daniel (kindergarten teacher)*

Unlike the previous three partnerships, Carol and Daniel’s discussions about instructional scaffolding were consistently contextualized in observation of students’ strengths and difficulties. Daniel believed that students should have opportunities to make mistakes and that texts should have enough challenge so that students needed to problem solve independently. He also evaluated his own teaching decisions in relationship to these ideas:

I should have gone over, at least mention, “Oh, look, he’s wading at the pool. He probably is learning to swim.” ...And *sea*, I intentionally didn’t want to say anything because I wanted to see if they could [read the word independently].

Carol generally agreed with Daniel’s analysis but also extended the conversation with explicit suggestions:

Carol: Just think, “Okay, they’re stuck [on a difficult word]. They’re not doing anything. I’d better teach them what to do.” Then you step back and teach them, try prompting...

Daniel: And have them almost verbalize, “I need to go back, start again, check the picture.” Okay.

Carol: It could be as simple as saying, “Kenny, you’re stuck. What are you going to do?” And if he looks at you with a blank face say, “Here’s what you do.”

During all three coaching sessions, Daniel provided analysis of the level of support that had been provided to students (both at Carol’s request and on his own initiative). When Daniel presented a question or hypothesis to Carol, she often posed a relevant, analytic question back to Daniel (see illustrative example of a coded coaching session segment for Carol and Daniel, Table 6). Such questions were prompts for Daniel to reflect on what had occurred during the lesson, rather than indications that Carol felt something had gone wrong:

Daniel: There were enough known words in the text, words that they could go on, and actually did go on and just try to read through it. So I think that's probably, at least to me it's fine. And they were probably where they need to be.

Carol: And yet, you, did they still have enough reading work that they had to do?

Daniel: I think there was some work with that, yes. I think yesterday Tyler was just kind of on a different wavelength.

Carol: Okay. Well maybe he just wasn't paying attention yesterday, was he? Which is typical behavior for him.

Daniel: Yeah. But I felt like he was engaged and he was really monitoring himself.

Table 3  
*Timeline of Prominent Ideas/Hypotheses: Kate and Sherrie*

Cycle	Coaching Sessions	Coach Interviews	Teacher Interviews
1	T: Students need one type of prompt at a time.	Texts will be too hard if not introduced well.	Students need less supportive book introduction so they will do more of their own thinking.
	C: Students need higher level of scaffolding.	Book introductions need to be preplanned and written.	Students need a mini-lesson on retelling story events in sequence in each lesson.
	T: Some words cannot be solved by meaning.	Prompting should emphasize use of meaning and structure cues.	
	T: Should choose text that has a good sense of story.		
2	T: Should prompt not to point with fingers to improve fluency.	Well-planned book introductions create a strong meaning base for students.	Students need a mini-lesson on retelling story events in sequence in each lesson.
	C: Students needed more scaffolding for the word <i>inside</i> .	High levels of support through book introductions and prompting will result in strong student progress.	Discussing related memories and letting students start to read is enough support.
	C: Good fluency makes word identification easier.		
	T: Should teach something that all students need to learn.	Teachers need to use a variety of types of prompts.	Students not ready for harder books because they don't use visual cues.
3	T: Omitted support to see if students could get on own.	Teachers have to be able to choose a teaching point based on students' needs.	Students need to retell collaboratively.
	C: Teaching strategies causes students to gain independence.	Students need more opportunities to apply new learning.	Vocabulary needs to be introduced conversationally rather than "predict and locate."
	C: Strategies need to be taught in a sequence.		
	C: Teaching strategies will result in strong progress than memorizing words.	Prompting only to initial visual information will not be enough support for students.	Some words not solvable by meaning/pictures and need lots of teacher prompting.
	T: Students need a way to remember specific words.		
	C: Book introductions allow students to hear difficult words in context.		

Daniel's analyses of the literacy behavior of students, often in response to a question posed by his coach, were detailed:

Carol: What do you see, not only Mary but other kids, doing as you were working with them? What are their strengths?

Daniel: Well, I feel like all four of them were really doing some good cross checking. They were looking for chunks on the run in the words, and they were confirming their guesses by looking at the pictures.

Carol and Daniel focused explicitly on the need to plan and modify instruction based on observation of students'

Table 4  
*Timeline of Prominent Ideas/Hypotheses: Rose and Charles*

Cycle	Coaching Sessions	Coach Interviews	Teacher Interviews
1	T: If students successful then level of scaffolding too high.	Effective scaffolding requires understanding of purposes across guided reading procedures.  Emergent readers do not yet need to be taught long and short vowel patterns.  Emergent readers need to be taught to focus on reading as gaining meaning.	Teaching without reference to vowel sounds makes it harder for students to read.  Important not to address too many new understandings.  Introductions should provide slight support so students can figure out words on their own.
	C: Texts without a clear story are difficult to teach.		
	C: Strategic behaviors for word solving need to be taught in order.		
	C: Important to choose texts with familiar experiences for students.		
	C: Book selection/introductions avoid need for extended sounding out.		
	C: Word identification strategies matched to students' current development cause stronger progress.		
2	T: Prompting needs to match to students' knowledge level.	Effective scaffolding requires understanding of purposes across guided reading procedures.  Most important for teachers to learn how to use meaning-based prompts for emergent readers.	Should teach to meaning first, and then phonics.  Knowledge of letter sounds causes stronger student independence.  Students need to be given hard tasks and not "babied."  In-the-moment teacher decisions better than planned.
	T: If students successful then level of scaffolding too high.		
	C: Texts without a clear story are difficult to teach.		
	T: Students should know letter sounds for word identification.		
	C: Prompting should focus on students' use of meaning.		
3	T: Omitted support to see if students could get on own.	Level of support in introductions need to be matched to most students' needs.  Word work should teach strategies students can use in today's lesson.	Minimum scaffolding should be provided so students can work independently.  Consistent use of prompts is helping students.
	C: Strategies need to be taught in sequence.		
	C: Attempting to remember words from introduction not effective.		
	T: Different strategies for words not solvable by pictures.		

responses. This focus supported their use of authentic, co-constructed discussions within which coach and teacher worked together to discuss specific instructional moves that might solve students' weaknesses. In the first interview, for example, Daniel expressed concern over kindergarten students' ability to benefit from guided reading lessons:

Developmentally, is that expecting too much, for them to be able to [cross check and integrate cues] and think about the sound? But yet it's something in the process. So I guess if you wait until, I don't know what tells you they're ready other than try it and see how they're doing and either stick with it or chuck it and go on.

By the third interview, Daniel articulated the need to modify instruction on the run based on his observation of students' needs rather than general developmental level:

They were stuck on [a word] and it was, "Okay. We'll work through that." And then we got the mag-

netic letters out and got the boards out, and did some work with that. So it made a real good connection. So had I [taught the word work component] before the lesson I wouldn't have known that was the place where they were going to need the extra help.

*Summary.* Carol and Daniel's discussions were grounded in observation of students' responses during lessons and in analysis of the instructional scaffolding that had been provided. Although Daniel was concerned that students have opportunities to read independently without scaffolding, Carol's interaction with him during coaching sessions supported his analysis and appeared to lead Daniel to useful reflection concerning the characteristics of the scaffolding needed by his students. These findings are summarized in Table 7.

### *Enacting Instructional Scaffolding*

Distinct differences were identified in the degree to which each coach/teacher partnership enacted instructional

Table 5  
*Timeline of Prominent Ideas/Hypotheses: Kristi and Melinda*

Cycle	Coaching Sessions	Coach Interviews	Teacher Interviews
1	C: Reading the book together was too much support.  C: If students don't encounter difficulty reading then the book introduction too supportive.	Effective scaffolding requires use of written plans.	Important to consider students' prior knowledge for book selection.  Need to choose books for specific teaching purposes.  Students need to be able to read sign words easily.  Student-to-student modeling for text reading is helpful.
2	T: Chose the book because had a word students need to learn.  C: Introduction needs to insure students know the gist of the text's plot.	Effective scaffolding requires use of written plans.  Effective scaffolding requires focus on specific teaching points.	Difficult balance between giving too much and not enough for introductions.  Students need to be able to read sight words easily.  Students' lack of life experiences interferes with progress.
3	C: Introduction needs to be preplanned and written out.  C: Introduction needs to provide explicit information on text's plot and language pattern.	Effective scaffolding requires use of written plans.	Difficult balance between giving too much and not enough for introductions.  Introduction shouldn't cause students to just mimic teacher's language.  Students need to be able to read sight words easily.

scaffolding during coaching conversations (see overall summary, Table 8). Although all four coaches did request teachers' analyses of the instructional support that had been provided during the observed lesson, three out of four coaches made such requests only for problems or concerns they had identified during their observation of the lesson and without explicit enactment of pedagogical reasoning. Carol, in con-

trast, supported Daniel's analysis of his students' responses to instruction, frequently responding to Daniel's questions with calls for further analysis. Carol and Daniel were also the only partnership in this study that consistently demonstrated frequent and detailed analyses of student literacy behavior and connected this information explicitly to their discussions of needed instruction.

Table 6  
Coded coaching session excerpt: Carol and Daniel

Participant	Transcript	Coding
Carol:	You were wondering about these children being able to begin to use beginning visual cues when they were stuck. Did you see any evidence of that during this lesson?	Coach providing replaying Previous coaching  Coach requesting analysis Student literacy behavior
Daniel:	Bits and pieces. But I'm not sure they put it together.	Teacher providing analysis Student literacy behavior
Carol:	What did you see them just beginning to do?	Coach requesting analysis Student literacy behavior
Daniel:	Well, I saw, I think I saw each one of them, when they got to the word, I mean it was very obvious that they did check the pictures, so that's good. They're using some cross checking there. If, like in the case of [the word] <i>resting</i> , some of them said, kind of looked back and forth and said, "Sleeping" and it was like, "I can live with that", and they closed the book. So it made sense and it was one word. Now <i>laying on the floor</i> , I saw Jamie do that, or "laying down" I think he said. And he stopped and he noticed that it didn't fit, but he didn't know what to do about it. He was aware of the error but he had, he had no direction.	Teacher providing analysis Student literacy behavior
Carol:	So what would you do tomorrow with this group, say when you see them maybe knowing that they've made a mistake and not knowing what to do, stopping here?	Coach requesting analysis Needed instruction
Daniel:	I think I'll probably try to reiterate, "Okay, you know this, you know this but that didn't fit, that didn't match. Your finger ran out of words. What can you do about that?" And try to get them to say, "Well, I'll stop, go back, look at the picture, get your mouth ready." But I have to do more teaching with that. They didn't, they didn't get it.	Teacher providing description Needed instruction  Teacher providing analysis Needed instruction

Most contradictory ideas about instructional scaffolding remained largely unresolved during the study. Daniel's statements about instructional scaffolding, on the other hand, shifted from an early general concern that guided reading

instruction might place too many demands on young learners to modifications to his instruction based on observation and analysis of his students' emerging literacy behavior. Daniel appeared to benefit from his coach's use of instructional scaf-

Table 7  
*Timeline of Prominent Ideas/Hypotheses: Carol and Daniel*

Cycle	Coaching Sessions	Coach Interviews	Teacher Interviews
1	T: Students need to utilize visual cues integrated with meaning.	Students need to read on own.	Need to choose texts with just enough work for students to do.
	C: Students need to read independently in guided reading lessons.	Important to observe students' strengths and then modify teaching decisions.	Students should have opportunity to make mistakes.
	C: Book should have opportunities for student errors.	Teaching decisions need to be clearly focused on goals.	Students can teach each other.
	T: Books should provide opportunities to cross check meaning and initial letters.	Need to teach skills that students can apply immediately.	Students need to hear prompts repetitively and consistently.
2	T: Left some word identification work to see if could do it.	Prior knowledge includes concepts and information, not just knowledge of words.	Decide on goals by trying it and observing results.
	C: Need to support application of students' prior knowledge to new text.	Word work should be generative.	Student-to-student support may interfere with progress.
	T: Students should not be just appealing to/waiting for teacher.	Teaching should help students solve problems independently.	Students need to use strategies not just memorize words.
	C: Need to teach explicitly to what action students should take at difficulty.		Need to leave some problem solving for independent work.
3	T: Student-to-student support may interfere with achievement.	Planned decisions should be superseded by students' immediate needs.	Need to choose texts with familiar experiences and explain concepts and language.
	T: Need to stop pointing, read familiar text to improve fluency.	Teaching should help students solve problems independently.	Students should have opportunity to make mistakes.
	T: Some errors should be ignored by teacher.		Prompting should require students to take action to solve.
	T: Texts with conversation help improve texts.		Students can teach each other.
	T: Can be more effective to let students correct errors on own.		Teaching choices should be based on evidence of immediate need.

Table 8

*Summary of Analysis: Enacting Instructional Scaffolding*

Analytic Criteria	Kate and Sherrie	Rose and Charles	Kristi and Melinda	Carol and Daniel
Did the coach ask the teacher to analyze instructional support provided in the lesson?	Yes (generally only for specific concerns)	Yes (disagreed with teacher's assessment or did not wait for a response)	Yes (for errors in instructional procedures, as perceived by the coach)	Yes (analysis also provided on teacher's own initiative)
How extensive was the coach's analysis of students' literacy behavior?	Limited	Limited	Limited	Frequent and detailed
How extensive was the teacher's analysis of students' literacy behavior?	Limited	Limited	Limited	Frequent and detailed
Coach's description of needed instruction tied explicitly to evidence of student behavior?	Rarely	Limited	No	Yes
Teacher's description of needed instruction tied explicitly to evidence of student behavior?	No	No	No	Yes

folding demonstrating how to engage in explicit pedagogical reasoning, resulting in his more complex and functional understanding of the nature of instructional scaffolding for reading lessons.

A self-assessment rubric for coaches' use was formulated, based on the overall success of the coaching conversations engaged in by Carol and Daniel as well as differences in the ways that all four coaching partnerships enacted instructional scaffolding within this study (see Appendix). This rubric is designed to help coaches evaluate and improve their own coaching conversations, and may also serve as a basis for professional development for coaches.

## Discussion

In this study, coaching conversations and interview transcripts were analyzed to assess the ways in which coaches and teachers conceptualize instructional scaffolding for guided reading lessons and enact instructional scaffolding during coaching conversations. Results indicate that the enactment of instructional scaffolding within coaching conversations may be an important addition to expectations for effective coaching. Teachers' expertise for effective instructional scaffolding appeared to be assisted by coaching conversations that enacted instructional scaffolding, demonstrating an analytic, evidence-based approach to instructional problem solving. Such conversations appeared to result in integration across differences in teacher and coach viewpoints on aspects of

instructional scaffolding for guided reading lessons. Daniel's understanding of the characteristics of supportive teaching may have occurred because the coach and teacher engaged in pedagogical reasoning during coaching conversations.

Reflective experience with instructional scaffolding in action could have resolved the conflicts encountered by coaches and teachers in this study, beyond the too-simple concept "to help or not to help." Each coach's emphasis on supportive scaffolding was important and consequential for student achievement. Each teacher's concern for students to read more independently was also accurate and important. These ideas are not inherently contradictory; a teacher teaches for students' ability to read independently *by* providing supportive book introductions. This perspective requires teachers to help a child extend his reach, rather than to simply test for independence by routine withdrawal of needed support. Well-designed and implemented book introductions, then, should result in increased independence through students' internalization of effective ways of interacting with text.

Three of the coaches in this study appeared to rely primarily on descriptions of needed instruction originating from their own evaluation of the observed teaching and based on prescriptions for practice. Enacting instructional scaffolding as a functional aspect of coaching conversations, however, requires coaches to model and prompt, constructing a discussion centered around analysis of students' responses during instruction and problem solving for needed instructional improvement based explicitly on evidence of student needs.

---

This shift in the nature of coaching conversations requires coaches to maintain their stance as an expert while structuring advice and input for teachers within pedagogical reasoning. Decisions regarding needed instructional changes would be based on the coach and teacher's co-constructed evidence of student responses to instruction rather than arising from the coach's admonitions or evaluative statements. It is possible that coaching conversations that enact the processes of instructional scaffolding will help teachers internalize and act on these same processes when planning and delivering instruction.

Professional development opportunities for coaches can provide explanation and demonstration of the processes of instructional scaffolding. Useful activities might include:

- Demonstration and in-depth explanation of coaching sessions that enact instructional scaffolding effectively, beyond a standard format consisting of pre-conference, observation, feedback, and a written plan of action.
- Discussion and evaluation of videotaped coaching conversations, providing coaches with the practice needed to notice key aspects of effective coaching.
- Coaches' self-assessment of their own enactment of instructional scaffolding during coaching conversations, using the rubric developed from this study (see Appendix).

The results of the study contribute crucial information regarding the interaction between coaches and teachers in support of improved reading instruction. Limitations inherent to the study are likely to have affected the results, indicating areas in need of further research. A small sample size of four coach/teacher partnerships was utilized for this study. This sample size is appropriate for descriptive research and allowed for detailed data collection and analysis, but limits generalizability. Some aspects of the study's findings may vary by amount of teaching experience, for example, or differences in teachers' professional development. A teacher who has studied instructional scaffolding in clinical settings, for example, would be likely to conceptualize and discuss this complex concept with more depth and clarity. As in all qualitative studies, the design of this study did not allow for all variables to be controlled. The study did not analyze for teachers' ability to actually implement effective instructional scaffolding within guided reading lessons, nor of the effectiveness of guided reading instruction in general. It is likely that the coaches who volunteered to participate in this study were different in key ways than their colleagues. They may, for example, have felt more confident in their coaching expertise than did other reading coaches in the district.

Further research is needed in order to determine whether the findings of this study are typical of effective coaching. It will be important, for example, to verify whether coaching conversations are more effective when they include enactment of instructional scaffolding for other types of lessons beyond guided reading and with teachers possessing varying amounts and types of training and experience. Investigations

identifying additional aspects of coaching conversations associated with differential success in teachers' instructional expertise will also be crucial. It would be useful to study any differences in student achievement data that result from a variety of types of coach/teacher interaction, and further research should also investigate the effectiveness of the coaching rubric utilized in the current study. This research, coordinated with the present study's findings, will fill important gaps in our current knowledge of the characteristics of effective interaction between coaches and teachers. Without the availability and use of this knowledge base for the selection and training of coaches, consistent improvement in instruction and increased student achievement may continue to be an elusive expectation.

## References

- Altwerger, B., Arya, P., Jin, L., Jordan, N. L., Laster, B., Martens, P. ... Wiltz, N. (2004). When research and mandates collide: The challenges and dilemmas of teacher education in the era of NCLB. *English Education, 36*(2), 119-133.
- Bean, R. M. (2004). *The reading specialist. Leadership for the classroom, school, and community*. New York: Guilford Press.
- Darling-Hammond, L. (1999). Educating teachers for the next century: Rethinking practice and policy. In G. A. Griffin (Ed.), *The education of teachers. Ninety-eighth yearbook of the National Society for the Study of Education. Part I* (pp. 221-256). Chicago: The University of Chicago Press.
- Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading. Good first teaching for all children*. Portsmouth, NH: Heinemann.
- Fragar, A. M. (1994). Developing case knowledge in reading education. *The Journal of Reading, 37*(5), 392-398.
- Gibson, S. A. (2006). Lesson observation and feedback: The practice of an expert reading coach. *Reading Research and Instruction, 45*(4), 295-318.
- Gibson, S. A. (2010). Reading Recovery teacher expertise: Gaining and structuring content knowledge for early literacy intervention. *Literacy Teaching and Learning, 15*(1 & 2), 15-71.
- Keith, M. J. (1988, November). *Stimulated recall and teachers' thought processes: A critical review of the methodology and an alternative perspective*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Louisville, KY.
- Literacy Coaching Clearinghouse. (2009). *Self assessment for elementary literacy coaches*. Retrieved from <http://www.literacycoachingonline.org/tools.html>
- Meyer, D. K., & Turner, J. C. (2002). Using instructional discourse analysis to study the scaffolding of student self-regulation. *Educational Psychologist, 37*(1), 17-25.
- Many, J. E., Dewberry, D., Taylor, D. L., & Coady, K. (2009). Profiles of three preservice esol teachers' development

- 
- of instructional scaffolding. *Reading Psychology*, 30(2), 148-174.
- Neufeld, B., & Roper, D. (2003). *Coaching: A strategy for developing instructional capacity: Promises and practicalities*. Cambridge, MA: Education Matters.
- Palincsar, A. S. (1986). The role of dialogue in providing scaffolded instruction. *Educational Psychologist*, 21(1 & 2), 73-98.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park, CA: Sage.
- Peterson, D. S., Taylor, B. M., Burnham, B., & Schock, R. (2009). Reflective coaching conversations: A missing piece. *The Reading Teacher*, 62(6), 500-509.
- Poglinco, S. M., Bach, A. J., Hovde, K., Roseblum, S., Saunders, M., & Supovitz, J. A. (2003). *The heart of the matter: The coaching model in America's Choice schools*. Consortium for Policy Research in Education. University of Pennsylvania, Graduate School of Education.
- Rodgers, A. (2004). Some assembly required. Scaffolding in the classroom. In A. Rodgers & E. M. Rodgers (Eds.), *Scaffolding literacy instruction. Strategies for K-4 classroom* (pp. 162 -172). Portsmouth, NH: Heinemann.
- Rodgers, A., & Rodgers, E. M. (2007). *The effective literacy coach*. New York: Teachers College Press.
- Roehler, L. R., & Cantlon, D. J. (1997). Scaffolding: A powerful tool in social constructivist classrooms. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning* (pp. 6-42). Cambridge, MA: Brookline Books.
- Ross, P., & Gibson, S. A. (2010). Exploring a conceptual framework for expert noticing during literacy instruction. *Literacy Research and Instruction*, 49(2), 175-193.
- Ruddell, R. R. (2004). Researching the influential literacy teacher: Characteristics, beliefs, strategies, and new research directions. In R. Ruddell & N. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 979-997). Newark, DE: International Reading Association.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research. Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.
- Strong, M., & Baron, W. (2004). An analysis of mentoring conversations with beginning teachers: Suggestions and responses. *Teaching and Teacher Education*, 20, 47-57.
- Taylor, B. M., Pearson, P. D., Clark, K., & Walpole, S. (2002). Effective schools and accomplished teachers: Lessons from primary-grade reading instruction in low-income schools. *Elementary School Journal*, 101, 121-165.
- Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life. Teaching, learning, and schooling in social context*. Cambridge, MA: Cambridge University Press.
- Wood, D. (1998). *How children think and learn. The social contexts of cognitive development*. Malden, MA: Blackwell.
- Wood, D. (2003). The why? what? when? and how? of tutoring: The development of helping and tutoring skills in children. *Literacy Teaching and Learning*, 7(1 & 2), 1-30.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17, 89-100.

Appendix  
Coaching Conversation Assessment Rubric

	1 Ineffective	2 Improving	3 Expert
Description of literacy behavior of students during an observed guided reading lesson, provided by both coach and teacher and integrated into coaching conversation.	Focused primarily on students' affective behavior/attention during the lesson. Limited description of children's literacy behavior.	Coach describes students' literacy behavior, addresses children's use of strategies. Not well integrated with other aspects of the conversation.	Both coach and teacher describe students' use of strategies, and utilize to analyze teaching and determine suggestions for improvement.
Analysis of new literacy behavior/strategies needed by students.	Neither coach nor teacher extends description of student literacy behavior to next steps needed by individual students.	Coach describes next steps in development of strategic behavior. Description explicitly tied to students' literacy behavior.	Coach and teacher co-construct analysis of strategies needed by individual students, integrated into coaching conversation.
Replaying portions of instruction, integrated with analysis of students' literacy behavior and needs.	Little if any description of teaching moves/decisions, generally provided by coach and at points perceived as errors.	Coach describes teaching moves/decisions, primarily areas identified as concerns, and establishes connection to student needs.	Coach and teacher co-construct understanding of teaching moves/decisions, analyzing and describing specific aspects of students' literacy behavior.
Analysis of the type/level of instructional support provided, connected to students' literacy behavior.	Little or no analysis of instruction; emphasizes routine procedures rather than instructional support.	Coach provides evaluation of aspects of teaching; some connection to students' literacy behavior.	Coach and teacher co-construct analysis of teaching decisions, well connected to students' literacy behavior.
Description of specific changes in instructional interaction needed, integrated with students' literacy behavior/needs and analysis of instructional support provided.	Few if any specific suggestions or decisions regarding changes in instruction. Suggestions related to instructional routines rather than students' literacy behavior/needs or analysis of instructional support.	Coach provides specific suggestions for teacher's improved interaction with students, related to observed student literacy behavior. Needed instruction typically described by coach.	Coach and teacher co-construct specific suggestions for improved interaction with students, integrated with observation of student literacy behavior/needs and analysis of instructional support.

---

# *Mid-Western Educational Research Association*

## **34<sup>th</sup> Annual Meeting Call for Proposals**

**Proposal Deadline: *May 1, 2011***

### ***“What does it mean to be educated?”***

**October 12-15, 2011**

**Sheraton Westport Plaza Tower Hotel, St Louis, Missouri**

Ellen A. Sigler, Ed.D. Program Chair

[mwera2011@wcu.edu](mailto:mwera2011@wcu.edu)

The 2011 Annual Meeting of the Mid-Western Educational Research Association (MWERA) will be held in the greater St. Louis area, with an exciting program of invited speakers, workshops, and peer-reviewed papers presented in a variety of session formats.

The 2011 program will center upon this year’s theme: **“What does it mean to be educated?”** The gap between research and practice can only be bridged if we define what it means to be educated in a rich, useful way. What are the different philosophies and definitions of “educated” that inform our practice? At all levels, from preschool programs to doctoral programs, the program theme encourages us to explore our goals as educators, the purpose of our schools, and the expectations we have of those we teach. Attendees and presenters are invited to explore questions beneath the surface of the work done by both researchers and teachers.

The conference will feature dynamic speakers of interest to both researchers and practitioners. Teachers, administrators, and other school personnel are especially invited to come and share their school-based research and experiences at the 2011 MWERA conference.

We will be meeting at the **Sheraton Westport Plaza Tower Hotel**. The hotel is centrally located in St. Louis’s popular West Port Plaza area, placing us just minutes from some of the city’s most popular attractions. West Port Plaza offers access to more than 30 restaurants, exclusive shops, and entertainment. For more information go to <http://www.sheratonwestport.com/> and click on **Sheraton Westport Plaza Tower**.



---

## Ways to Participate

Anyone may **submit a proposal** for the MWERA 2011 Annual Meeting, whether or not that person is currently a member of MWERA. However, before the Annual Meeting, all presenters **MUST** be members in good standing with MWERA. To promote broader participation in the program, no one person should appear as a presenter on more than three proposals.

Division Chairs are also seeking MWERA members to serve as **Proposal Reviewers, Session Chairs, and Session Discussants**. Please contact the Program Chair if you are willing to serve. Finally, you can participate by attending the conference and encouraging colleagues and students to participate.

## Guidelines for Submitting a Proposal

Although it is desirable for proposals to address the theme of the Annual Meeting, it is not required. Proposals **MUST be submitted electronically**, using the submission process available through the MWERA website. Specific instructions for submission can be found at [www.mwera.org](http://www.mwera.org).

## Deadline for Proposal Submission

The proposal portal will be **open March 15, 2011** and all proposals must be submitted no later than **May 1, 2011**.

## Criteria for Proposal Review

Depending on the format and type of scholarly work being presented, appropriate criteria have been developed and will be used for the blind review process. These criteria include:

- a. Topic (originality, choice of problem, importance);
- b. relevance of the topic to the Division and to MWERA membership;
- c. contribution to research and education;
- d. theoretical/conceptual framework;
- e. analyses and interpretations (significance, implications, relationship of conclusions to findings, generalizability, or trustworthiness); and
- f. overall written quality (clarity of writing, logic, and organization).

Papers presented at MWERA are expected to be original work that has not previously been presented at any other meeting or published in any journal. Further, it is a violation of MWERA policy to promote commercially available products or services (except as exhibits) that go beyond the limits of appropriate scholarly or scientific communication. Individuals who wish to display educationally-related products or services should contact the Program Chair.

## Content Required for Proposals

NOTE: These are blind reviews. No identifying information should be included in the proposal or on the cover page.

### Abstract

The abstract should be 100-150 words. The abstracts of accepted papers will be published in the *MWERA 2011 Annual Meeting Program* and may be available on the MWERA website. Use clear, precise language, which can be understood by readers outside your discipline.

### Paper or Poster Session

Summaries for **Paper** and **Poster** proposals should be **approximately 2500 words, or about 4-6 pages in length** and explicitly address as many of the following as appropriate, preferably in this order: (a) objectives, goals, or purposes; (b) perspectives and/or theoretical framework; (c) methods and/or techniques (data source, instruments, procedures); (d) results and conclusions; and (e) educational and/or scientific importance of the work.

### Symposium, Workshop, Alternative Session, and Best Practices Forum

Summaries for **Symposium, Workshop, Alternative Session, and Best Practices Forum** proposals also should be **approximately 2500 words, or about 4-6 pages in length** and explicitly address as many of the following as appropriate, preferably in this order: (a) descriptive title; (b) objectives, goals, and purposes; (c) importance of the topic, issue, or problem; (d) explanation of the basic format or structure of the session, with a brief rationale for the format; (e) listing of the presenter(s), by number not name for blind review (e.g., "Presenter 1"), with an explanation of each person's relevant background and role in the session; and (f) anticipated audience and kind of audience involvement. Limited program space may be available for these types of sessions. A specific session time period may be requested; however, session length is dependent on room availability and is under the discretion of the program chair.

## Session Descriptors for Proposals

Please be certain to use the approved MWERA descriptors in completing your proposal. These descriptors are located on the "Annual Meeting Information" tab of the MWERA website [www.mwera.org/information.html](http://www.mwera.org/information.html) and as part of the submission process.

## Session Format Descriptors

### Paper Presentation

Paper sessions are intended to allow presenters the opportunity to make short, relatively formal presentations in which they summarize their papers to an audience. Three to

---

five individual papers dealing with related topics are grouped into a single session running 1 hour 20 minutes. Each paper presentation is allowed approximately 15 minutes (depending on the number of presentations in a given session) to present the highlights of the paper. In addition, a Session Discussant is allowed approximately 10-15 minutes, following all papers, for comments, synthesis, and/or constructive feedback. A Session Chair moderates the entire session. Presenters are expected to provide complete copies of their papers to all interested audience members.

### **Poster**

Poster sessions are intended to provide opportunities for interested individuals to participate in a dialogue with both the presenter(s) and other interested individuals. Presenters are provided an area in which to display a small, table-top Poster, ancillary handouts, or other table-top A/V materials. Interested individuals are free to move into and out of these poster presentations as they wish. Presenters are expected to make available complete copies of the paper on which the poster was focused.

### **Symposium**

A symposium is intended to provide an opportunity for examination of specific problems or topics from a variety of perspectives. Symposium organizers are expected to identify the topic or issue, identify and ensure the participation of individual speakers who will participate in the session, prepare any necessary materials for the symposium, and facilitate the session. It is suggested, though not required, that the speakers or symposium organizer will provide interested individuals with one (or more) papers relevant to, reflective of, or drawn from the symposium.

### **Workshop**

Workshops are intended to provide an extended period of time during which the workshop leader helps participants develop or improve their ability to perform some process (e.g., how to provide clinical supervision, using the latest features of the Internet, or conduct an advanced statistical analysis). Workshops may be up to 1.5 hours and organizers are responsible for providing all necessary materials for participants.

### **Alternative Session**

The form, topics, and format of alternative sessions are limited only by the imagination and creativity of the organizer. These options are intended to afford the most effective method or approach to disseminating scholarly work of a variety of types. Proposals for alternative sessions must include a brief rationale for the alternative being proposed and will be evaluated on their appropriateness to the topic and audience, their ability to meet the limitations of time, space, and expense for MWERA, and the basic quality or value of the topic. The organizer of alternative sessions is responsible for

all major participants or speakers, developing and providing any necessary materials, and chairing the session.

### **Best Practices Forum**

The “Best Practices” sessions provide opportunities for individuals or groups to present “best” or “promising” practices impacting both K-12 and higher education. These sessions highlight unique and innovative programs that have demonstrated promise for improving and enhancing educational practice. Presenters will be grouped by similar topics to facilitate discussion among the groups and audience. Presenters are expected to make available complete copies of the paper on which the “Best Practices” session focused.

## **Expectations of Presenters**

Presenters whose papers have been accepted to a session with a Session Chair and/or Session Discussant are **expected to upload** a completed version of their conference paper through the MWERA website **no later than September 15, 2011**.

MWERA reserves the right to reproduce and distribute summaries and abstracts of all accepted proposals, including making such works available in a printed or electronic form, through the MWERA website, and in various promotional materials, and in press releases promoting the Annual Meeting and the organization. As a condition of acceptance, all authors of papers accepted to the 2011 Annual Meeting explicitly grant MWERA the right to reproduce their work’s summary and/or abstract in these ways. Such limited distribution does not preclude any subsequent publication of the work by the author(s).

Authors of accepted proposals assume the ethical and professional responsibility to appear at the Annual Meeting and to participate in their presentation or assigned session. When circumstances preclude the author(s) from doing so, it is the responsibility of the author(s) to arrange a suitable substitute to present the accepted material and to notify the Program Chair in advance, or as soon as possible.

---

## **Dates to Remember**

March 15, 2011	Proposal portal opens
May 1, 2011	Proposal Submission Deadline
July, 2011	Notification of acceptance to first listed presenter/proposal submitter
September 11, 2011	Last day for conference hotel guaranteed room-rate
September 16, 2011	Accepted Papers uploaded for discussant review
October 12-15, 2011	Conference dates for MWERA in St. Louis

---

---

## Questions

Questions about proposals, the electronic submission process, or the meeting in general should be directed to the Program Chair:

**Ellen A. Sigler, Ed.D.**

**Department of Psychology**

**Western Carolina University**

[mwera2011@wcu.edu](mailto:mwera2011@wcu.edu)

## Division Information

### **A—Administration, Organization, & Leadership**

This division is concerned with research, theory, development, and the improvement of practice in the organization and administration of education.

Larry Cook [lcook@ashland.edu](mailto:lcook@ashland.edu) Ashland University

### **B—Curriculum Studies**

This division is concerned with curriculum and instructional practice, theory, and research.

Jill Phipps [phipps@uakron.edu](mailto:phipps@uakron.edu) The University of Akron

### **C—Learning & Instruction**

This division is concerned with theory and research on human abilities, learning styles, individual differences, problem solving, and other cognitive factors.

Larry Tinnermen [Larry.Tinnerman@indstate.edu](mailto:Larry.Tinnerman@indstate.edu) Indiana State University

### **D—Measurement & Research Methodology**

This division is concerned with measurement, statistical methods, as well as both quantitative and qualitative research methods, as applied to educational research.

Dustin Derby [dustin.derby@palmer.edu](mailto:dustin.derby@palmer.edu) Palmer College of Chiropractic

### **E—Counseling & Human Development**

This division is concerned with the understanding of human development, special education, and the application and improvement of counseling theories, techniques, and training strategies.

Tracy Stuckey-Mickell [tmickell@ehe.osu.edu](mailto:tmickell@ehe.osu.edu) The Ohio State University

### **F—History, Historiography (to include Philosophy)**

This division is concerned with the findings and methodologies of historical research in education. In addition, this division also welcomes papers concerning educational thought and philosophies applied to educational practice TBD.

Ellen Sigler [esigler@wcu.edu](mailto:esigler@wcu.edu) Western Carolina University.

### **G—Social Context of Education**

This division is concerned with theory, practice, and research on social, moral, affective, and motivational characteristics and development, especially multicultural perspectives.

Ting Li [lit@muohio.edu](mailto:lit@muohio.edu) Miami University

### **H—Research, Evaluation, & Assessment in Schools**

This division is concerned with research and evaluation to improve school practice, including program planning and implementation.

Caroline Prior [capryor@siue.edu](mailto:capryor@siue.edu) Southern Illinois University

### **I—Education in the Professions (to include the Scholarship of Teaching and Learning - SOTL)**

This division is concerned with educational practice, research, and evaluation in the professions (e.g., medicine, nursing, public health, business, law, and engineering). In addition, this division also welcomes papers on the scholarship of teaching and learning, which involves the scholarly inquiry into student learning which advances the practice of teaching in post secondary programs.

Anne George [george@sxu.edu](mailto:george@sxu.edu) Saint Xavier University

### **J—Postsecondary Education**

This division is concerned with a broad range of issues related to two-year, four-year, and graduate education.

Eric Mansfield [EA-mansfield@wiu.edu](mailto:EA-mansfield@wiu.edu) Western Illinois University

### **K—Teaching & Teacher Education**

This division is concerned with theory, practice, and research related to teaching at all levels and in-service and pre-service teacher education, including field experience supervision and mentoring.

Tasha Almond [tashaA@mtplains.com](mailto:tashaA@mtplains.com) Mountain Plains Evaluation

### **L – Educational Policy & Politics**

This division is concerned with educational policy as well as political, legal, and fiscal matters related to education.

Lynda Leavitt [lleavitt@lindenwood.edu](mailto:lleavitt@lindenwood.edu) Lindenwood University

**For membership information <http://www.mwera.org>**

## *Politics, Economics, and Testing: Some Reflections*

Michael J. Feuer  
George Washington University

Good afternoon, and thank you for such a warm welcome and kind introduction. It is a pleasure to be here, and I've been looking forward to visiting with you—and seeing Columbus—ever since Cindy proposed this many months ago.

My topic today is assessment and accountability, surely not new words or concepts to anyone here, although perhaps my arguments will provoke some new thinking. Let me start with some data that probably will sound familiar, if not in its detail then its gloomy underlying message. This relates to some “recent” test results:

Out of 57,873 possible answers, students answered only 17,216 correctly and accumulated 35, 947 errors in punctuation in the process. Bloopers abounded: one child said that rivers in North Carolina and Tennessee run in opposite directions because of the will of God. (U.S. Congress, Office of Technology Assessment, 1992, p. 109)

If you're wondering how you missed this important news item, don't worry: it's not from the *most* recent NAEP or SAT or Ohio Achievement Assessment. Rather it's from one of the first instances of large scale written educational testing, *circa 1840*, a time of great reform in American schooling, a period that later became known as the Common School reform movement and was associated with Horace Mann and others who spent their lives trying to broaden the franchise of educational opportunity and raise standards, all at the same time. [Let me digress just for a moment here to cite the great historian of education, Lawrence Cremin, who noted in a brilliant and short book he wrote just before his untimely death, the idea that we could (and should) raise quality standards and increase access simultaneously was a uniquely American ideal and one that we are still, in many ways, pursuing. Cremin's (1990) book should be required reading for anyone contemplating venturing into the turbulent world of education reform...].

But back to my story line...which is about the importance of history in considering contemporary educational challenges. We sometimes forget that some of our most vexing problems are, in their fundamental aspects, not new. We are a relatively young country (to paraphrase from Tom Lehrer's memorable line about Mozart, by the time the US National Academy of Sciences was founded in 1863, the first King of England had already been dead for about 1000 years...). But we do have history here, and our history of education is marinated in flavorful juices of the great American experiment with divided government, with a certain excep-

tionism that steered us away from other systems that had seduced so many other societies. Our allergy to centralized authority, coupled with a deeply held aspiration for fairness, are two elements in our unique political culture that have had and continue to have great effect on education policy, reform, and learning.

Part of my message today is a simple one but I hope not simplistic: Our penchant for accountability and our appetite for standardized testing are, in the language of statistics and psychometrics, highly collinear.

But first I want to address a specific aspect of the history, as it relates to testing and accountability. It is sometimes tempting to demonize the testing community for all kinds of perceived evils: bringing us the wondrous frustrations of multiple choice test items that seem to bear little relation to what we really value in teaching and learning, being so ready and willing to market more and more tests that can be scored at greater and greater speed, and for not being terribly concerned with the deeper meanings of test results or their behavioral consequences as long as the results meet certain standards of statistical reliability. We find it easy and convenient to blame the test makers for everything from adverse impact in higher education to the horrors of teaching-to-the-test in K-12. I've actually heard one good friend of mine, in a rather extreme fit of anger, attribute a teacher's suicide to NCLB requirements for student testing!

It's all rather easy, and somewhat enjoyable, this test bashing, and I admit at times I've tasted the Kool-Aid. But let's not forget (and as a recovering economist I cannot forget) that there is usually a *demand* side that at least partially explains why certain strange or undesirable things appear on the market. In this case, i.e., the emergence of uniform written exams, the forces that converged to enable and propel testing as perhaps the most persistent and arguably powerful tool for assessment of educational quality and governance of educational change, had its roots in fundamental aspects of the unique experiment in democracy that was taking shape in our new republic. Why should we be surprised, really, that by 1975 one of the great minds of mental measurement and educational assessment, was lamenting five decades of controversy over mental testing while noting, perhaps immodestly but certainly with scientific validity, that psychometrics had become one of the greatest contributions of psychology to human affairs (Cronbach, 1975)?

We had better recognize that this tension would not have been possible if there hadn't been, for a long time and

---

for many legitimate reasons, a powerful demand side in the production and distribution of tests, an appetite for standardization that had its roots in the coinciding principles of democratic accountability and efficiency in the expansion of educational opportunity.

In sharing the news of how poorly students performed on that 1840 assessment and of how charmingly wrong some of their answers really were I could be making a perhaps simpler point, namely that rumors of the golden age—that elusive and transitory period in history when things were fine as compared to how awful things have become—are more than a tad exaggerated. Now, it has been empirically documented frequently (most recently in an extraordinary book by two Harvard economists, Goldin & Katz, 2010) that at least until the last quarter of the 20<sup>th</sup> century our remarkable educational system was, indeed, in something of a golden age, largely responsible for advancing the general economic and social welfare of our nation and for uplifting the quality of life and standard of living to levels well above any other nation in the world. (It is important that we keep this historical record in mind as we contemplate the future, and though much of the doom and gloom rhetoric based on cross-sectional evidence from international comparative assessments is exaggerated, there is reason to fear the ill effects of complacency borne of prior success.)

But my main reason for recalling the 19<sup>th</sup> century experience with testing is to make a different historical point: it is to emphasize that standardized educational tests have been a staple of public accountability in education for almost two centuries, and that from their inception they have been popular devices used for both good and mischief. Horace Mann and his partners in the great reform movement were not only brilliant social reformers intent on expanding the educational franchise, but they were shrewd politicians too, who understood long before the ascendance of professional communications experts and policy wonks that by including certain questions on the tests they could expose the failures of school masters they were battling with, and, as one of our preeminent educational historians noted, use testing as a “bludgeon of reform...” (Tyack, 1974). In a word, if you think some teachers and principals are feeling pressured by NCLB testing, you are right: but based on the historical evidence one cannot help think that today’s test-based accountability pales in its ferocity when compared with the earliest episodes of the “bludgeoning...”

We’ve been testing for a long time. My point is that it’s not so surprising when viewed in the context of the American experiment. There is a deeply American quality to this reliance on tests: they were a remarkable invention of social engineering in large part because they did not appear to require a tradeoff between efficiency and fairness—they rather spectacularly seemed to achieve both goals at once. I would argue that standardized testing became a symbol of the aspiration for fairness and universal access that distinguished American schools from European and Asian schools.

Moreover, as the tests grew more sophisticated, both in their format and in models for scoring and interpreting of results, they increasingly were viewed as tools of rational—scientific—management. Let me elaborate just a bit on this comment and tie testing to more generic properties of technology and society. In a country and culture already beginning to exhibit a certain fascination with the possibilities of technology—which would of course characterize the extraordinary transformation of the American society and economy over the remainder of the century—here was one, standardized testing, with genuinely *dual* uses. The duality, at the time, was already comprised of measurement (i.e., describing what the kids are learning) and reform (i.e., motivating change to improve their learning). And since then that duality has blossomed from two branches into a rather more complex system with multiple purposes, multiple designs, and a highly complex interweaving of goals and constraints that makes most rational policy analysts run for something simpler (mapping the origins of the universe, for example.)<sup>1</sup>

Let me try to underscore some of the key ideas embedded in this brief historical prelude:

- Neither NCLB nor its recent antecedents (Goals 2000: remember that?) are new attempts to rely (and perhaps over-rely) on testing as a technology of reform, nor is the evidence of arguably irreconcilable multiple uses of test results;
- Tests—like most if not all technologies—are imperfect, which means that some results will overstate and other results will understate the “true” state of a child’s learning or potential;
- The fact that we continue to rely on tests is to a large extent attributable to our unflagging pursuit of at least some “objectivity” in the way we evaluate teaching and learning, which is rooted in the framing principles and philosophy of the American democratic experiment and, in particular, our aversion to centralized authority;<sup>2</sup> and
- What has been missing from the often heated debate over assessment and its multiple uses has been a kind of rational and dispassionate analytical framework for assessing its benefits and costs, perhaps similar to the analytical frames we apply to other complex phenomena in which there needs to be attention to both the good and the bad, a framework that could perhaps inform policy makers and the public about the strengths and limitations of testing and stimulate the kind of research needed to increase the benefits and reduce the costs.

---

1 A Nobel-prize winning physicist once confessed that after working on education reform for a few years he decided to go back to estimating the ages and chemical composition of the planets... which he said was much easier.

2 For a description of how American policy makers at times envy their counterparts in more centralized systems, see my discussion of French Education Minister Claude Allegre’s visit to the National Academy of Sciences, in Feuer, M. (2006), *Moderating the Debate: Rationality and the Promise of American Education*. Cambridge: Harvard Education Press.

---

This leads me to the main lesson I'd like to impart today, namely that we need to start thinking about the future of testing and accountability through a lens of potential benefits, potential costs (or risks), and perhaps most important, the pursuit of reasonable rather than optimal solutions to the problems of testing and accountability. I'm going to revert to some core principles of economics in advancing us toward such a framing of the issues.

It is a staple of economic theory that individually rational and self-interest seeking behavior can lead to disastrous or at least seriously suboptimal social outcomes. Anyone who has driven on a highway and has confronted the frustration of "rubbernecking," for example, has first hand experience with the failure of individual rationality in terms of its collective results (see for example, Schelling, 1974).

What does traffic flow have to do with testing and accountability? In a nutshell, the fact that certain behaviors, or technologies, lead to unintended or undesirable consequences, is not, in itself, a sufficient basis for banning the technology; rather, understanding the sources of what are sometimes called "externalities" in the literature of political economy, is an important foundation upon which to build appropriate policy remedies.<sup>3</sup> The lesson is that

- there have always been and continue to be justifiable arguments for accountability generally and for the use of tests as one tool of accountability;
- there are unintended negative consequences of testing for accountability that need to be anticipated as best as possible; but
- the undesired risks or costs associated with testing as a tool of accountability need to be weighed against the potential and measureable benefits as well as against the counterfactual case of eliminating testing from the toolbox of acceptable accountability practices.

So, just as we would not prohibit either traffic flows or the rights of drivers to look out their windows, we should not lurch toward a prohibition of testing just because it obviously (and not so obviously) entails downside risks and some unarguably bad behavior. A good example of public policy analysis that hinges on this approach to dealing with benefits and costs is in the environmental movement. Strategies for remedying the ill effects of individually-self interested behavior that results in water and air pollution have evolved from the naïve view that damaging the environment was in some ways analogous to crimes warranting rigid and coercive policing, to the development of more sophisticated political, legal, regulatory and incentives-based approaches. A prominent economist working in this area offered this contrast:

...the police power approach ... is appropriate when a certain kind of behavior is perceived as a terrible social threat and it is felt the behavior must

be stopped even at great cost...but there is a world of difference between hijacking [and other such crimes] and pollution. Hijacking is a threat to life and property without redeeming features, whereas pollution is a by-product of thousands of individual decisions in the course of very desirable activities—production and consumption of commodities and services. Hijacking should be prevented if possible, whereas with pollution, the goal is to induce people to continue the desirable activities in ways that reduce and alter environmental discharges.... (Mills, 1978, p. 204)

Embedded here is the notion that simply prohibiting polluting behavior is likely to be inefficient, counterproductive, and insensitive to negative consequences that could be even more damaging than the pollution itself. A range of strategies have been devised over the years, with varying success, all aimed at inducing changes in behavior and collectively reducing the pace and magnitude of a perceived and real set of externalities. Regulatory approaches, for example, can be costly to design and implement, and though still in wide use have exhibited mixed levels of success; variations on taxation schemes, which are intended to curb polluting behavior by imposing monetary charges, have become more popular, to economists at least, although such programs also can be costly to design and enforce.

All these initiatives share a basic proposition, namely that *the goal of reducing a negative externality requires attention to benefits and costs—in the estimation of the effects of the polluting technology and the effects on the economy and society of curbing the pollution, and in the estimation of the costs associated with designing and implementing the policy strategy itself.*

Perhaps this schema help us untangle the problems of testing and accountability. The main point is that test based accountability systems have benefits and costs, and I'll start with the latter. Here is an abridged list of the things that can and do go wrong when tests are used inappropriately:

1. Tests are imprecise tools of estimation that provide only a partial view of selected aspects of what students know ("domain sampling"). Using tests as a basis for more comprehensive judgments is usually inappropriate.
2. Tests alone offer preliminary clues, at best, as to how students learned whatever it is they demonstrate on the test. Inferences about teachers, schools, principals, class size, and other possible causes require substantially more data than score reports.
3. Most of the tests available "off the shelf" are not well-suited to providing teachers with useful information on the cognitive or intellectual barriers their students face, the special work they need to improve, or the ways teachers can shape their lessons to help kids overcome specific learning gaps.
4. When test results are used as a basis for making significant decisions ("high stakes" decisions) the validity of

---

<sup>3</sup> This argument is expanded in Feuer, M. "Externalities of Testing: Lessons from the Blizzard of 2010," *Measurement*, 8: 59–69, 2010.

---

the scores can be compromised.

5. As a corollary to #4, teaching to the test is usually a bad idea, no matter what the test looks like, unless of course we don't care much about the validity or reliability of the information the test was originally designed to produce.
6. Decisions based on any cut score methodology will result in misclassification, assuming tests are imperfect estimators of the underlying domain of interest. (Most people worry about false negatives, i.e., kids erroneously being identified as "below basic" when in fact they're not. Adverse impact issues arise from this type of error. But as important is the problem of false positives, i.e., kids (or schools) that are branded as passing when in fact they're not.)
7. Using tests as the sole basis for measuring adequate yearly progress can lead to huge numbers of schools being misclassified, even when the source of their failure can be quite random. The effects of such misclassification on resource allocation, student mobility, parental support for schools, and morale can be costly in ways we don't really know how to measure.
8. Too much emphasis on test results naturally leads to distortions in the way both good and bad teachers allocate their time. We simply don't know how many good teachers will (a) develop reactive strategies that undermine their otherwise good instincts, (b) find ways to game the system just so they can go on with their good teaching, or (c) give up and leave the system to only those teachers for whom the testing doesn't make much difference!
9. Tests used to compare schools across states ought to be designed with enough similarity of content and format to permit valid comparisons. Reconciling this simple dictum, distilled from the literature on test equivalence and linking, with the historical and contemporary insistence on state and grass roots control of curricula and pedagogy is a full-time job.
10. There is a risk that as tests become both more important and more similar across states and jurisdictions they will become a de facto national curriculum. Few Americans seem ready for that.

Against this impressive array of good reasons to curb our enthusiasm for testing, what can possibly be said in its defense? I will offer a few general answers and some more directly tied to our current situation.

1. The alternative to standardized assessment of student learning is a return to subjectivity and intuition, neither of which should be viewed as a curse except if they are attributes of decisions that a) would benefit from more rigor and precision and b) are the basis for actions that can seriously affect children, teachers, or the public trust. Although we've all been in classrooms where inspired teachers are doing wonderful things, relying on a "know it when you see it" criterion for evaluating teaching and learning would be insulting to the profession not

to mention hazardous to the learning opportunities of generations of children.

2. Given the complexities embedded in the words "education" and "learning," it is important to agree on at least some basic approximations and on some metrics to inform parents and others of whether anything of value is actually taking place. A culture that is capable of digesting and interpreting, with exquisite subtlety, the massive quantities of statistics that are collected about, for example, major league baseball, has clearly expressed its appetite for quantitative information about progress of education.
3. When designed and implemented properly, tests can provide useful information to teachers, principals, and school officials striving for improved policies and practices. The fact that test results are often expropriated for uses that go beyond their technological capacity and beyond the aims for which they have been validated is not, in itself, a sufficient argument against their use for the purposes for which they were designed and validated.
4. Without some agreed-upon quantitative benchmarks, the good embedded in so many of our schools and the high quality of professionalism exercised by so many of our teachers will a) be suspect and b) not become the basis for learning and improvement elsewhere, where it is needed most. It's not just that test scores provide a certain kind of braking function on public and political jockeying and the impulse to make extravagant claims of success prematurely; it goes the other way too, in terms of providing evidence of genuine progress that can be the foundation for scaling up progress beyond specific cases.
5. The inverse of point 4 should be obvious: without agreed-upon metrics the concept of public accountability is fundamentally undermined. Recall the initial use of uniform examinations at the birth of the common school reform, i.e., the idea of giving parents on "both sides of the tracks" information about how their children were faring. Identifying schools or school systems that are in trouble, using well designed tests of student achievement, should be an acceptable basis for further investigation, and most important, design of programmatic or policy remedies.

And now for a few more specific arguments relevant to our current situation:

- The bad news about NCLB notwithstanding, there are some positive results that are perhaps under-reported. For example:

NCLB's focus on students with low achievement seems to have had some short-term positive effects. The percentage of schools meeting Adequate Yearly Progress (AYP) targets increased in 2003-04 from the year before in most states, and the recently released National Assessment of Educational Progress (NAEP) long-term

---

trend scores have shown some narrowing of achievement gaps. (Linn, 2005, p. 1)

- On the issue of whether teachers understand and are able to align their teaching to state standards, the news is mixed, but on the plus side some research has been illuminating: Many district and state superintendents and teachers have applauded the move toward greater alignment of curriculum to state standards. On the very important matter of the achievement gap, which has been a persistent problem and one that NCLB explicitly seeks to affect, there is also mixed news but on the positive side there is evidence that percentages of students scoring proficient have risen and that gaps between subgroups have narrowed in most states at the elementary, middle, and high school levels, although in a notable minority of cases gaps have widened (see for example, Dietz & Roy, 2010).
- A more comprehensive set of studies point also to significant progress in the narrowing of the achievement gap as a result not specifically of NCLB but of the more general standards movement of which NCLB is the latest example (Gamoran, 2008).

Is there a grand lesson here? Let me suggest that at least three basic conclusions are worthy of consideration by education researchers eager to contribute to improved policy making. First, there is a hardy appetite in the policy world for credible and reliable information derived from empirical study, and we should be proud of our community for its diligence in the pursuit of answers to complex questions. Second, the most interesting questions are, indeed, too complex to expect definitive or optimal solutions, and our goal should be to provide reasonable, rather than perfect recommendations, based on appropriate rather than exhaustive deliberation.<sup>4</sup> And finally, there is merit in analyzing reforms from the standpoint of their potential (and measurable) benefits, intended and unintended, along with their actual risks and downside effects. Our overarching goal as

---

<sup>4</sup> This is the essence of Herbert Simon's definition of "procedural rationality." See Feuer (2006) for application to education policy and research.

a community should be to engage with policy makers and politicians who are entrusted with making the tough decisions, humbly and carefully outline the pluses and minuses of any particular action, and offer our scientific expertise toward the design and implementation of programs that can help all our children learn.

## References

- Cremin, L. (1990). *Popular education and its discontents*. New York: Harper and Row.
- Cronbach, L. (1975). Five decades of controversy over mental testing. *American Psychologist*, 30(1), 1-14.
- Dietz, S. & Roy, M. (2010). *How many schools have not made adequate yearly progress under the No Child Left Behind Act?* Washington, DC: Center on Education Policy.
- Feuer, M. (2010). Externalities of testing: Lessons from the blizzard of 2010, *Measurement*, 8, 59-69.
- Feuer, M. (2006). *Moderating the debate: Rationality and the promise of American education*. Cambridge: Harvard Education Press.
- Gamoran, A. (2008). *Standards-based reform and the poverty gap*. Washington, DC: Brookings Institution.
- Goldin, C. & Katz, L. (2010). *The race between education and technology*. Cambridge: Harvard University Press.
- Linn, R. (2005). *Fixing the NCLB accountability system*. UCLA: CRESST.
- Mills, E. (1978). *The economics of environmental quality*. New York: Norton.
- Schelling, T. (1974). *Micromotives and macrobehavior*. New York: Norton.
- Tyack, D. (1974). *The one best system*. Cambridge: Harvard University Press.
- U.S. Congress, Office of Technology Assessment (1992, February). *Testing in America's schools: Asking the right questions*. Retrieved from <http://govinfo.library.unt.edu>

---

# *Facilitating a Faculty Learning Community: Determining Consensus using Q Methodology*

Susan Ramlo  
The University of Akron

## *Abstract*

*With plans to improve a Technical Report Writing course, writing faculty and engineering-technology faculty formed a faculty learning community (FLC). Although discussions were often productive, it was often difficult to gauge consensus and differing views among the group members. In a previous study, Q methodology, a measure of subjectivity, was used to facilitate discussions among a diverse set of faculty (Ramlo, 2005). Similar to that study, the researcher found here that determination of consensus and differing opinions in the FLC improved dialogue, built an improved sense of community, and led to improvement of the course.*

During the 2007-2008 academic year, Technical Writing faculty and Engineering Technology faculty formed a faculty learning community (FLC) as part of a Northeast Region Ohio Learning Network (OLN) grant. The focus of this particular FLC was to re-make a Technical Report Writing (TRW) course such that students would become more engaged in the course with an increased sense of the relevance of technical writing in their engineering technology careers. However, adapting a course in one department to meet students' and faculty's perceived needs in another department can be a difficult task. A variety of opinions existed about the course objectives, about students' abilities to learn and write, about teaching best practices, and about the course content. The FLC consisted of faculty from two distinct disciplines, English and Engineering. This paper focuses on how Q methodology, a measure of subjectivity, was used to investigate the views of a diverse set of faculty and how this investigation facilitated the revision of the Technical Report Writing course.

## **Background on the Problem**

Milton Cox began faculty learning communities at Miami University in the 1990s and the concept has spread, although the highest concentration remains in Ohio (Cox & Richlin, 2004). Cox and Richlin described FLCs as cross-disciplinary and consisting of a group of six to fifteen faculty and staff. Such groups are not simply committees or task forces. Instead, the FLC members engage in an active, collaborative, yearlong commitment to enhance teaching and learning. Yet compared to typical faculty reading circles (Ramlo & McConnell, 2008), seminars, or brown-bag lunch discussion groups, FLCs offer a teaching and learning experience that is more structured and intense with a greater focus on developing a sense of community (Cox, 2004).

For an FLC, the expectations involve the community's pursuit of a specific project related to teaching and often require efforts to build a sense of community within the group (Cox, 2004). Dewey (1916) said, "Men live in a community

in virtue of the things which they have in common; and communication is the way in which they come to possess things in common" (p. 5). Unfortunately, when a diverse set of faculty come together, they often do not possess things in common. These differences may make determining how the FLC members will effectively work together and share their views difficult (Cox & Richlin, 2004). Witte and Engelhardt (2004) stated that a group, even with a variety of views represented, must collaborate in order to accomplish their goals. Such collaboration must be democratic in that the various views represented need to be heard and those involved must feel a sense of empowerment (Clark et al., 1996).

## **The Technical Report Writing Course**

Technical Report Writing (TRW) was originally designed to serve all of the engineering technology (ET) programs (seven Associate degrees and five Bachelor degrees in ET) at The University of Akron to satisfy our accreditation requirements. However, a review of programs now using the TRW course, which qualifies as an English general education course, revealed that TRW now serves a wide range of associate programs, from Criminal Justice Technology to Business Technology. Perhaps because of this diverse student population, the TRW course has focused more on the quality of student writing, such as organization and flow, than true technical writing. However, at the same time, engineering technology faculty stated that they wanted the course to focus on critical thinking, analysis, and authentic applications for engineering and engineering-technology professionals. Both technical writing and ET faculty were concerned that 25% of engineering technology students and 33% of all students taking TRW receive a D, F, or withdrawal from this course. In addition, the results of 39 student interviews demonstrated that, overall, students did not feel engaged in the TRW course nor did they feel the course was relevant to their engineering technology field of study. Yet one of the major program goals for each of these engineering technology programs is that: Students will demonstrate the ability to communicate effec-

---

tively with written, oral, and visual means in both technical and non-technical settings. Such skills make our graduates more employable (The National Commission on Writing, 2004) and assist in maintaining our engineering technology program accreditation.

It was apparent from the beginning of the OLN grant writing process, that all of the learning community members agreed that communication skills were important for the engineering technology students. However, FLC members did not agree on how to best accomplish improving students' communication skills nor did they agree on the focus or content of the course. The division of ideas was most apparent from the suggestion that the engineering technology faculty should teach a technical writing course instead of the "English" faculty. Another suggestion, also from the engineering technology faculty, was to replace TRW in the engineering technology programs with English Composition II. The technical writing faculty suggestions were greatly based upon maintaining the status quo.

### Consensus & Views in an FLC

With two sets of faculty from two different departments and with different professional backgrounds (English versus Engineering), it was not surprising that a variety of views about the project and revision of the TRW course appeared during discussions. Yet the FLC members had to determine how to effectively work together and share their views while moving toward consensus about the revisions to the course. Efficient group techniques that build a cooperative and consensus-building environment are necessary, especially in diverse groups (Witte & Engelhardt, 2004). Certainly a diverse group membership can create difficulties with group process and the development of consensus (Knight et al., 1999). Angelo and Cross (1993) describe a variety of strategies to assess the FLC process self-reflection, focus group interviews, meeting notes, projects, portfolios, faculty presentations, and surveys. The dynamics of our group were compounded by the fact that the FLC facilitator was one of the engineering technology faculty. After months of discussion, the FLC and facilitator needed a means of determining consensus of the group as well as differing opinions. Q methodology has been used in other studies to reveal this type of information and facilitate discussions among diverse groups (Focht, 2004; Ramlo, 2005)

### Participants & Demographics of this Study

The FLC consisted of five full-time engineering technology faculty and five full-time technical writing faculty. In addition, 11 engineering technology students (six seniors and five juniors) also participated in this study because they were seen as key stakeholders. Certainly a large group of students demonstrated, during earlier interviews related to engineering technology program assessments, strong opinions about the current writing course, especially regarding relevance of the course for engineering technology majors. Some of

their input during those discussions was used to develop statements used in the Q methodology study here. All of the students who participated in the Q methodology study had worked in their engineering technology field. Participants only identified themselves as students (including class rank) or faculty (TRW or ET) to provide anonymity and, therefore, freedom to express opinions since this was important for the purpose of the study.

### Purpose of the Study

The FLC facilitator had used Q methodology previously to determine consensus about the creation of a School of Technology (Ramlo, 2005). Like that situation, the FLC consisted of a variety of faculty, some of whom were more vocal about their views than others. Discussions were productive but did not reveal the consensus and distinctions between the different views of the entire group. Previously, Q methodology was used to prepare facilitators for dialogue regarding a topic through revealing consensus and differences of opinion (Focht, 2004; Ramlo, 2005). The purposes of this study were to use Q methodology as a means of determining consensus as well as differing opinions about the revision of the writing course and then use these to facilitate discussions within the FLC. In other words, we investigated how determining consensus and differing opinions would enable the FLC facilitator to select talking points for further discussions about the course improvements. This process was thus aimed at not only helping to build a better sense of community within the group, but also allowing for improved dialogue and commencement about specific course changes.

### Supporting Literature & Methodology

Q methodology is a measure of subjectivity in that it allows the various views on a topic to be determined (Brown, 1980, 2008; Stephenson, 1953). This method has been used in a variety of studies within higher education (Ramlo, 2005, 2006/2007, 2008; Ramlo, McConnell, Duan, & Moore, 2008; Ramlo & Nicholas, 2010). One of these studies, by Ramlo (2005), used Q to determine the various views and consensus among faculty pertaining to the possible creation of a School of Technology at a large, public, metropolitan university. In that study, the results from performing the Q study assisted faculty to address organizational change. Similarly, in this study, Q methodology was chosen because we sought to reveal the multiple views on revising the writing course along with consensus among those views. Alternative methods, such as Likert scale evaluations, lead to the loss of meaning (McKeown, 2001). This loss of meaning is based upon the Likert-scale survey's focus on generalizing to one overall view of the topic whereas Q methodology's purpose is in revealing the various views that exist within a specific population (McKeown, 2001; ten Klooster, Visser, & de Jong, 2008). Q methodology is often considered a mixed research methodology because it blends qualitative and quantitative aspects of research (Newman & Ramlo, 2010; Stenner &

Stainton-Rogers, 2004). Tashakkori and Teddlie (2009) have called this type of methodology inherently mixed.

Any Q study begins with the selection of items to be sorted by the participants. These items often come from interviews, focus groups, written communications, and other similar sources. Q allows participants to provide their perspectives by sorting items, typically statements related to the topic, into a distribution such as the one shown in Figure 1, used for this study. Because the sorting process involves interpretation of the items by the sorters and each are judged relative to the others based upon this interpretation, validity is not a consideration within Q methodology studies. In other words, no external criterion for a person's point of view exists and, therefore, the issue of validity of Q sorts does not apply. Similarly, operational definitions are not meaningful in Q because the researcher's view of the items is independent of the determination of the views of the participant (Brown, 1980). For this reason, interviews or participants' written comments are often used to help sorters expand on their interpretations of the items as well as their specific item placements. This type of input from the sorter assists the researcher in interpreting the factors (Brown, 1980; McKeown & Thomas, 1988). It is also important to note that the test-retest reliability of the Q sorts has been shown to be 0.80 or higher (Brown, 1980).

In this study, 47 statements were taken directly from either communications between the researcher, FLC members, and students. Many of these communications were emails

between the parties about different aspects of the Technical Writing course while others were informal interviews by the researcher. These communications included discussions related to the course's learning objectives, student expectations, faculty expectations, and difficulties in teaching the course. Some of these discussions focused on typical Technical Report Writing syllabi including the types of assignments that were, at that time, part of the course. The 47 statements selected offered a diverse set that attempted to represent the communicability on the Technical Report Writing course. The list of statements is in the appendix. Members of the FLC and a small selection of students sorted these selected 47 statements based upon their views of the Technical Report Writing course. Students were included here because of their communications with the researcher and the selection of statements that came from those communications. Several of the participants commented that the sorting process helped them reflect upon the course revision and current technical writing course. Participants have made similar comments in other studies by the researcher (Ramlo, 2005; Ramlo, 2006/2007).

The participants' sorts were analyzed to determine the views about the writing course. In Q methodology the analyses involve statistical analyses such as correlation and the calculation of factor scores. In Q methodology persons are correlated into factors based upon their sorts, as opposed to items being correlated in R factor analysis (Brown, 1980; McKeown & Thomas, 1988; Stephenson, 1953). It is impor-

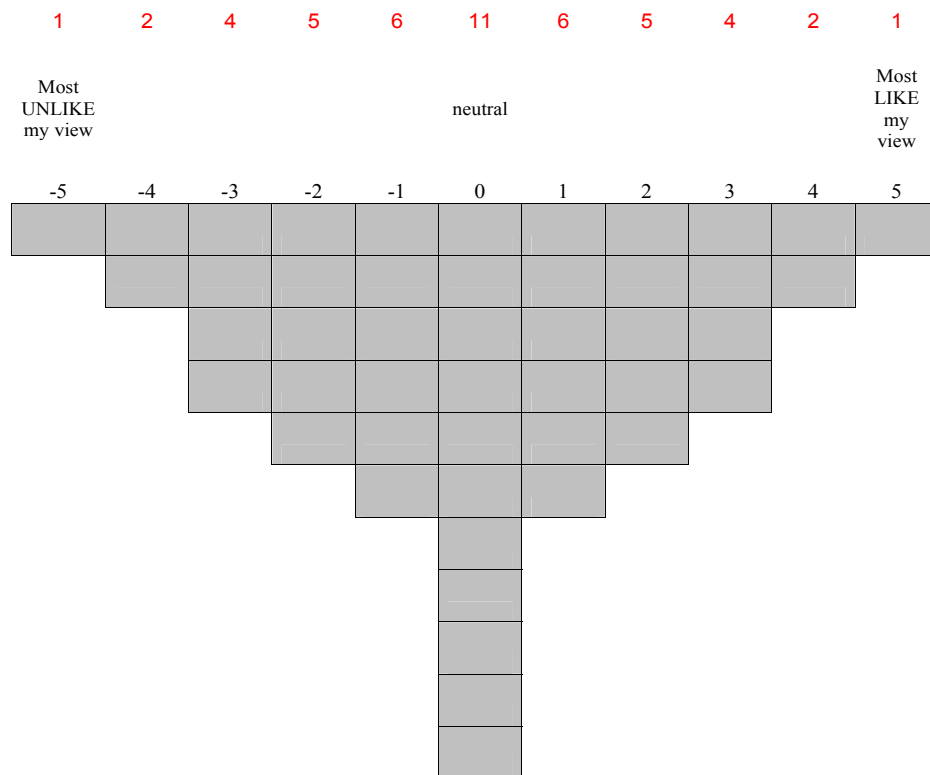


Figure 1. The grid used for the distribution of the Q sorts by each participant

tant to note that in Q, the sample size is the number of items sorted, not the number of persons sorting. Thus the sample size in this study is 47 because the participants sorted 47 statements. The factor analyses in this study used centroid as the factor extraction method because that is the preferred extraction method for Q methodology studies (Brown, 1980; Newman & Ramlo, 2010; Stephenson, 1953).

The factors determined within Q methodology studies represent different views about a topic. Those represented by the same factor hold similar views. In addition, there are often similarities among the different views/factors. The analyses of the sorts, therefore, also produce tables of consensus. These are the statements that are similar across the various factors based upon statistical analyses. In the school of technology study (Ramlo, 2005), consensus and distinctions among the views about the creation of a school of technology on a large, Midwest public university were determined using Q methodology. The participants in that study commented that using Q methodology helped them feel empowered because Q revealed the different perspectives instead of allowing those who were the most vocal to drive the discussions and direction of this organizational change. In addition, revealing the perspectives and consensus facilitated dialogue and collaboration among the group's membership. The results of that study encouraged the researcher to use Q methodology to facilitate the FLC's improvements to the writing course.

## Results

The factor analysis in this study revealed three factors, each representing a different opinion about the Technical Report Writing course. The factor matrix is shown in Table 1 with Xs indicating those who are represented by each of the three factors/perspectives. In the table, the QSORT column includes the coded Q sorter information. Engineering technology faculty members are coded as F-ET sorters. F-TRW is the code for technical writing faculty. The remaining codes are for ET students who were either juniors or seniors.

All five of the engineering technology faculty and four out of five of the technical writing faculty are represented by Factor 1. This factor also includes seven students: four seniors and three juniors. Factor 2 represents one TRW faculty person and another two students, both juniors. The remaining factor represents one student's view, a senior. Although in some cases, a factor describing one person may be of interest because that person may hold an influential role, for instance, in this study the researcher decided to focus on the first two factors in part because those factors included the most influential participants—the TRW faculty.

The Q methodology analyses result in a number of tables that are used to interpret the factors. Participant characteristics and written comments were also used here for interpretation of the views and consensus. Tables 2 and 3 include the six most like my view statements and the six most unlike my view statements, respectively, from the sort grid that represents this perspective. The statements in grid

Table 1  
*Factor Matrix with an X Indicating a Defining Sort*

Sorter #	QSORT	Factor Loadings		
		1	2	3
1	F-ET1	0.68 X	-0.37	0.22
2	F-TRW1	0.65 X	0.25	-0.18
3	F-ET	0.80 X	0.17	0.21
4	F-TRW2	0.65 X	0.30	-0.19
5	F-TRW3	0.47	0.73 X	-0.09
6	F-ET3	0.73 X	0.20	0.05
7	F-ET3	0.69 X	0.21	0.11
8	F-ET4	0.66 X	-0.19	-0.09
9	F-TRW4	0.51 X	0.24	-0.24
10	Senior	0.60 X	-0.13	0.31
11	SENIOR2	0.56 X	0.03	0.07
12	SENIOR3	0.13	-0.26	0.56X
13	JUNIOR	0.28	0.68 X	-0.07
14	JUNIOR2	0.52 X	-0.31	0.24
15	SENIOR4	0.65 X	-0.19	0.08
16	SENIOR5	0.26	-0.01	0.09
17	JUNIOR3	0.68 X	-0.12	0.09
18	JUNIOR4	0.16	-0.46 X	0.27
19	SENIOR4	0.62 X	-0.23	0.02
20	JUNIOR5	0.58 X	-0.36	-0.03
21	F-TRW5	0.64 X	0.37	-0.11
% explained variance		33%	11%	4%

*Note:* F-ET represents engineering technology faculty; F-TRW represents technical writing faculty; students are represented by either JUNIOR or SENIOR codes. X indicates that this sort is selected as represented by this factor.

positions 5 (statement 38), 4 (statement 46), 3 (statements 29, 30, and 41), -3 (statements 45 and 21) indicate this view is focused on creating multiple, small authentic activities that focus on the integration of data presentation, discussion and analysis for the TRW course. Statements 14 (grid position 4), 3 (in grid position -3), and 15 (position -5) indicate that this view believes that students need to develop strong communication skills.

These views were also supported by the comments made by those represented by Factor 1 regarding their most extreme placements (+5 and -5) of the statements. One Factor 1 sorter comment revealed that the sorting process helped him reflect on the course and potential changes. All of the sorter comments for those represented by this view focused on the need for students to be good communicators and to have real-world experiences that are practical and useful. Because of this views' stress on authentic learning and the importance of communication skills, the researcher named this factor "Authentic assignments to promote the development of strong communication skills for students."

Tables 4 and 5 contain the top six "most like" and "most unlike my view" of the TRW course for Factor 2. These statements indicate that this factor also believes that TRW should stress real-world technical writing assignments but, unlike Factor 1, statements regarding revision of the course appear within the more neutral area in the center of the sorting grid. Thus, this view appears more ambivalent about the course revisions than Factor 1. Yet the top six "most like my view"

Table 2

*Factor 1—Top Six Most Like My View of the Technical Report Writing Course*

Item #	Q sample Statement	Grid position	z-score
38	A Technical Writing course should integrate presentation of analytical data and results, including graphs, with the writing content.	5	1.917
46	Real-world technical writing assignments would make students see the relevance of technical writing skills.	4	1.874
14	Professionals must know how to communicate to both technical and non-technical personnel.	4	1.788
29	Students need to learn how to present data and results within a variety of "technical" reports.	3	1.731
30	Students need to learn how to discuss data and analyses within a variety of "technical" reports.	3	1.725
41	A variety of small assignments would be better than a few small and several large assignments for teaching technical writing.	3	1.083

Table 3

*Factor 1—Top Six Most Unlike My View of the Technical Report Writing Course*

Item #	Q sample Statement	Grid position	z-score
45	Creating a marketing flyer is an important assignment for a Tech Report Writing course.	-3	-1.160
21	Only the engineering technology students need to learn how to write a conclusion.	-3	-1.343
3	Quality writing isn't that important to students in Science, Technology, Engineering, and Mathematics (STEM).	-3	-1.483
36	A large (15-20 pages) writing assignment that is worth half the total points available is appropriate for a technical report writing course.	-4	-1.527
20	Tech Report Writing is fine just the way it is; no changes to the course are needed.	-4	-1.757
15	Communication skill is overrated and most students won't have to do much writing in their professions.	-5	-2.088

statements for this view appear to support the continuation of what have been typical TRW course assignments (statement 39 about email writing in position 4, statement 40 about writing memos in position 3, and statement 33 about electronic research techniques in position 3 all of which came from TRW faculty syllabi). Thus, this view appears to be accepting of maintaining current practice within the TRW course. Similarly, this factor's most unlike statements indicate rejection of students' negative opinions of TRW that came directly from student interviews (statements 27, 8, and 10 in positions -3, -3, and -5, respectively). Therefore, this factor was named "TRW is a good course and we are ambivalent about the potential course revisions."

Factor 2 is bipolar in that two of the loaders had positive factor loadings (one TRW faculty person and one junior) while one, a junior, had a negative factor loading. Thus, the negative loader would have an inverted statement grid compared to the other two sorters represented by Factor 2. In other words, for this junior represented by Factor 2, Table 4 represents the top six most unlike my view statements and Table 5 represents the top six most like my view statements. Thus, the negative loader has a negative view of the TRW course as well as the current course materials yet is also ambivalent about the potential of course revisions.

Although three distinct views about the TRW course emerged in this study, Table 6 contains the eleven consensus

statements that do not distinguish between any pair of the three factors. The grid-positions for these eleven statements, for each factor, are also given in the table. Most of these consensus statements listed are within the neutral areas of the representative factor grids (-1, 0, 1). However, two of the statements are in more prominent grid positions for each of the factors. The placement of statement 9 at the -3 position for Factors 1 and 2 as well as at the -1 position for Factor 3, indicates that each of these views believe that the D, F, and withdraw rates can be improved with improvements to the course. Consensus statement 46 at the 4 position for each of the three factors indicates that introducing authentic assignments is unanimously viewed as the way to improve relevance, learning, and pass-rates within the TRW course. The consensus about the development of such activities became the focus of the FLC for the remainder of the project after revealing the results of the study to the FLC members at our first gathering in January 2008. At that meeting, the above results were presented and the FLC discussed how these views were related to the remainder of our time as an FLC and our work related to revising the writing course.

As Focht (2004) indicated, the results of the Q analyses allowed for facilitation of this dialogue regarding the TRW course by revealing consensus and differences of opinion of both faculty and students. Even difficult discussion topics, like students' opinions about the TRW course being irrelevant

to their course of study, appeared to be facilitated through the revealing of consensus and differing opinions determined through Q methodology. In addition, these discussions during January 2008 allowed our FLC to better discuss the direction

or directions for the FLC. The results were also occasionally referred to in subsequent meetings however the majority of the discussions related to the Q study results were during that January meeting.

Table 4  
*Factor 2—Top Six Most Like My View of the Technical Report Writing Course*

Item #	Q sample Statement	Grid position	z-score
12	Assignments in Tech Report Writing represent the types of writing students will be required to perform in their professional lives.	5	2.454
46	Real-world technical writing assignments would make students see the relevance of technical writing skills.	4	1.820
39	Students should learn proper email writing in technical report writing.	4	1.731
34	Students need to be able to present effective technical reports orally and that should be taught in Technical Report Writing.	3	1.434
40	The writing of memos (memoranda) is an important aspect of a technical writing course.	3	1.373
33	Students need to know electronic research techniques / methods.	3	1.328

Note: These statements represent the top six most like my view for the negative factor loader since this is a bipolar factor.

Table 5  
*Factor 2—Top Six Most Unlike My View of the Technical Report Writing Course*

Item #	Q sample Statement	Grid position	z-score
9	No matter what is done, a lot of students will continue to receive a grade of D, F or withdraw from the Tech Report Writing course.	-3	-1.352
27	Students believe that the Technical Report Writing course is a waste of their time.	-3	-1.373
8	English Comp II would serve the students better than Technical Report Writing.	-3	-1.467
7	There is no real difference between the English Comp and Technical Report Writing courses.	-4	-1.539
25	The Technical Writing course is more about writing than technical writing.	-4	-1.584
10	There's no real substance in the Technical Report Writing course assignments.	-5	-2.056

Note: These statements represent the top six most like my view for the negative factor loader since this is a bipolar factor.

Table 6  
*Consensus Statements that Do Not Distinguish Between Any Pair of Factors*

No.	Statement	Factor 1 Position	Factor 2 Position	Factor 3 Position
5	Students should be required to get a C- or better in English Comp as a pre-requisite to taking Tech Report Writing.	1	0	2
6	What is presented and how it is taught should be consistent across all sections of Technical Report Writing.	0	-1	1
9	No matter what is done, a lot of students will continue to receive a grade of D, F or withdraw from the Tech Report Writing course.	-3	-3	-1
11	Assignments in Tech Report Writing assist students in developing critical thinking skills.	0	0	0
13	Writing a resume and cover letter are important assignments in Tech Report Writing.	1	2	0
16	Students learn how to write conclusions in Tech Report Writing.	0	1	1
17	Students don't realize how much writing they'll need to do in their professions.	3	1	0
18	Tech Report Writing focuses too much on the quality of the writing and not enough on the content of the writing.	0	-1	1
31	Students need to know how to write an abstract for their professional lives.	0	0	0
35	Writing essays is important for teaching technical writing skills.	-1	-2	0
46	Real-world technical writing assignments would make students see the relevance of technical writing skills.	4	4	4

---

For instance, the FLC had previously begun the creating a wiki that would focus on Tharp and Gallimore's (1988) seven ideas for assisting learning and other learning strategies for teaching TRW. Based upon the Q results, including the consensus statement regarding the need for more authentic technical writing assignments, the FLC members decided that it was best to solely focus on the creation of the authentic writing assignments instead of also branching into learning strategies. This discussion, as well as others that day, revealed that the Q results helped our FLC obtain a greater sense of community and allowed our group to focus on specific tasks related to the course's improvement, including the creation of authentic learning activities. These activities were created and used within numerous sections of TRW. In addition, because of discussions that followed the Q methodology study, several sections of TRW were team taught with a technical writing instructor and a statistician/ math instructor. This combination has allowed further enhancements related to data analysis, results, and presentation, activities that the writing faculty felt uncomfortable with initially.

### Conclusions & Implications

Although discussions were ongoing, the dynamics of our group often made it difficult to determine whether or not the FLC was moving toward developing a more unified view for the direction for revising the TRW course. Cox (2004) stated that FLC members must determine how to effectively work together and share their views. An important aspect of effectively working together is discovering consensus and distinguishing opinions so that dialogue can be facilitated (Focht, 2004; Ramlo, 2005). In this study, Q methodology was used to determine consensus and distinguishing opinions about the revision of a technical writing course.

Using Q methodology, the researcher was able to identify factors that represent the various perspectives about the TRW course as well as consensus. The analysis of the sorts revealed that our FLC had indeed moved toward a more unified view of the TRW course and the direction to take for its revision. Specifically, this view represented agreement about revising the TRW course to make it more relevant to students including the creation of authentic learning assignments. In addition, according to written statements from the sorters, the sorting process allowed students and faculty to better reflect on the course and how to improve it.

Overall, the Q results enabled our group to understand that we had reached an agreement about the direction for revising the TRW course which was not evident through our group discussions. Reflecting on the various opinions and consensus about TRW, facilitated our mission to improve this course and confirmed findings by Focht (2004) and Ramlo (2005).

This experience was similar to the study by the researcher mentioned previously where she determined consensus and differing opinions about the development of a School of Technology (SOT) on at a large, public, metropolitan

university (Ramlo, 2005). That study also demonstrated the ability of Q methodology to allow for the determination of consensus and different opinions. In the SOT study, like this FLC study, the revealing of consensus and differing views allowed the faculty to better discuss potential changes and their implications for students and faculty. The SOT study, like the FLC study, demonstrates how Q methodology can be used to facilitate change and improve discussions among stakeholders where all voices can be heard, not just those who speak the loudest or the most frequently. The researcher suggests that FLCs use Q methodology to facilitate the communications among the faculty in the community and to assist the FLC to move forward for implementing changes instead of attempting to determine consensus and opinion through meetings and other discussions which do not always allow for such determinations. As Cox (2004) suggested, FLCs must determine how to discuss and share their views and this study demonstrates that Q methodology can be a useful tool for this endeavor.

### References

- Angelo, T. A., & Cross, K. (1993). *Classroom assessment techniques: A handbook for college teachers* (2<sup>nd</sup> ed.). San Francisco: Jossey-Bass.
- Brown, S. R. (1980). *Political subjectivity: Applications of Q methodology in political science*. New Haven, CT: Yale University Press.
- Brown, S. R. (2008). Q methodology. In L. M. Given (Ed.), *The Sage encyclopedia of qualitative research methods* (pp. 700-704). Thousand Oaks, CA: Sage.
- Clark, C., Moss, P. A., Goering, S., Hener, R. J., Lamar, B., Leonard, D., ... Wascha, K. (1996). Collaboration as dialogue: Teachers and researchers engaged in conversation and professional development. *American Educational Research Journal*, 33, 193-231.
- Cox, M. D. (2004). Introduction to faculty learning communities. *New Directions for Teaching & Learning*, 97, 5-23.
- Cox, M. D., & Richlin, L. (2004). *Building faculty learning communities*. San Francisco: Jossey-Bass.
- Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. New York: The Macmillan Company.
- Focht, W. (2004). *Q and policy sciences research*. Presentation as part of a panel discussion at the International Society for the Scientific Study of Subjectivity Annual Conference, Athens, GA.
- Knight, D., Pearce, C. L., Smith, K. G., Olian, J. D., Sims, H. P., ... Flood, P. (1999). Top management team diversity, group process, and strategic consensus. *Strategic Management Journal*, 20(5), 445-465.
- McKeown, B. (2001). Loss of meaning in Likert scaling: A note on the Q methodological alternative. *Operant Subjectivity*, 24, 201-206.

- 
- McKeown, B., & Thomas, D. (1988). *Q methodology*. Newbury Park, CA: Sage.
- Newman, I., & Ramlo, S. (2010). Using Q methodology and Q factor analysis to facilitate mixed methods research. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (2nd ed., pp. 505-530). Thousand Oaks, CA: Sage.
- Ramlo, S. (2005). An application of Q methodology: Determining college faculty perspectives and consensus regarding the creation of a school of technology. *Journal of Research in Education, 15*(1), 52-69.
- Ramlo, S. (2006/2007). Student views of learning in an introductory college physics course: A study using Q methodology. *Operant Subjectivity, 30*(1 / 2), 52-63.
- Ramlo, S. (2008). *Student perspectives on learning physics and their relationship with learning force and motion concepts: A study using Q methodology*. Paper presented at the International Society for the Scientific Study of Subjectivity / Q Methodology Conference, Hamilton, ON.
- Ramlo, S., & McConnell, D. (2008). Perspectives of university faculty regarding faculty reading circles: A study using Q methodology. *The Journal of Faculty Development, 22*(1), 25-32.
- Ramlo, S., McConnell, D., Duan, Z., & Moore, F. (2008). Evaluating an inquiry-based bioinformatics course using Q methodology. *Journal of Science Education and Technology, 17*(3), 219-225.
- Ramlo, S., & Nicholas, J. B. (2010). In-service science teachers' views about learning physics after a one week workshop. *Human Subjectivity, 1*, 109-120.
- Stenner, P., & Stainton-Rogers, R. (2004). Q methodology and qualiquantology: The example of discriminating between emotions. In Z. Todd, B. Nerlich, S. McKeown, & D. D. Clarke (Eds.), *Mixing methods in psychology* (pp. 101-120). Hove, NY: Psychology Press.
- Stephenson, W. (1953). *The study of behavior: Q-technique and its methodology*. Chicago: University of Chicago Press.
- Tashakkori, A., & Teddlie, C. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage.
- ten Klooster, P. M., Visser, M., & de Jong, M. D. T. (2008). Comparing two image research instruments: The Q-sort method versus the Likert attitude questionnaire. *Food Quality & Preference, 19*(5), 511-518.
- Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge, UK: Cambridge University Press.
- The National Commission on Writing. (2004). *Writing: A ticket to work... or a ticket out: A survey of business leaders*. Unpublished manuscript. Retrieved from [http://www.writingcommission.org/prod\\_downloads/writingcom/writing-ticket-to-work.pdf](http://www.writingcommission.org/prod_downloads/writingcom/writing-ticket-to-work.pdf)
- Witte, E. H., & Engelhardt, G. (2004). Towards a theoretically based group facilitation technique for project teams. *Sozialpsychologie, 52*, 2-24.
- 

### Appendix – List of 47 statements from Q sample

---

1. Technical Report Writing students find this course relevant to their future professions.
  2. Teaching technical writing is the sole responsibility of those teaching English Comp and Tech Report Writing.
  3. Quality writing isn't that important to students in Science, Technology, Engineering, and Mathematics (STEM).
  4. Academic freedom is important & writing faculty should be able tailor assignments, assessments, & course goals as they believe is appropriate.
  5. Students should be required to get a C- or better in English Comp as a pre-requisite to taking Tech Report Writing.
  6. What is presented and how it is taught should be consistent across all sections of Technical Report Writing.
  7. There is no real difference between the English Comp and Technical Report Writing courses.
  8. English Comp II would serve the students better than Technical Report Writing.
  9. No matter what is done, a lot of students will continue to receive grades of D, F or withdraw from the Tech Report Writing course.
  10. There's no real substance in the Technical Report Writing course assignments.
  11. Assignments in Tech Report Writing assist students in developing critical thinking skills.
  12. Assignments in Tech Report Writing represent the types of writing students will be required to perform in their professional lives.
  13. Writing a resume and cover letter are important assignments in Tech Report Writing.
- 

*(Appendix is continued on the next page.)*

14. Professionals must know how to communicate to both technical and non-technical personnel.
15. Communication skill is overrated and most students won't have to do much writing in their professions.
16. Students learn how to write conclusions in Tech Report Writing.
17. Students don't realize how much writing they'll need to do in their professions.
18. Tech Report Writing focuses too much on the quality of the writing and not enough on the content of the writing.
19. Engineering Technology faculty should just teach their own Technical Writing course.
20. Tech Report Writing is fine just the way it is; no changes to the course are needed.
21. Only the engineering technology students need to learn how to write a conclusion.
22. Sets of specific activities are needed for each type of program (criminal justice, eng tech, etc) served by Tech Report Writing.
23. Students should enroll in different sections of Tech Report Writing based upon their intended major.
24. Tech Report Writing currently meets the writing / communication skill goals of the programs it serves.
25. The Technical Writing course is more about writing than technical writing.
26. Many students in Technical Writing have poor spelling and grammatical skills.
27. Students believe that the Technical Report Writing course is a waste of their time.
28. Technical Writing faculty need to gain some real-world technical writing experience.
29. Students need to learn how to present data and results within a variety of "technical" reports.
30. Students need to learn how to discuss data and analyses within a variety of "technical" reports.
31. Students need to know how to write an abstract for their professional lives.
32. Students need to know how to write lengthy technical reports for their future professions.
33. Students need to know electronic research techniques / methods.
34. Students need to be able to present effective technical reports orally and that should be taught in Technical Report Writing.
35. Writing essays is important for teaching technical writing skills.
36. A large (15-20 pages) writing assignment that is worth half the total points available is appropriate for a technical report writing course.
37. A writing course should allow re-submissions of work for a potentially higher grade (teaching for mastery).
38. A Technical Writing course should integrate presentation of analytical data and results, including graphs, with the writing content.
39. Students should learn proper email writing in technical report writing.
40. The writing of memos (memoranda) is an important aspect of a technical writing course.
41. A variety of small assignments would be better than a few small and several large assignments for teaching technical writing.
42. Teaching how to make Table of Contents and Table of Figures is important for a technical writing course.
43. The ability to write a "letter of transmittal" is an important skill for students.
44. Students in technical fields should only learn how to write for other technical colleagues.
45. Creating a marketing flyer is an important assignment for a Tech Report Writing course.
46. Real-world technical writing assignments would make students see the relevance of technical writing skills.
47. Teaching students how to include statistical analyses into a technical report is important for a Tech Report Writing course.

---

# *The Perfect Storm: How Policy, Research, and Assessment Will Transform Public Education*

James H. McMillan  
Virginia Commonwealth University

In the 2000 movie *The Perfect Storm*, as you are probably aware, there is an unusual convergence of several critical weather factors that set the stage for a destructive outcome that takes both property and lives. It has become a popular metaphor to describe how events come together in a unique way to have an exceptional influence on something, typically a negative impact. In education today there is also a perfect storm, one that won't affect property directly but will influence the lives of millions of students. It is interesting that Arne Duncan, the United States Secretary of Education, used the storm metaphor in a speech in June of 2009 (Duncan, 2009). Here is what he said:

Let me start by talking about the unique, historic, and powerful opportunity we have to transform public education. We have a perfect storm for reform: We have:

- The Obama effect;
- Leadership on the Hill and in the unions;

- Proven strategies for success; and
- The *Recovery Act* providing \$100 billion.

Of course Duncan's remarks are not about anything destructive, unless you argue, like some have, that he is talking about the destruction of locally controlled education. He clearly thinks that the above factors are coming together in a positive way.

I want to focus on a different kind of perfect storm, one which is bringing several factors together that will create what I believe will be a destructive force for student learning. My contention is that there are three powerful influences that are coming together that will shape public education in the future—policy and politics, research, and assessment. What is argued is that we will soon have national standards, national tests, a national curriculum, and value-added teacher and school evaluation (see Figure 1).

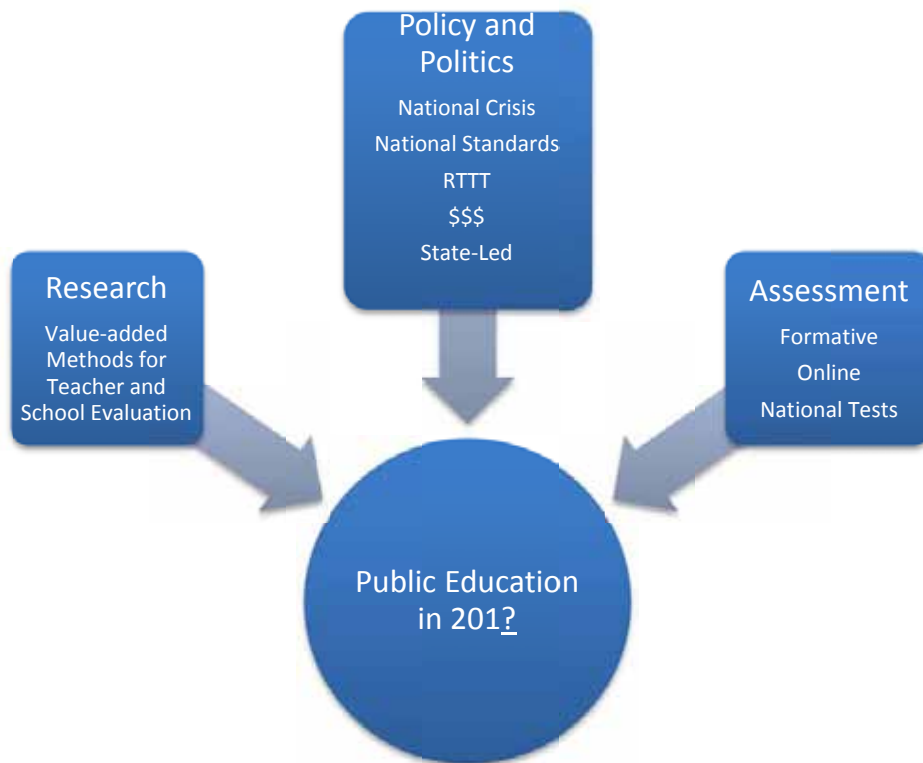


Figure 1. The Perfect Storm for Public Education

---

I believe the effect of these efforts is predictable based on previous experience; the effect on teacher evaluation is less clear. I will review each of these three factors, with a discussion of why they are detrimental, then list a few things we can do as educational researchers and assessment experts to mitigate the negative effects.

## Policy and Politics

Here we need to return to national-level policy and politics. In that same speech last June, Duncan also made the following points:

- The genius of our system is that much of the power to shape our future has, wisely, been distributed to the states instead of being confined to Washington.
- Our best ideas have always come from state and local governments.
- On so many issues . . . the states are often leading the way.
- We think that every state should set internationally benchmarked standards and assessments that prepare students for success in the workforce and college.

This does not sound like anything that portends an increasingly federalized system of education. Indeed, the administration has repeatedly indicated that the effort to develop national standards is not a federal initiative, citing state-led efforts of the National Governors Association and the Council of Chief State School Officers to develop Common Core Standards. The Common Core website uses the phrase “Common Core *State* Standards Initiative” (emphasis added). We now have such standards in mathematics and English/language arts (without naming specific pieces of literature to be read), which have been adopted by 34 states plus the District of Columbia. The standards are supported by common sense (yes, it makes some sense to have the same learning standards for all students), as well as by statements from influential individuals and some research. Chester Finn, Jr. based recent comments on a study undertaken by the Fordham Institute (Carmichael, Martino, Porter-Magee, & Wilson, 2010). He recently said, “The United States is approaching a set of agreed-upon national standards of a core of its K-12 curriculum, and I think that’s a healthy thing for the country” (Sawchuk, 2010). (Note his inclusion of both “national standards,” in contrast to language used by federal agencies, and “curriculum”).

The Fordham study has received favorable press with its conclusion that the Common Core standards in English/language arts are clearer and more rigorous than current standards in 37 states, and math standards in 39 states. On July 22, 2010, CNN based their article; *National learning standards make the grade*, in part on a favorable review of the Fordham study, saying that setting national standards is “gaining momentum according to an official of an educational think tank that compiled a national study comparing standards”

(Holland, 2010). All the states were given grades (with few receiving a letter grade of A that reflected the highest score) based on content and rigor, and clarity and specificity. These judgments were made by only two language arts experts and three mathematics experts, not exactly what we would hope for in rigorous, systematic, and unbiased research. Nevertheless, this study is cited as evidence that the Common Core will raise standards in most states.

While the effort to develop the Common Core has been headed by the NGA and CCSSO, adoption of them has been encouraged, one could say, by federal rules that tie much needed money for the states to agreement to use the Common Core. Both the Race to the Top grant competition, a \$4.35 billion pot of money, and Title I funding (\$14 billion) tie chances of funding to adoption of the Common Core. Another \$320 million pot has been awarded to two organizations to develop national tests of the standards (more on that later). Also, \$250 million in the Recovery Act is for improving statewide data systems, and the budget of Institute for Educational Services (IES) has been increased more than \$70 million from 2009. At the state level, data systems are being developed to track students and integrate resources such as teacher credentials and fiscal information with student outcomes. In a June 8, 2009, address to the annual IES research conference Arne Duncan said:

We want to know whether Johnny participated in an early learning program and completed college on time and whether those things have any bearing on his earnings as an adult.

Hopefully, some day, we can track children from preschool to high school and from high school to college and college to career. We must track high-growth children in classrooms to their great teachers and great teachers to their schools of education.

In other words, there has been an active federal role in promoting national standards and tests. It is a policy decision with clear consequences.

Another strong political factor is that the President has emphasized the interdependence between schooling and the economic recovery, without question a serious issue for all. In July 2010, President Obama emphasized that reforming education is the “economic issue of our time... It’s an economic issue when we know countries that out-educate us today will outcompete us tomorrow” (Calmes, 2010). Thus, education is in crisis and needs to be fixed (not too different from assertions made to justify NCLB). This is further supported by international comparisons.

All of these factors suggest that we may be racing to adopt rushed reforms, without careful research to know what will happen to education when these reforms are adopted. To be sure, as indicated below, research is part of the picture, and here is one area that we need to have our voices influencing.

---

## Research on Value-Added Models and Factors Influencing Test Score Variability

There is little doubt that value-added research models will be used to judge teacher and school performance. To a certain extent, the notion that teacher effectiveness can be measured by how much their students' scores improve by the end of the year makes sense and is easy to explain. In other words, how much have students learned in this class or school? The logic of this is compelling; why not judge teachers on gain scores, not according to the same set standards for all students? Wouldn't this be fair? Teachers would be compared on a more level playing field. Perhaps, but there are significant barriers to the use of value-added models.

The allure of value-added models is that factors such as family background, school resources, class size, previous achievement and a host of other variables can be used to isolate the effect of the teacher by comparing student expected growth (hopefully based on several years of data) to actual growth. But how this is accomplished is critical. The value-added model developed by Bill Sanders and used in several states has not been fully evaluated with an external review because part of it is "cloaked in proprietary secrecy" (Eckert & Dabrowski, 2010, p. 89). This lack of transparency and resulting appropriate external review is concerning, to say the least. It is related to another trend with value-added models. Some are developed by econometricians, individuals who can crunch numbers but may not have a good understanding of the nature of the data, limitations of the data, and consequences of reporting formats within school contexts. In the Value-Added Research Center at the University of Wisconsin at Madison researchers use the words "value-added productivity" (emphasis added), which suggests a business rather than education perspective. The models can be very complex and difficult to understand, and the manner in which results are reported is critical. In California, for instance, value-added scores for grades 3-5 were recently reported in the *Los Angeles Times* for the Los Angeles Unified School District; an economist and education researcher from the Rand Corporation did the analysis. Rank ordered results for every teacher (6,000 total) and school were included. The results were norm-referenced, so you could easily see how an individual teacher or school stacked up, and of course there had to be teachers at the bottom of the curve, no matter what improvement of scores. There are appropriate cautions about interpreting the results, including a statement that small percentile differences are not significant:

Value-added scores are estimates, not precise measures, and readers should not place too much emphasis on small differences in teacher percentiles...both sampling error and measurement error contribute to the variability of the estimated teacher effects...the teacher's "true" rank falls in a range

around each point estimate...the range of potential values for math was plus or minus 7 at the 20<sup>th</sup> and 80<sup>th</sup> percentiles (*Los Angeles Times*, 2010).

The problem is that the initial results were not reported with the standard error of measurement intervals, only one year of data was reported, and no other indicators of teacher effectiveness were included. Reporting data for the value-added system in Tennessee is obtuse and difficult to understand. Researchers at Vanderbilt University (NCPI, 2009) have used a simplified value-added model for linking student test scores with performance pay, but there is still a need to report results so that interpretations are appropriate.

Another consideration is how well value-added normative data fit with standards-based education. There is a clear record of research about the implications of norm-referenced evaluation. The logic of standards-based education, which has become the basic model of school reform, is criterion-referenced. But in standards-based models, student background is not controlled. If schools with high and low socioeconomic student populations show the same achievement, it is difficult to know if the standards are too easy, teachers in the low SES schools are terrific, or if teachers in the high SES schools are terrible. There is some development of status-based accountability based on test scores, as well as efforts to combine norm-referenced value-added data with status data (e.g., in Colorado) (Betebenner & Linn, 2009).

It will be interesting to gauge public reaction to reporting value-added results. In the September 2 issue of the *Wall Street Journal*, an editorial was titled "Teachers for Cover-ups." It targeted the Los Angeles teachers' union for objecting to the reporting of the scores, printing "Unions tell the *L.A. Times* to stop reporting test results." As could be expected, the *Wall Street Journal* defended the reporting of the scores and ranking. My hunch is that value-added results will be embraced by most non-educators and some educators, even with the caveat that standardized test scores signal but one of many important schooling outcomes, but we will see.

I believe there are several important issues with value-added models, beyond reporting of results, that need further research. One is preparing tests with sufficient "stretch" so that there is not a ceiling effect (Koedel, 2010), something that is common with standards-based assessments. This is needed to allow high scoring students room to improve. But to do this has obvious implications for the make-up of the test.

Another research-related factor to consider is how much teachers can actually influence the variability in student performance on standardized tests. Consider all the factors that influence student achievement on these tests that are outside the control of the teacher (e.g., general ability, native language, friendships, parental support, siblings, previous achievement, attendance, summer experiences, curriculum, district testing policies). This doesn't leave much that differences between teachers can influence. Schochet and Chiang

---

(2010) claim research has shown that 90% of the variability in student achievement is determined by student-level factors other than what the teacher can control (at the same time many claim that the teacher is by far and away the most important school-related factor to student achievement). Consequently, a limited amount of the remaining variability can be attributed to the unique contributions of an individual teacher (as differentiated from what any teacher provides). While value-added models help adjust for such differences, there is simply no way to fully account for these differences in a systematic manner.

## Assessment

There are several developments in the assessment field that will fuel national assessments. These developments are driven by an unprecedented convergence of three factors: substantial federal funding, “voluntary” participation in determining common state standards, and advances in technology. Ironically, research on the impact of formative classroom assessment has generated interest in making large-scale tests more responsive to student learning and relevant to instruction. This is clearly reflected in the RTTT funding of \$350 million in grants to support the development of a “new generation” of “multi-state” comprehensive assessments. Two groups have been funded with approximately \$160 million for four years for development of the comprehensive systems (SMARTER Balanced Assessment Consortium [31 states] and the Partnership for the Assessment of Readiness for College and Careers [PARCC; 26 states]). The “new” assessment systems must go beyond summative assessment and include an integrated set of performance assessments, as well as interim assessments that are described as “through-course,” accomplished during the school year. While this new emphasis on formative assessment is noteworthy and appropriate, it will be interesting to see how it can be achieved.

At issue is whether it is possible to use benchmark testing for what has been carefully and clearly defined as a process or series of steps used in formative assessment (William & Leahy, 2007; Brookhart, 2007; Popham, 2008). Consider the 2006 definition used by the Council of Chief State School Officers:

Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students’ achievement of intended instructional outcomes.

Note this definition includes *during instruction*, *providing feedback*, and *ongoing teaching*. These are characteristics not often associated with large-scale testing. William and Leahy (2007, p. 31) point out “a ‘formative assessment’ that predicts which pupils are likely to fail the forthcoming state-mandated test is not formative unless the information from

the test can be used to improve the quality of the learning within the system.” Popham (2008, p. 6) has recently made the same point in his definition of formative assessment, which emphasizes that formative assessment is a “planned process” in which evidence is used so that teachers “adjust their ongoing instructional procedures” or students “adjust their current learning tactics.” It is assessment with these characteristics that, according to the research, improves student learning.

It seems to me that what is being proposed is quite different from what is defined as formative in the context of on-going instruction. I’m not sure what to call it to differentiate it from a more instructionally relevant definition. Maybe something like “quasi-formative” would work, or maybe such assessments should be called “summative/formative” tests since they look like mini-summative tests that can provide limited feedback to teacher and students. Maybe we will all be pleasantly surprised, but the task is daunting.

The difference between what the classroom assessment literature contains about formative assessment and these “new” assessments is important because the evidence that formative assessment makes a difference in achievement is based on the definition that includes on-going, feedback, and immediacy. Empirical evidence that formative benchmark testing has a positive impact on student learning is both limited and mixed. For example, some research suggests that targeted instruction can lead to improvements in student test scores (Lachat & Smith, 2005; Nelson & Eddy, 2008; Trimble, Gay & Matthews, 2005; Yeh, 2006) as well as proficiency in reading and mathematics (Peterson, 2007). However, empirical investigations that utilized quasi-experimental approaches have found no significant differences between schools using benchmark assessments and comparison schools not using such tests (Henderson, Petrosino & Guckenburger, 2008; Niemi, Wang, Wang, Vallone, & Griffin, 2007). There is also little evidence that interim tests can be used to determine whether students are on track to successfully complete the end-of-year assessment (Brown & Coughlin, 2007).

The rhetoric of “new generations” assessments is appealing with its emphasis on interactive assessment items that require “higher order” thinking skills and the use of artificial intelligence to score open-ended responses. Both proposals include the development of online digital resources to improve teaching and learning, including professional development materials, all aligned to national standards. There is even consideration of combining interim assessments with a year-end assessment to reach a final student score.

The list of objectives upon which the new assessments are based is impressive if daunting (Center for K-12 Assessment & Performance Management, 2010):

- Aligned with national standards.
- Lower cost (hence online tests).

- Formative as well as summative.
- Fast turnaround (hence online tests).
- Use of adaptive test delivery.
- Assessment of problem solving with multi-step simulations.
- Greater accommodations for students with disabilities and ELL students.

At issue with all of this is whether the new assessments will reflect older or newer research about how learning occurs and cognition. Traditional large-scale assessments tend to reflect learning theory that emphasizes fragmented knowledge and limited conceptions of cognition. More recent research on learning and cognition has emphasized the mental structures needed for problem-solving and the organization of knowledge so that it is useful. Knowledge is constructed and stored so that it can be easily retrieved, depending on context, the nature of the task, and previous learning. It is a matter of knowing when, where, and how to use knowledge, not simply demonstrating what is known and understood. Hence, students need to develop sophisticated understandings of how core concepts and explanations are applied to decision-making and problem-solving. Research on constructivism and learning progressions provides a basis for developing assessments on this more sophisticated idea

of learning (Pellegrino, 2009).

Can the currently funded assessment development projects reflect more contemporary theories of learning and cognition? It will depend in large part on the nature of the standards that are assessed. The current plan to utilize through-course assessments throughout the school year is a step in the right direction, as is the emphasis on more constructed-response and performance assessments. It will be interesting to see if this emphasis reflects more recent learning theory or whether it becomes a series of mini-summative assessments, like what is now occurring with interim assessments.

### Error (there is more than what you are led to believe)

There are several sources of error, both systematic and random, that must be considered for the next generation of accountability tests. For many years we in the research and measurement community have known about the deleterious effects of using standardized test scores to judge teacher effectiveness. One of the best insights was offered by Donald Campbell (1979). His conclusions have become known as “Campbell’s law,” and it is relevant for many fields, including education (Rothstein, 2008). This is what he asserted:

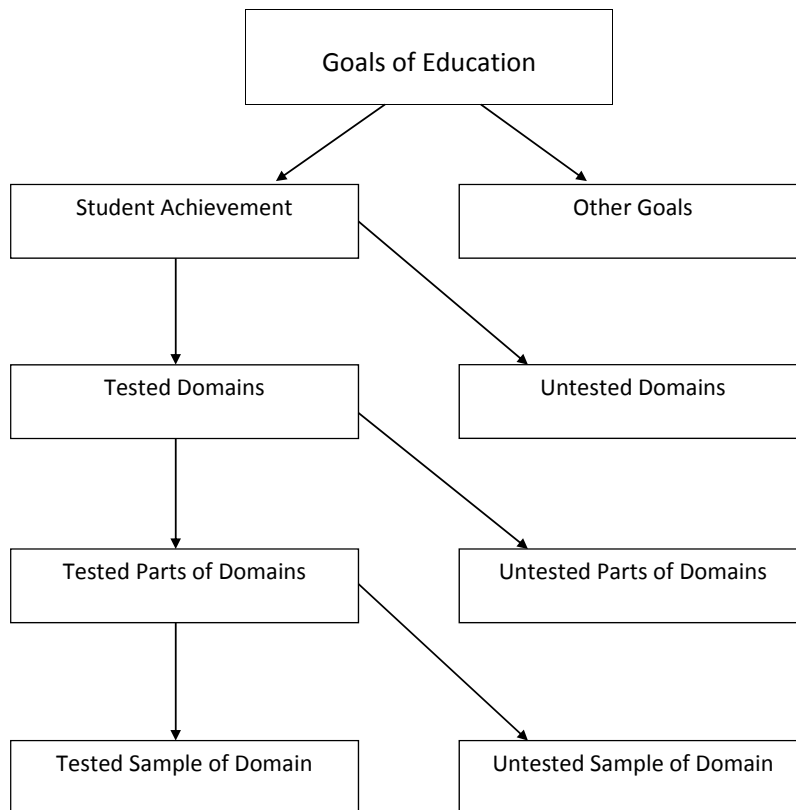


Figure 2. The Effects of Sampling That Narrow What is Tested (Adapted from Koretz, 2010).

---

The more any quantitative social indicator is used for social decision-making [e.g., teacher effectiveness], the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor (p. 85).

Furthermore, he stated that:

From my own point of view, achievement tests may well be valuable indicators of general school achievement under conditions of normal teaching aimed at general competence. But when test scores become the goal of the teaching process they both lose their value as indicators of educational status and distort the educational process in undesirable ways ... Achievement tests are, in fact, highly corruptible indicators (p. 85 )

A primary cause of Campbell's law is that there is incomplete and imperfect measurement of *desired* outcomes. The factor that makes this measurement incomplete is sampling error. The amount of sampling error is realized by considering all stages of sampling to get to the final test. Figure 2 illustrates these steps. At the outset only certain goals of education are selected, namely mathematics and reading/language arts achievement goals. Then achievement goals are limited to those domains of achievement that are sampled. Once domains are identified, parts of the domain are sampled, and then there is a sample of each part.

When you consider teacher effectiveness, similar sampling takes place, except now the achievement results, based on incomplete sampling, are used as an indicator of teacher effectiveness. In other words, only a *sample* of how "effective" the teacher has been is measured. There is error associated with sampling, and many desirable teacher benefits are not included (e.g., influencing a student to stay in school, developing a positive attitude toward reading, enhancing prosocial skills).

The sampling dynamic leads to the corruption of education by shifting resources allocated to tested subjects. Koretz and colleagues call this *between-subjects reallocation* (Koretz & Hamilton, 2006; Koretz, McCaffrey, & Hamilton, 2001), and summarize evidence to document the effect.

The amount of error that results from sampling must be added to two additional sources of error – measurement error and cohort effects. Measurement error is well described if under-reported. Typically a single source of measurement error is included, and that is most commonly internal consistency. Even high internal consistency reliability estimates, however, result in a fair amount of error in making final determinations such as pass/fail, or for teachers – adequate/inadequate. This is illustrated nicely with some data from the Virginia Standards of Learning test results. According to technical manuals, the overall amount of likely misclas-

sification is typically about 10% for 5<sup>th</sup> grade math. There is a 4% false negative result, just attributed to the measurement error. If similar statistics result when making decisions about teachers (adequate/inadequate), 4 of 100 teachers could be unfairly terminated.

A recent IES report addresses misclassification error rates using value-added data when measuring teacher and school performance (Schochet & Chiang, 2010). Using simulations, they estimate that the total percentage of misclassified teachers using three years of data is about 26%. That is, about 26% are false positives and false negatives. One in four teachers are misclassified.

Cohort effects are very difficult to control. Obviously, in any given year a teacher may have more or less able and motivated students. Students seem to come together as a group in some classes but not others; some students "lose" more knowledge over the summer than other students. Teacher-student relationships vary. More students are absent for some classes. There is more in migration of students for some classes. Changing the criteria for student assignments to different teachers may be important. Every teacher knows that every class is unique, even if there is random assignment of students to each teacher. These factors are identified by Kane and Staiger (2002) as random differences across classrooms. Cohort effects are very real and are only adjusted by presenting many years of data, with the assumption that these effects eventually even out.

Another consideration that results in error in our conclusions about student learning and teacher effectiveness is the well-documented *test inflation* factor. Test inflation occurs when increases in scores do not match increases in actual student knowledge and understanding. As we have seen with NCLB, percentages of students judged to be proficient keeps climbing (though now we're seeing some ceiling effects). The question is whether the increase in scores is an indicator of student achievement or reflects on many factors that result in higher scores without the associated gain in achievement. This is essentially a validity issue. What inferences are appropriate about student learning?

Research on test inflation has documented large exaggerations of improving accountability test scores (Koretz, 2008). The best recent illustration includes examples of studies that show more improvement on state-level high-stakes test scores than on NAEP. For example, research on scores from Kentucky in the 1990s showed significant gains on the KIRIS over three years, with no improvement on NAEP. Similar patterns were found in Texas. But even standardized achievement tests many years ago showed test inflation when scores at the end of several years use of the same standardized test resulted in lower scores on the newly standardized version of the test (which then would show gradual improvement each year). What happens is that over time teachers focus instruction on what is on the test, use

---

classroom test items that are similar to what is used on the high-stakes tests (e.g., more multiple-choice items), tend to use test items that they remember from the high-stakes test, enhance students' test-wise capabilities, cheat, coach, read items to students, give extra time to finish the test, teach writing to be consistent with the scoring rubric, and excessive drilling on knowledge tested. The goal is higher test scores, not greater student knowledge, understanding, and problem-solving ability. Teachers may also focus instruction on "bubble" students, ones who are close to passing, with less emphasis for very high performing as well as very low performing students.

Three things seem inevitable – 1) there will be more testing; 2) the stakes will be higher; and 3) there will be greater standardization across states. This will inevitably lead to more test prep and teaching to the tests. The prospect of a school and state performing poorly on national tests will generate considerable motivation to do whatever is needed to improve test scores, leading to test score inflation and less emphasis on what is not on the test. The current considerable influence of test-based accountability on teaching and learning seems poised to become even more powerful. There will be significant pressure on teachers to focus on what is tested.

### Surviving the Storm

The movement to national standards and tests is powerful. We are now desensitized to high-stakes testing and have the technical capability to use complex approaches to teacher and school evaluation. So if the "Perfect Storm of Reform" is coming, what can we do to minimize the destruction it could wreak on student learning? I believe the following are things we can do with assessment and research that can have a positive impact.

#### *Assessment Development*

- Get involved immediately in the construction of high-stakes tests to ensure that these tests are developed with sufficient attention to validity, reliability, and fairness (the three pillars of educational measurement), and that important, high-level standards for learning are assessed (e.g., inference, problem-solving, deep understanding). This should include developing tests that provide the correct types of evidence that can be shown to be appropriate for evaluating teachers. We also need to get involved with state tests and reporting options.
- Employ multiple methods of assessment, even if this is less cost efficient, perhaps on a sampling basis (matrix sampling).
- Emphasize the need for standards and tests to be compatible with contemporary learning theory.
- Become involved in state test design and reporting options.
- Monitor the integrity of data systems and encourage data that can examine trends over several years.

#### *Evaluating Teachers and Schools*

- Measure, "count," and report everything that is important in defining teacher effectiveness.
- Emphasize that value-added models of teacher effectiveness are at best only a general indicator of teacher effectiveness, and that more assessment may be warranted as a follow-up to verify and identify more specific areas of concern. There is error that needs to be accounted for, and using norm-referenced analyses may distort the differences between teachers.
- Report all important school data together, not just value-added scores, to provide context and a balanced perspective on school effectiveness. Context would include such "input" factors as student socioeconomic status, size, teacher characteristics, and special programs. Contextual information should also be presented in displaying teacher effectiveness data. Do not come up with single grades for schools.
- Consider results from a single test as an *indicator* or *snapshot* that requires further evidence.
- Combine value-added with status-based approaches.
- Monitor unintended consequences and factors influencing test inflation.

#### *Reporting*

- Report and explain confidence intervals and standard errors of measurement. These are not so technical that parents and others can't understand. The concept of margin of error is well understood once explained (hopefully, though this is based only my own experience).
- Avoid reporting of scores of small student subgroups.
- Avoid reporting of single year "growth." Use several years of data longitudinally to indicate stability over time.
- Be suspicious of large gains in any one year.
- Use plain, nontechnical language
- Present concise summaries.
- Utilize graphs and charts.
- Provide guidance for how to use the results.

#### *Other*

- Involve parents in the development, reporting, and use of assessment results.
- Conduct research on the impact of assessments on instruction and student achievement.
- Provide on-going teacher and administrator professional development to ensure accurate, uniform understanding of how to use results.
- Keep a close eye on econometricians and other quasi-educators.
- Use policy issues as examples in instructing preservice teachers and school administrators, and focus

---

professional development on assessment and research principles and issues that are critical in the appropriate interpretation and use of assessment data.

- Gather data that are locally relevant and meaningful.

### Summary

In summary, bring on the storm! We are equipped and motivated to fight for what is right for our students and the system of education in our country. We can't be complacent during this critical time of establishing national standards and national tests. By understanding and communicating important principles of research and assessment we can work with politicians and others to influence policy. The next few years will be both exciting and daunting, but just as we tell our students to be engaged in learning, we need to be engaged in efforts to establish policy that will affect our profession and students.

### References

- Betebner, D. W., & Linn, R. L. (2009). *Growth in student achievement: Issues of measurement, longitudinal data analysis, and accountability*. Paper presented at the Exploratory Seminar: Measurement Challenges Within the Race to the Top Agenda. Retrieved from <http://www.k12center.org/publications.html>
- Brookhart, S. M. (2007). Expanding views about formative classroom assessment: A review of the literature. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 43-62). New York: Teachers College Press.
- Brown, R. S., & Coughlin, E. (2007). *The predictive validity of selected benchmark assessments used in the Mid-Atlantic Region*. Washington, DC: U.S. Department of Education, Institute of Educational Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Available <http://ies.ed.gov/ncee/edlabs>
- Calmes, J. (2010). *Obama defends education program*. Retrieved from [http://www.nytimes.com/2010/07/30/education/30obama.html?\\_r=2&ref=todaypaper&utm\\_source=Newsletter](http://www.nytimes.com/2010/07/30/education/30obama.html?_r=2&ref=todaypaper&utm_source=Newsletter)
- Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2, 67-90.
- Carmichael, S. B., Martino, G., Porter-Magee, K., & Wilson, W. S. (2010). *The state of state standards—and the common core—in 2010*. Washington, DC: The Thomas B. Fordham Institute.
- Center for K-12 Assessment & Performance Management. (2010). *Enhancing assessments to support teaching and learning: Next generation assessment systems proposed under Race to the Top program*. Retrieved from <http://www.k12center.org/rsc/pdf/15051-K12Cntr-RTTbro-Digital.pdf>
- David, J. L. (2010). Using value-added measures to evaluate teachers. *Educational Leadership*, 67(8), 81-82.
- Duncan, A. (2009, June). Address to the Governors Educational Symposium. Retrieved from <http://www.ed.gov/news/speeches/2009/06/06142009.pdf>
- Duncan, A. (2009). *Robust data gives us the roadmap to reform*. Washington, DC: U.S. Department of Education.
- Eckert, J. M., & Dabrowski, J. (2010). Should value-added measures be used for performance pay? *Phi Delta Kappan*, 91(8), 88-92.
- Henderson, S., Petrosino, A. & Guckenbug, S. (2008). *A second follow-up year for "Measuring How Benchmark Assessments Affect Student Achievement"* (REL Technical Brief, REL 2008-No. 002). Regional Educational Laboratory Northeast & Islands.
- Holland, S. (2010). National learning standards make the grade. Retrieved from <http://www.cnn.com/2010/US/07/21/national.learning.standards/index.html?hpt=C2>
- Kane, T., & Staiger, D. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, 16(4), 91-114.
- Koedel, C., & Betts, J. (2010). Value added to what? How a ceiling in the testing instrument influences value-added estimation. *Education Finance and Policy*, 5(1), 54-81.
- Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard Press.
- Koretz, D. (2010). *Implications of current policy for educational measurement*. Princeton, NJ: Educational Testing Service.
- Koretz, D., & Hamilton, L.S. (2006). Testing for accountability in K-12. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 531-578). Westport, CT: Praeger.
- Koretz, D., Mccaffrey, D., & Hamilton, L. (2001). *Toward a framework for validating gains under high stakes conditions* (CSE Technical Report No. 551). Los Angeles: University of California, Center for the Study of Evaluation.
- Lachat, M. & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed at Risk*, 10(3), 333-349.
- Los Angeles Times* (2010). *Los Angeles teacher ratings: FAQs & About*. Retrieved from [http://projects.latimes.com/value-added/faq/#difference\\_value\\_added\\_api](http://projects.latimes.com/value-added/faq/#difference_value_added_api)
- National Center on Performance Incentives (NCPI). (2009). *Frequently asked questions: About our work in metropolitan Nashville Public Schools*, retrieved from [www.performanceincentives.org/about\\_ncpi/faq\\_nash.asp](http://www.performanceincentives.org/about_ncpi/faq_nash.asp)
- Nelson, M. & Eddy, R. (2008). Evaluative thinking and action in the classroom. In T. Berry & R. Eddy (Eds.), *Consequence of No Child Left Behind for Educational Evaluation: New Directions for Evaluation*, 177, 37-46.
- Niemi, D., Wang, J., Wang, H, Vallone, J., & Griffin, N. (2007). *Recommendations for building a valid benchmark assessment system: Second report to the Jackson Public*

- 
- Schools*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation.
- Pellegrino, J. W. (2009). *The design of an assessment system for the Race to the Top: A learning sciences perspective on issues of growth and measurement*. Paper presented at the exploratory seminar: Measurement challenges within the Race to the Top agenda. Retrieved from: <http://www.k12centere.org/publications.html>
- Peterson, J. (2007). Learning facts: The brave new world of data-informed instruction. *Education Next*, 7(1), 36-42.
- Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Rothstein, R. (2008). *Holding accountability to account: How scholarship and experience in other fields inform exploration of performance incentives in education*. Nashville, TN: Vanderbilt University, National Center on Performance Incentives.
- Sawchuk, S. (2010, August 8). States rush to adopt common standards as RTTT content ends. *Education Week*. Retrieved from [www.edweek.org](http://www.edweek.org)
- States rush to adopt Common Standards as RTT contest ends. *Education Week, August 11*, 8.
- Schochet, P. Z., & Chiang, H. S. (2010). Error rates in measuring teacher and school performance based on student test score gains. Washington, DC: National Center for Educational Evaluation and Regional Assistance (NCEE 2010-4004), Institute of Education Sciences.
- Trimble, S., Gay, A., & Matthews, J. (2005). Using test score data to focus instruction. *Middle School Journal*, 36(4), 26-32.
- Wiliam, D., & Leahy, S. (2007). A theoretical foundation for formative assessment. In J. H. McMillan (Ed.). *Formative classroom assessment: Theory into practice* (pp. 24-42). New York: Teachers College Press.
- Yeh, S. (2006). High-stakes testing: Can rapid assessment reduce the pressure? *Teachers College Record*, 108(4), 621-661.

---

## Mid-Western Educational Researcher

### Call for Manuscripts—2010-2011

The *Mid-Western Educational Researcher* is a peer-reviewed scholarly journal that publishes book reviews, point-counterpoints, and articles relating related to educational issues, practice, and research. While the journal considers work based on action research, this work must be relevant to the MWER readership and demonstrate rigor according to standard research conventions and practices. Membership in the MWERA is not required in order to submit a manuscript for review. If your manuscript is accepted for publication, a membership to the MWERA is required. The *MWER* (ISSN 1056-3887) is published quarterly by the Mid-Western Educational Research Association through The Ohio State University. Additional special issues are published as needed.

The journal is inviting manuscripts for review not submitted for consideration or published elsewhere in the following four sections:

#### **Book Reviews**

---

Reviews of newly released or forthcoming books or other media (e.g., film, documentary, blog, software program) that are of interest to the *MWER* readership. Reviews should include: a complete citation in APA format, a rationale for why a review of the book or media is warranted, a summary of the book or media, an analysis of strengths and weaknesses, and a statement about the significance of the book or media for educational researchers and/or practitioners. A brief summary of the reviewer's background and reasons for interest in the book or media should be included in a separate paragraph. Reviews should be between 5-10 pages. Questions regarding this section should be directed to Susan Kushner Benson at [snk@uakron.edu](mailto:snk@uakron.edu).

#### **Point-Counterpoint**

---

In this section issues of educational, pedagogical, or methodological significance will be explored from different viewpoints. Submissions should include two related manuscripts. The author(s) of the first manuscript presents a point of view on an issue and the author(s) of the second manuscript provides a different point of view on the same issue. To submit an idea for the Point-Counterpoint section, please submit a short proposal outlining the topic to be discussed, its significance, and who will be writing the point and counterpoint. Once the proposal is accepted, the point and counterpoint must be submitted as one document to the Editors within 60 days for peer review. Each manuscript should be no more than 15 pages in length including the title page and references. Questions regarding this section should be directed to Sandra Spickard Pretzman at [ssandra@uakron.edu](mailto:ssandra@uakron.edu).

#### **Research**

---

This section seeks research-based manuscripts addressing diverse issues in education. The journal does not subscribe to any particular methodology, and publishes quantitative, qualitative, mixed methods, and meta-analysis studies. Manuscripts should include an abstract of 150 words or less and be no more than 30 pages in length including title page, abstract, references, tables, and figures. All tables and figures must be in a workable text format. Questions regarding this section should be directed to Xin Liang at [liang@uakron.edu](mailto:liang@uakron.edu) or Jennifer Milam at [jenn.milam@uakron.edu](mailto:jenn.milam@uakron.edu).

#### **Graduate Student Scholarship**

---

This section is devoted to publishing graduate student research, literature reviews, and theoretical and methodological discussions. The journal does not subscribe to any particular methodology, and publishes quantitative, qualitative, mixed methods and meta-analysis studies. To submit a work to this section, the graduate student must be the sole or primary author of the work and must be a graduate student at the time of submission. Verification of graduate student status will be required if the manuscript is accepted for publication. Manuscripts should include an abstract of 150 words or less and be no more than 30 pages in length including title page, abstract, references, tables, and figures. All tables and figures must be in a workable text format. Questions regarding this section should be directed to Kristin Koskey at [koskey@uakron.edu](mailto:koskey@uakron.edu).

#### **General Formatting Guidelines**

All submissions must follow the following set of guidelines:

- All submissions must be a WORD document
- Inclusion of a title page consisting of the author's name, affiliation, address, e-mail, phone, and fax. This information should ONLY be provided on the title page.
- Headings, citations, tables, digital object identifier names (doi; see ch. 7.01), and figures should conform to the *Publication Manual of the American Psychological Association* (6th ed.)
- All text should be double spaced in 12-point Times New Roman font (tables can be in no less than 10-point font)
- 1½" margins on all sides
- 3-5 suggested keywords
- If human subjects were used, provide a statement indicating IRB approval

#### **Submit all materials via e-mail to:**

**[mwer@uakron.edu](mailto:mwer@uakron.edu)**

The subject line must include MWER followed by the section submitting to (e.g., MWER Book Review). Submissions will be subjected to a peer review process.

All manuscripts will be acknowledged electronically upon receipt. Please note that authors are responsible to submit manuscripts that are free of grammatical and mechanical errors. Manuscripts will be initially screened for format and fit. Appropriate manuscripts will be subjected to blind peer review. Reviews take approximately 4-8 weeks from the time of submission. The editors reserve the right to make minor modifications to produce a more concise and clear article. Authors acknowledge by virtue of their submission to the journal that they will consent to have their work available internationally through the EBSCO portal, as per agreement with the MWERA.



Non-profit postage paid at Columbus, Ohio, with permission of David Andrews, Dean, College of Education and Human Ecology, The Ohio State University.

Send address change to: Aimin Wang at [wanga@muohio.edu](mailto:wanga@muohio.edu)



# Mid-Western Educational Research Association Membership Application

The mission of the Mid-Western Educational Research Association is threefold: 1) to disseminate educational research conducted in the central states and provinces of North America, 2) to promote a collegial research culture in the region, and 3) to provide a forum for mentoring the research skills of graduate students and junior faculty members. Membership is open to teachers, administrators, graduate students, and educational researchers from any school, college, or university; from business or industry; or from governmental or private agencies and organizations. The annual dues (Jan-Dec) include a subscription to the publication and a reduced registration fee for the conference

**JOURNAL:** *Mid-Western Educational Researcher* (4 issues per year)

**Annual Conference:** **October 12-15, 2011**  
**At the Sheraton Westport Hotel, St. Louis, MO**

For conference information, please contact:  
**Ellen Sigler, Ph.D.**  
**MWERA-2011 Program Chair; Western Carolina University**  
**Office: (828) 227-3320 ; E-mail: esigler@ohio.edu**

Professional Member: \$60    Graduate Student: \$30    Life Member: \$450.00  
Printed Journal: \$30

## MWERA MEMBERSHIP APPLICATION

Name (First, MI, Last): \_\_\_\_\_

Mailing Address: \_\_\_\_\_  
\_\_\_\_\_

Home Phone: (\_\_\_\_) \_\_\_\_\_ Office Phone: (\_\_\_\_) \_\_\_\_\_

Email: \_\_\_\_\_ MWERA Division Preferences \_\_\_\_\_

Highest Degree Earned: \_\_\_\_\_ Institution / Employer: \_\_\_\_\_

Students: Please enclose verification of student status

\_\_\_\_ In addition to my dues, I am paying the \$30 for a printed copy of the journal \$ \_\_\_\_\_

\_\_\_\_ In addition to my dues, I am making the following unrestricted donation to MWERA: \$ \_\_\_\_\_

\_\_\_\_ Enclosed is my check, money order, or cashier's check to MWERA for \$ \_\_\_\_\_

\_\_\_\_ Please bill my Visa/MasterCard # \_\_\_\_\_

Expiration date: \_\_\_\_\_ security code: \_\_\_\_\_ in the amount of \$ \_\_\_\_\_

Name on card: \_\_\_\_\_

Signature: \_\_\_\_\_

**Send to: Sharon McNeely, Ph.D, MWERA ED    PO Box 34421    Chicago, IL 60634-0421**  
**E-mail: mwera.ed@gmail.com**



The Ohio State University  
2500 Kenny Road  
Columbus, Ohio 43210

Nonprofit Org.  
U.S. Postage  
PAID  
Columbus, Ohio  
Permit No. 711